

UNIVERSIDAD DE SEVILLA

DEPARTAMENTO DE
PSICOLOGÍA EXPERIMENTAL

**INNOVACIONES METODOLÓGICAS
EN LA EVALUACIÓN
DE LA FORMACIÓN CONTINUA**

TESIS DOCTORAL

SUSANA SANDUVETE CHAVES
DIRECTOR: DR. SALVADOR CHACÓN MOSCOSO

Sevilla, 2008

UNIVERSIDAD DE SEVILLA

DEPARTAMENTO DE
PSICOLOGÍA EXPERIMENTAL

INNOVACIONES METODOLÓGICAS EN LA EVALUACIÓN DE LA FORMACIÓN CONTINUA¹

TESIS DOCTORAL

SUSANA SANDUVETE CHAVES
DIRECTOR: DR. SALVADOR CHACÓN MOSCOSO

Sevilla, 2008

¹ La realización de este trabajo ha sido posible gracias a la participación como miembro investigador del grupo oficial del Plan Andaluz de Investigación de la Junta de Andalucía *Innovaciones metodológicas en evaluación de programas* (HUM-649) en los siguientes proyectos y convenios:

A. Proyectos de investigación:

- Proyecto de I+D del Ministerio de Educación y Ciencia, *Calidad del diseño y estimación del impacto en la evaluación de la transferencia en programas de formación continua* (SEJ 2004-05360/EDUC).
- Proyecto de Investigación de Excelencia en equipos de la Junta de Andalucía, *Calidad del diseño y estimación del impacto en la evaluación de la transferencia en programas de formación continua* (P06-HUM-01458).

B. Convenios de colaboración con las siguientes instituciones:

- UNIVERSIDAD DE SEVILLA (CENTRO DE FORMACIÓN Y PERFECCIONAMIENTO DEL PERSONAL DE ADMINISTRACIÓN Y SERVICIOS). Proyecto: *Sistemas de evaluación de los programas de formación continua de la Universidad de Sevilla* (SI-29/00; SI-44/01; SI-36/02; SI-68/03; 91/04; 30/05; SI-37/06).
- DIPUTACIÓN DE SEVILLA (SECCIÓN DE FORMACIÓN CONTINUA). Proyecto: *Evaluación de la Formación Continua promovida por la Diputación de Sevilla* (SI-94/99; SI-84/00; SI-108/01; SI-23/03; SI-39/04; SI-87-05; SI-114/06).
- INSTITUTO ANDALUZ DEL DEPORTE (DEPARTAMENTO DE FORMACIÓN). Proyecto: *Evaluación de la Formación Continua promovida por el Instituto Andaluz del Deporte* (SI-74/05; SI-75/05; SI-76/05; SI-77/05; SI-35/06).



Departamento de
Psicología Experimental

Salvador Chacón Moscoso, Profesor Titular del Departamento de Psicología Experimental de la Universidad de Sevilla

CERTIFICA: Que el trabajo titulado *Innovaciones metodológicas en la evaluación de la formación continua* realizado por Dña. Susana Sanduvete Chaves bajo mi dirección reúne las condiciones necesarias para ser presentado como Tesis Doctoral.

Lo que firmo en Sevilla, a 21 de diciembre de 2007.

Facultad de Psicología. Universidad de Sevilla. Campus Ramón y Cajal. C/ Camilo José Cela, s/n. 41018. Sevilla (España). Teléfono: 954557689. Fax: 954551784.

A mis padres

Agradecimientos

A mi familia, en especial a mis padres y hermanos, por su apoyo incondicional.

A Alfonso, por su paciencia y comprensión.

A mis amigos y amigas, que aportan algo de sentido a mi vida.

A mis compañeros del dojo Tenshinkan y, en especial, a mi Maestro Higashio, por todo lo que de ellos he aprendido y porque, poco a poco, casi sin darme cuenta, se fueron haciendo un hueco en mi vida hasta convertirse en parte fundamental de ella.

A la gente de Formación Continua de Diputación de Sevilla, los que allí están ahora y los que por allí han pasado en estos años en que hemos convivido, porque son mucho más que compañeros de trabajo.

A Juanan y Juan Ramón del Centro de Formación del Personal de Administración y Servicios de la Universidad de Sevilla, claros ejemplos de competencia, entrega y, sobre todo, ilusión en el trabajo y en la vida.

A Aurora, del Instituto Andaluz del Deporte, que creyó en mí.

A Julio, Fulgencio y demás colaboradores del equipo de meta-análisis de la Universidad de Murcia por su disponibilidad y amabilidad.

A mis compañeros del grupo Mila, Josemi, José Antonio Lozano y José Luis sin cuya ayuda inestimable, especialmente en los últimos momentos, me hubiera sido muchísimo más difícil terminar este trabajo.

A Pablo, porque con su espíritu crítico me ayuda a crecer.

A José Antonio Pérez Gil, por estar ahí cuando más lo necesité.

Un agradecimiento muy especial a Salvador Chacón, infatigable en el trabajo además de gran persona, sin cuyo apoyo la realización de este trabajo hubiera sido imposible.

Por último, a todos aquellos que aún no he nombrado que, estando ahora o habiendo pasado por mi vida en algún momento, de manera más o menos consciente me han ayudado tanto a mi crecimiento y desarrollo personal.

*“Si das pescado a un
hombre hambriento lo nutres
durante una jornada. Si le enseñas a
pescar, le nutrirás toda la vida”*

Lao Tsé

ÍNDICE DE CONTENIDOS

	Pág.
Introducción	17
1ª PARTE: ASPECTOS TEÓRICOS Y ESTADO DE LA CUESTIÓN	21
Capítulo 1. Formación y evaluación: modelos teóricos más utilizados	23
1. Formación: definición, utilidad y objetivos	25
2. La evaluación como mejora de los programas	28
2.1. Evaluación de programas: concepto y modelos teóricos	29
2.1.1. Definición de evaluación y principales características	29
2.1.2. Principales modelos tradicionales en evaluación de programas	33
2.1.2.1. Modelo de planificación de la evaluación	33
2.1.2.2. Evolución del modelo de planificación de la evaluación.....	35
2.1.2.3. Acreditación	37
2.1.2.4. Modelo dirigido a la consecución de objetivos.....	37
2.1.2.5. Modelo orientado hacia el consumidor (de meta libre)	38
2.1.2.6. Modelo experimentalista de evaluación	38
2.1.2.7. Modelo de evaluación centrado en el cliente	39
2.1.2.8. Modelo CIPP o evaluación orientada hacia la toma de decisiones.....	40
2.1.2.9. Modelo de evaluación basado en el contexto político.....	41
2.1.2.10. Fases de evaluación: el proceso de evaluación como un continuo	41
2.2. Evaluación de programas de formación continua	49
2.2.1. El modelo de Parker	49
2.2.2. El modelo de “Bell System”	49
2.2.3. El modelo de etapas	50
2.2.4. El modelo de los componentes	51
2.2.5. Modelo integrador de la evaluación de la formación.....	51
2.2.6. Directrices generales sobre cómo llevar a cabo la evaluación de planes de formación según Connolly	53
2.2.7. Recomendaciones para evaluar planes de formación según Komras	53
2.2.8. Evaluación de la formación continua de recursos humanos.....	54
2.2.9. Evaluación de la formación según Goldstein	55
2.2.10. Evaluación de la formación en el sector público	55

2.2.11. El modelo de los cuatro niveles	58
2.2.12. El quinto nivel: retorno de la inversión	62
3. Conclusiones	63
Capítulo 2. Elaboración de una escala para medir la calidad de estudios primarios	75
1. Introducción	77
2. Primera fase: compilación de ítems existentes en la literatura sobre calidad en estudios primarios y aplicación exploratoria.....	79
2.1. Compilación de ítems sobre calidad en estudios primarios	79
2.1.1. Método.....	79
2.1.1.1. Muestra	79
2.1.1.2. Instrumentos	80
2.1.1.3. Procedimiento	80
2.1.2. Resultados.....	80
2.2. Aplicación exploratoria.....	81
2.2.1. Método.....	81
2.2.1.1. Muestra	81
2.2.1.2. Instrumentos	82
2.2.1.3. Procedimiento.....	82
2.2.2. Resultados.....	83
3. Segunda fase: estudio de validez de contenido	83
3.1. Método	84
3.1.1. Muestra	84
3.1.2. Instrumentos	84
3.1.3. Procedimiento	85
3.2. Resultados	85
4. Tercera fase: elaboración de una versión depurada de la escala y aplicación exploratoria	90
4.1. Elaboración de la escala (versión depurada)	90
4.1.1. Método.....	90
4.1.1.1. Muestra	90
4.1.1.2. Instrumentos	91
4.1.1.3. Procedimiento.....	91
4.1.2. Resultados.....	91

4.2. Aplicación exploratoria.....	94
4.2.1. Método.....	94
4.2.1.1. Muestra	94
4.2.1.2. Instrumentos	94
4.2.1.3. Procedimiento.....	95
4.2.2. Resultados.....	96
5. Doble proceso deductivo-inductivo: elaboración de una versión integradora de la escala	97
5.1. Método	97
5.1.1. Muestra.....	97
5.1.2. Instrumentos	97
5.1.3. Procedimiento	97
5.2. Resultados	97
6. Discusión.....	114
7. Conclusiones	133
Capítulo 3. Estado de la cuestión en evaluación de la formación continua: aplicación de la escala de calidad	137
1. Introducción	139
2. Método	139
2.1. Muestra	140
2.2. Instrumentos	140
2.3. Procedimiento.....	140
3. Resultados.....	141
4. Discusión.....	160
5. Conclusiones	175
Capítulo 4. Eficacia de los programas de formación continua: un estudio meta-analítico.....	179
1. Introducción.....	181
2. Método	181

2.1. Muestra	181
2.2. Instrumentos	182
2.3. Procedimiento.....	182
3. Resultados	186
3.1. Descriptivos.....	186
3.2. Eficacia	189
3.3. Variables moderadoras.....	190
4. Discusión	193
5. Conclusiones	198

2ª PARTE: INNOVACIONES METODOLÓGICAS APLICADAS A CADA UNA DE LAS FASES DE EVALUACIÓN.....203

Capítulo 5. Evaluación antes de las acciones formativas: detección de necesidades209

1. Introducción	211
2. Primera fase: pre-evaluación	215
2.1. Objetivos	215
2.2. Método	215
2.2.1. Muestra.....	215
2.2.2. Instrumentos	216
2.2.3. Procedimiento	217
2.3. Resultados	217
2.4. Discusión.....	218
3. Segunda fase: evaluación principal.....	219
3.1. Objetivos	219
3.2. Método	219
3.2.1. Muestra.....	219
3.2.2. Instrumentos	220
3.2.3. Procedimiento	223

3.3. Resultados	223
3.4. Discusión.....	224
4. Tercera fase: post-evaluación	227
4.1. Objetivos	227
4.2. Método	227
4.2.1. Muestra.....	227
4.2.2. Instrumentos	228
4.2.3. Procedimiento	228
4.3. Resultados	229
4.4. Discusión.....	231
5. Diseño del Plan de Formación.....	233
5.1. Objetivos	233
5.2. Método	233
5.2.1. Muestra.....	233
5.2.2. Instrumentos	234
5.2.3. Procedimiento	234
5.3. Resultados	235
5.4. Discusión.....	235
6. Estudio comparativo de técnicas de escalamiento.....	236
6.1. Objetivos	236
6.2. Método	236
6.2.1. Muestra.....	236
6.2.2. Instrumentos	237
6.2.3. Procedimiento	237
6.3. Resultados	250
6.4. Discusión.....	253
7. Conclusiones	254
Capítulo 6. Evaluación durante las acciones formativas: diseño e implementación.....	259

1. Introducción	261
2. Método	262
2.1. Muestra	262
2.2. Instrumentos	262
2.3. Procedimiento.....	263
3. Resultados.....	264
4. Discusión.....	265
5. Conclusiones	266
Capítulo 7. Evaluación después de las acciones formativas (I). Satisfacción.....	269
1. Introducción	271
2. Método	272
2.1. Muestra	272
2.2. Instrumentos	273
2.3. Procedimiento.....	275
3. Resultados.....	276
4. Discusión.....	282
5. Conclusiones	285
Capítulo 8. Evaluación después de las acciones formativas (II). Hacia la evaluación de los niveles superiores: conocimiento y transferencia.....	289
1. Evaluación del conocimiento.....	289
1.1. Introducción	291
1.2. Método	291
1.2.1. Muestra.....	291
1.2.2. Instrumentos	293
1.2.3. Procedimiento	295
1.3. Resultados	298
1.4. Discusión.....	300

2. Evaluación de la transferencia.....	305
2.1. Introducción	305
2.2. Método	306
2.2.1. Muestra.....	306
2.2.2. Instrumentos	306
2.2.3. Procedimiento	307
2.3. Resultados	308
2.4. Discusión.....	309
3. Conclusiones	314
Capítulo 9. Conclusiones y desarrollo futuro.....	317
REFERENCIAS BIBLIOGRÁFICAS	329
Índice de autores	345
Índice de términos.....	349
ANEXOS	i
Anexo I. Ítems que contienen las escalas de medición de la calidad del diseño más frecuentemente referenciadas.....	iii
Anexo II. Estudio de validez de contenido de ítems para medir la calidad metodológica de las intervenciones.....	vii
Anexo III. Procedencia de los ítems de la escala de medición de la calidad de estudios primarios	xix
Anexo IV. Resultados obtenidos en el primer análisis bibliográfico exploratorio para determinar las características que solían presentar los estudios primarios en los ámbitos psicológico, social y educacional	xxiii
Anexo V. Ítems que obtuvieron índices de congruencia iguales o mayores que 0.5 en distintas combinaciones de los aspectos/constructos valorados	xxvii

Anexo VI. Resultados obtenidos en el segundo análisis bibliográfico exploratorio para determinar las características que solían presentar los estudios primarios en el ámbito de la formación continua	xxxii
Anexo VII. Escala de calidad utilizada para el análisis de las características de los programas de formación continua (versiones adaptadas para los diseños observacionales y encuestas, los cuasiexperimentales y los experimentales)	xxxvii
Anexo VIII. Resultados obtenidos en la priorización de acciones formativas en Diputación de Sevilla: comparación entre métodos	xlix
Anexo IX. Número de empates obtenidos por los distintos métodos de escalamiento	lxi
Anexo X. Cuestionario disponible durante la realización de cada acción formativa	lxv
Anexo XI. Listado de verificación (check list) para formadores y coordinadores en los tres momentos: antes, durante y después de la acción formativa	lxix
Anexo XII. Cuestionario de satisfacción	lxxiii
Anexo XIII. Modelo de informe referente a la calidad métrica de una prueba de evaluación de conocimientos dirigida al profesorado que la elaboró	lxxvii
Anexo XIV. Posibles instrumentos para evaluar transferencia	lxxxi

ÍNDICE DE TABLAS

	Pág.
Tabla 1.1. Resumen de las percepciones sobre la formación (Instituto de Formación y Estudios Sociales, 2002).....	26
Tabla 1.2. Diferencias principales entre la evaluación formativa y sumativa (adaptada de Brown y Gerhardt, 2002).....	32
Tabla 1.3. Ventajas e inconvenientes de la evaluación interna y externa (Anguera y Chacón, en prensa, a).....	33
Tabla 1.4. Resumen de los modelos generales de evaluación presentados.....	64
Tabla 1.5. Resumen de los modelos de evaluación de la formación continua presentados.....	67
Tabla 2.1. Características de las escalas de medición de la calidad más comunes...81	81
Tabla 2.2. Índices de representatividad, utilidad y viabilidad del dato obtenidos en cada uno de los ítems de la escala de calidad depurada	86
Tabla 2.3. Porcentaje de ítems con índices iguales o superiores a 0.5	89
Tabla 2.4. Proveniencia de cada uno de los ítems que compusieron la versión depurada de la escala de calidad.....	91
Tabla 2.5. Versión integradora de la escala de medición de la calidad en estudios primarios.....	101
Tabla 2.6. Método y resultados en cada fase de elaboración de la escala de medición de la calidad en estudios primarios.....	116
Tabla 2.7. Evolución acaecida hasta llegar a la versión integradora de la escala de calidad.....	118
Tabla 3.1. Resultados descriptivos de los ítems componentes de la escala de calidad	141
Tabla 3.2. Medias obtenidas en el índice de calidad respecto al tipo de estudio	163
Tabla 3.3. Resultados de la comparación por parejas con la prueba de Scheffé.....	164
Tabla 3.4. Puntos negativos y propuestas para mejorar la calidad de los estudios primarios.....	177
Tabla 4.1. Características descriptivas de los estudios incluidos en el meta-análisis.....	186

Tabla 4.2. Nivel global medio de eficacia en el posttest y en el seguimiento.....	189
Tabla 4.3. ANOVA de las variables moderadoras cualitativas	191
Tabla 4.4. Análisis de regresión de las variables moderadoras cuantitativas.....	192
Tabla 5.1. Resumen del proceso de detección de necesidades en Diputación de Sevilla.....	213
Tabla 5.2. Resumen del proceso de detección de necesidades en el Instituto Andaluz del Deporte	214
Tabla 5.3. Número de acciones formativas recogidas en la fase de pre-evaluación a través de la pregunta abierta del cuestionario de satisfacción	218
Tabla 5.4. Ejemplo de una viñeta del cuestionario de detección de necesidades.....	221
Tabla 5.5. Número de acciones formativas solicitadas y las que finalmente conformaron el plan a lo largo del tiempo en las distintas organizaciones.....	224
Tabla 5.6. Número de participantes en la fase de post-evaluación desde que se implementó el cuestionario como instrumento de recogida de información	228
Tabla 5.7. Ejemplo de una viñeta del cuestionario de priorización	228
Tabla 5.8. Número de acciones formativas solicitadas en cada área y número de acciones en el Plan definitivo del IAD para el año 2007.....	234
Tabla 5.9. Acciones formativas priorizadas por cada área en Diputación de Sevilla en los distintos años.....	237
Tabla 5.10. Ejemplo de priorización (método de Dunn-Rankin)	238
Tabla 5.11. Transformación de valores (método de Dunn-Rankin)	238
Tabla 5.12. Matriz de datos (método de ordenación por rangos)	240
Tabla 5.13. Suma de los órdenes asignados a cada estímulo (método de ordenación por rangos).....	240
Tabla 5.14. Matriz de frecuencia (método de los intervalos sucesivos)	241
Tabla 5.15. Matriz de frecuencias acumuladas (método de los intervalos sucesivos)	241
Tabla 5.16. Matriz de proporciones acumuladas (método de los intervalos sucesivos)	242
Tabla 5.17. Puntuaciones típicas (método de los intervalos sucesivos).....	242

Tabla 5.18. Amplitud de intervalos (método de los intervalos sucesivos).....	242
Tabla 5.19. Matriz de datos (método de los intervalos aparentemente iguales).....	245
Tabla 5.20. Matriz de frecuencias acumuladas (método de los intervalos aparentemente iguales).....	245
Tabla 5.21. Resultados obtenidos con el método de intervalos aparentemente iguales y con el método de intervalos sucesivos, tras la transformación.....	249
Tabla 7.1. Evolución de la participación en la evaluación de la satisfacción	272
Tabla 7.2. Resultados globales de las preguntas de formato cerrado del cuestionario de satisfacción en las distintas organizaciones a lo largo del tiempo ...	277
Tabla 7.3. Resultados de las preguntas de formato cerrado del cuestionario de satisfacción referidas a la coordinación.....	278
Tabla 7.4. Resultados de las preguntas de formato cerrado del cuestionario de satisfacción referidas al profesorado	278
Tabla 8.1. Participación en la evaluación de conocimiento en las distintas organizaciones a lo largo del tiempo	291
Tabla 8.2. Índices de fiabilidad y discriminación de las pruebas de evaluación de los conocimientos utilizadas en el Centro de Formación del PAS	298
Tabla 8.3. Acciones formativas donde se encontraron diferencias significativas en el estudio de las medias obtenidas en las pruebas de conocimiento	298
Tabla 8.4. Resultados globales de las preguntas de formato cerrado del cuestionario de transferencia en el Centro de Formación del PAS (año2006).....	308
Tabla 9.1. Limitaciones encontradas, mejoras ejecutadas y otras mejoras planificadas para el futuro	320

ÍNDICE DE FIGURAS

	Pág.
Figura 1.1. Validez interna y externa según Cronbach, adaptado de Cook, Campbell y Peracchio (1990)	35
Figura 1.2. Tipos de validez según Chacón y Shadish (en prensa), adaptado de Shadish, Cook y Campbell (2002a)	36
Figura 1.3. Esquema no-lineal de evaluación (Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999)	42
Figura 1.4. Diseño con grupo control no equivalente y prueba previa y posterior. .	47
Figura 1.5. Diseño de cohortes básico.	47
Figura. 1.6. Diseño de discontinuidad en la regresión.....	48
Figura 1.7. Características del modelo de los cuatro niveles de evaluación de Kirkpatrick (1999), adaptada de Phillips (1990)	62
Figura 3.1. Tipo de estudio.....	146
Figura 3.2. Grupo de comparación	147
Figura 3.3. Criterios de inclusión y exclusión de las unidades de la muestra.....	147
Figura 3.4. Azar	147
Figura 3.5. Diseño.....	148
Figura 3.6. Muestra.....	148
Figura 3.7. Mortalidad global.....	148
Figura 3.8. Mortalidad diferencial	149
Figura 3.9. Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones	149
Figura 3.10. Seguimiento	150
Figura 3.11. Momentos de medida	150
Figura 3.12. Medidas que aparecen en todos los momentos de registro.....	151
Figura 3.13. Variables dependientes normalizadas	151

Figura 3.14. Enmascaramiento del evaluador	151
Figura 3.15. Enmascaramiento del usuario	152
Figura 3.16. Enmascaramiento del formador (como evaluador interno cuando no hay intervención)	152
Figura 3.17. Homogeneidad de la intervención o del proceso de registro cuando no hay intervención.....	153
Figura 3. 18. Definición de los constructos.....	153
Figura 3.19. Métodos estadísticos para inferir los valores perdidos.....	153
Figura 3.20. Tamaño de efecto especificado.....	154
Figura 3.21. Índice de calidad	154
Figura 3.22. Índice estadístico calculado	155
Figura 3.23. Diferencias estadísticamente significativas.....	155
Figura 3.24. Índice de variabilidad facilitado.....	155
Figura 3.25. Número de participantes en cada grupo	156
Figura 3.26. Número de grupos en el estudio	156
Figura 3.27. Exclusiones tras medidas posteriores.....	156
Figura 3.28. Rango de edad especificado	157
Figura 3.29. Media de edad.....	157
Figura 3.30. Periodo de estudio	158
Figura 3.31. Intensidad del tratamiento/registro.....	158
Figura 3.32. Unidades de intervención o de registro	158
Figura 3.33. Área formativa	159
Figura 3.34. Campo de intervención.....	159
Figura 3.35. Tipo de publicación.....	160
Figura 3.36. Índice de calidad tras excluir los estudios teóricos	162
Figura 3.37. Índice de calidad distribuido según el tipo de diseño	163

Figura 1 (2ª introducción). Esquema no-lineal de evaluación (Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999)	206
Figura 5.1. Orden de priorización de las acciones formativas del área de conocimiento B “Económico-presupuestaria” en el año 2004 en Diputación de Sevilla	230
Figura 5.2. Orden de priorización de las acciones formativas del área de conocimiento A “Gestión y administración deportiva” en el año 2006 en el IAD	230
Figura 5.3. Límites de las categorías (Método de los intervalos sucesivos)	243
Figura 5.4. Ordenación resultante (Método de los intervalos sucesivos)	244
Figura 5.5. Cinco intervalos de igual amplitud (Método de los intervalos aparentemente iguales).....	245
Figura 7.1. Evolución del porcentaje de participación en la evaluación de la satisfacción con las acciones formativas.....	273
Figura 7.2. Medias obtenidas en la evaluación de la satisfacción con los cursos en el Centro de Formación y Perfeccionamiento del PAS	278
Figura 7.3. Medias obtenidas en la evaluación de la satisfacción con los cursos en Diputación de Sevilla	279
Figura 7.4. Medias obtenidas en la evaluación de la satisfacción con los cursos en el Instituto Andaluz del Deporte	279
Figura 7.5. Medias obtenidas en la evaluación de la satisfacción respecto a la coordinación en Diputación de Sevilla.....	280
Figura 7.6. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en el Centro de Formación y Perfeccionamiento del PAS.....	280
Figura 7.7. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en Diputación de Sevilla	281
Figura 7.8. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en el Instituto Andaluz del Deporte.....	281
Figura 8.1. Porcentaje de acciones formativas donde se llevó a cabo la evaluación de conocimientos en las distintas organizaciones a lo largo del tiempo.....	292
Figura 8.2. Evolución del porcentaje de acciones formativas que presentaron diferencias estadísticamente significativas en el estudio de evaluación de Conocimientos	299

Figura 8.3. Diseño de un solo grupo con prueba previa y posterior.....	302
Figura 8.4. Diseño con grupo control no equivalente y prueba previa y posterior ..	304
Figura 8.5. Medias obtenidas en la evaluación de la transferencia en el Centro de Formación del PAS en el año 2006	309
Figura 8.6. Diseño con medida previa y múltiples medidas posteriores	312

La **formación continua** puede ser definida como un conjunto planificado de acciones cuya función es aportar a los participantes conocimiento y/o habilidades que permitan la mejora del desempeño en sus puestos de trabajo (Buckley y Caple, 1991).

Actualmente, las distintas instituciones públicas y privadas conceden a este tipo de formación una gran importancia dentro del marco del desarrollo económico en España acorde con el del resto de Europa, tal como se desprende de las cifras invertidas desde el Fondo Social Europeo (2007).

Se entiende que el enriquecimiento de las cualidades profesionales de los trabajadores que con la formación continua se busca, promueve el empleo estable y la mejora del servicio que el trabajador ofrece, lo que se traduce en el término tan en boga “desarrollo sostenible”.

En este contexto, y con este trabajo de tesis doctoral se plantea que la evaluación es considerada imprescindible no sólo para analizar y demostrar el valor de la formación, sino también para su planificación e implementación. Como cualquier programa, es fundamental tratar de demostrar empíricamente que está funcionando desde un referente adecuado con los modelos teóricos existentes.

Concretamente en programas de formación, esto es especialmente necesario porque se invierten muchos recursos, de manera obligada por la normativa actual en el caso de las instituciones públicas. Y lograr sacar lo máximo de estos programas tiene consecuencias muy positivas a nivel individual, tanto profesional como personal, para los trabajadores y quienes se benefician de su trabajo, y a nivel organizacional para la empresa en la que trabajan.

Este trabajo pretende aportar directrices claras para llevar a cabo tanto el diseño como la evaluación de programas de formación continua; así, en la primera parte, tras una exhaustiva revisión teórica, la elaboración de una versión integradora de una escala de medición de la calidad de estudios primarios, la aplicación de dicho instrumento en el ámbito de la formación continua y un meta-análisis, se intenta determinar qué aspectos metodológicos habrían de tenerse en cuenta para optimizar los programas de formación y en qué sentido se podrían mejorar; y en la segunda parte del trabajo, se presenta un ejemplo práctico en distintos planes de formación que trata de aportar evidencias empíricas de la validez de la escala previamente mencionada.

Concretamente, en la **PRIMERA PARTE** se presentan cuatro capítulos.

En el **capítulo 1**, tras un estudio de la literatura sobre evaluación de programas en general y en particular sobre evaluación de la formación publicada hasta septiembre de 2006, se observa cómo los constructos de las variables sobre las que se actúa en procesos de formación continua raramente son delimitadas y/o definidas operativamente. Esta situación es consecuencia de varios aspectos:

- La acumulación, en gran medida desorganizada, de modelos teóricos tanto sobre evaluación en general como sobre formación continua en particular.

- Los modelos teóricos son generales y por tanto difíciles de adaptar para su aplicación a condiciones particulares de intervención; presentan dimensiones genéricas, no contextualizadas en ámbitos representativos de intervención, en lugar de delimitar variables operativas.

Con la intención de aportar información organizada, sistematizada y presentada de la manera más operativa posible para que resulte útil para elegir en qué modelo basar la evaluación en función del contexto, se presentan los principales modelos al uso considerando los siguientes aspectos, siempre que se expliciten (información disponible en <http://innoevalua.us.es>):

- El **objetivo** del modelo definiendo operativamente las dimensiones que plantea.
- Las **fases** (metodología) para su desarrollo y la operacionalización de cada una de ellas.
- La **utilidad** del modelo; dónde puede aplicarse.

Y para concretar aún más, se especifica para cada modelo (siempre que la información sea dada) el acrónimo “u t o s t_i” (Chacón y Shadish, en prensa, adaptado de Shadish, Cook y Campbell, 2002a). Se determinan por tanto los usuarios a los que el modelo puede adaptarse (units), la intervención que se propone (treatment), los resultados que se recogen (output), el contexto (setting) y el momento de medida (time).

En este capítulo se llega a la conclusión, de manera general, a tres aspectos concretamente: a) los modelos carecen de operativización de las variables que entran en juego en la evaluación de programas; b) los instrumentos que se utilizan no suelen presentar estudios de validez; c) las evaluaciones suelen ser sumativas o de resultados (no se evalúa el proceso muy a menudo). Pero, ¿qué ocurriría si se les preguntara acerca de esto a los expertos? ¿Qué variables considerarían claves para determinar la calidad de un programa de intervención? ¿La calidad del dato recogido? ¿El número de medidas tomadas y los momentos de medida? ¿Serán estos datos viables?

Para responder a esto, en el **capítulo 2** se elaboró una escala de medición de la calidad metodológica de estudios primarios, siguiendo un procedimiento deductivo-inductivo al perfilar el instrumento desde la bibliografía, posteriormente tras la obtención de datos empíricos en distintos contextos, de nuevo volviendo al estudio de la bibliografía y, así, sucesivamente.

Concretamente, los pasos que se siguieron fueron los siguientes: se recogieron, según la bibliografía, los ítems que miden la calidad de las intervenciones; se hizo un primer estudio exploratorio aplicando el instrumento a programas de intervención psicológica; se realizó posteriormente un estudio de validez de contenido donde expertos en diseño y evaluación de programas, calidad, y revisiones sistemáticas y meta-análisis, puntuaron tres dimensiones (representatividad, utilidad y disponibilidad del dato); se realizó otra versión de la escala donde se incluyeron nuevos ítems propuestos por expertos y donde se aclararon los conceptos que provocaron dudas; se realizó otra aplicación del instrumento, esta vez en el ámbito de la formación continua;

y, finalmente, se llevaron a cabo las modificaciones necesarias para que el instrumento elaborado fuera aplicable a todo tipo de estudio sin diferenciación.

Esta última característica es la mayor novedad aportada cuando se compara esta escala con otras previamente elaboradas y publicadas, ya que la literatura en este ámbito presenta un claro sesgo hacia los estudios experimentales. Sin embargo, la última versión de la escala no se presenta como algo definitivo y cerrado, sino como la propuesta más novedosa dentro de un proceso deductivo-inductivo que aún se encuentra en marcha.

En el **capítulo 3**, la versión más reciente de la escala es aplicada a los trabajos encontrados referidos a formación continua para vislumbrar en qué aspectos se podría incidir para mejorar estos programas. Una vez detectados los puntos débiles y enumeradas las posibles razones de la existencia de dichas debilidades, se proponen mejoras para optimizar los programas de formación continua.

Pero, ¿son eficaces los programas de formación continua publicados? Si la respuesta es afirmativa, ¿qué variables están relacionadas con su éxito? ¿Coincide esta información con la hasta ahora recogida proveniente de los modelos teóricos de evaluación, la bibliografía recogida referida a calidad, la opinión de los expertos en la materia y la evaluación de los programas de formación publicados?

Para responder a estas preguntas, en el **capítulo 4** se realiza un estudio meta-analítico con las publicaciones que tratan sobre formación continua que presentan la información suficiente para su inclusión. Se determina el grado de eficacia presentado por estos programas y las variables moderadoras que provocan la heterogeneidad entre los resultados encontrados.

A continuación, en la **SEGUNDA PARTE** de este trabajo, y siguiendo las fases de evaluación de programas propuesta por Anguera y Chacón (en prensa, a) y Alvira (1991), también recogido por el Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla (1999); tomando en detección de necesidades el modelo de Chacón, Lara y Pérez (2002c) adaptado de Witkin y Altschuld (1995); y los pasos de la evaluación de resultados según Kirkpatrick (1999), se presenta una visión más práctica de lo hasta ahora tratado a través de su aplicación a la formación continua realizada por tres organizaciones distintas: Diputación de Sevilla, el Centro de Formación del Personal de Administración y Servicios de la Universidad de Sevilla y el Instituto Andaluz del Deporte. Se trata no sólo de describir la evaluación en estas instituciones, sino además de proponer mejoras para el futuro y aportar información **operativizada** y concreta para que cualquier interesado pudiera aplicar a cualquier plan de formación estas actuaciones.

Concretamente, en referencia a la evaluación llevada a cabo **antes** de la realización de las acciones formativas, el **capítulo 5** aporta una descripción del estudio de detección de necesidades donde se explica el diseño utilizado y el proceso seguido para la recogida de los datos. Como innovación, se analizan éstos con distintos métodos de escalamiento para estudiar cuál, siendo el más sencillo, resulta más útil para la posterior toma de decisiones basada en la evidencia. Concretamente, los métodos utilizados son el de Dunn-Rankin (Arce, 1994) y tres métodos elaborados por Thurstone

que parten de la Ley de juicios categóricos (Barbero, 1993): intervalos sucesivos, intervalos aparentemente iguales y ordenación por rangos.

El **capítulo 6** hace referencia a la evaluación **durante** la realización de las acciones formativas. Concretamente, se aporta una descripción del procedimiento a llevar a cabo para tratar de encontrar posibles puntos que mejorar en el mismo momento en que se detectan y se muestran posibles instrumentos de recogida de esta información.

Los siguientes dos capítulos hacen referencia a la evaluación de resultados **después** de finalizar la realización de las acciones formativas. En el **capítulo 7**, se describe la evaluación de la satisfacción de los participantes en las acciones formativas y se trata de detectar las debilidades del proceso y los resultados y, a partir de ahí, proponer posibles mejoras para solventarlas.

En el **capítulo 8**, se presenta la descripción de la evaluación de los conocimientos adquiridos por los participantes; entre otras cuestiones, se presentan unas posibles directrices a seguir para elaborar pruebas de medición de conocimientos; además, se presenta una visión descriptiva de la evaluación de la transferencia de lo aprendido al puesto de trabajo donde, entre otras cuestiones, se aporta un posible cuestionario como instrumento de recogida de información.

El **capítulo 9** recoge las conclusiones obtenidas tras la realización de este trabajo, sirve como reflexión acerca de los resultados obtenidos y presenta posibles vías de desarrollo futuro.

1ª PARTE:

**ASPECTOS TEÓRICOS
Y ESTADO DE LA CUESTIÓN**

CAPÍTULO 1. FORMACIÓN Y EVALUACIÓN: MODELOS TEÓRICOS MÁS UTILIZADOS

1. Formación: definición, utilidad y objetivos.....	25
2. La evaluación como mejora de los programas	28
2.1. Evaluación de programas: concepto y modelos teóricos	29
2.1.1. Definición de evaluación y principales características	29
2.1.2. Principales modelos tradicionales en evaluación de programas	33
2.1.2.1. Modelo de planificación de la evaluación.....	33
2.1.2.2. Evolución del modelo de planificación de la evaluación.....	35
2.1.2.3. Acreditación	37
2.1.2.4. Modelo dirigido a la consecución de objetivos	37
2.1.2.5. Modelo orientado hacia el consumidor (de meta libre)	38
2.1.2.6. Modelo experimentalista de evaluación.....	38
2.1.2.7. Modelo de evaluación centrado en el cliente	39
2.1.2.8. Modelo CIPP o evaluación orientada hacia la toma de decisiones	40
2.1.2.9. Modelo de evaluación basado en el contexto político	41
2.1.2.10. Fases de evaluación: el proceso de evaluación como un continuo	41
2.2. Evaluación de programas de formación continua	49
2.2.1. El modelo de Parker	49
2.2.2. El modelo de “Bell System”	49
2.2.3. El modelo de etapas	50
2.2.4. El modelo de los componentes	51
2.2.5. Modelo integrador de la evaluación de la formación.....	51
2.2.6. Directrices generales sobre cómo llevar a cabo la evaluación de planes de formación según Connolly	53
2.2.7. Recomendaciones para evaluar planes de formación según Komras	53
2.2.8. Evaluación de la formación continua de recursos humanos.....	54
2.2.9. Evaluación de la formación según Goldstein	55
2.2.10. Evaluación de la formación en el sector público	55
2.2.11. El modelo de los cuatro niveles	58
2.2.12. El quinto nivel: retorno de la inversión	62
3. Conclusiones	63

CAPÍTULO 1. FORMACIÓN Y EVALUACIÓN: MODELOS TEÓRICOS MÁS UTILIZADOS

1. FORMACIÓN: DEFINICIÓN, UTILIDAD Y OBJETIVOS.

Un **programa de intervención** se define como un conjunto específico de acciones y recursos materiales diseñados e implantados organizadamente en una determinada realidad social con el propósito de resolver algún problema que atañe a un conjunto de personas (Fernández-Ballesteros, 1995).

Las **características** principales que muestra cualquier programa de intervención son (Anguera y Chacón, en prensa, a):

- **Planificación:** antes de llevar a cabo un programa, es muy importante que su diseño esté especificado por escrito de manera muy concreta, pues esto ayudará a detectar posibles discrepancias de lo ejecutado con lo previamente diseñado y permitirá el análisis de posibles aspectos a mejorar en busca de la consecución de los objetivos propuestos.
- **Sistematización:** supone el análisis pormenorizado de todos los aspectos que intervienen en el programa, sin por ello caer en una visión miope de la situación; muy al contrario, esta forma de proceder ayuda a tener una visión holística de las actuaciones, a la vez que da capacidad para detectar qué cuestiones singulares podrían mejorarse.
- **Delimitación de recursos humanos y materiales:** de nada serviría planificar una intervención de manera sistemática y con alto grado de especificación si, a la hora de la puesta en marcha, se cae en la cuenta de que la ejecución supone unos costes tanto materiales como de recursos humanos excesivamente elevados a los que no se puede hacer frente. Por ello, es fundamental que, en el momento de la planificación, todas las actividades se acompañen de una aproximación real de los costes que su realización va a conllevar, logrando de este modo programas realistas y pragmáticos, más que utópicos e inalcanzables. Por supuesto que es necesario marcarse metas ideales pero, para llegar a ellas, es preciso ir paso a paso con los pies en el suelo.
- **Temporalización:** no basta con tener claro qué objetivos se pretenden alcanzar; también se ha de determinar en cuánto tiempo se quieren conseguir; porque, al igual que es posible que la falta de recursos humanos y materiales lleve un programa al traste, también puede ocurrir por falta de tiempo. Se recomienda que no sólo los objetivos estén temporalizados, sino que también las actividades concretas a través de las cuales se quieren conseguir esos objetivos estén delimitadas temporalmente y con un orden de ejecución determinado.

Concretamente, un **programa de formación continua**, que es la temática concreta que en este trabajo se trata, es un esfuerzo sistemático y planificado por diseñar un conjunto específico de acciones formativas con el propósito de potenciar el conocimiento, las técnicas y actitudes de los participantes en relación a aquellos

aspectos que les sean útiles para mejorar el desempeño en sus puestos de trabajo (Buckley y Caple, 1991).

En el actual contexto complejo y cambiante en el que vivimos, la formación continua es una **necesidad**. Nuevas circunstancias en todos los ámbitos nos llegan constantemente (un nuevo programa informático, el aprovechamiento de otros recursos naturales en busca del desarrollo sostenible, nuevas problemáticas sociales como por ejemplo la llegada masiva de inmigrantes a costas españolas, etc.). Ante esto, los profesionales se ven en la necesidad de “reciclarse”, de actualizar sus conocimientos, para potenciar al máximo sus capacidades con miras a mejorar la realización de sus tareas laborales. Y la organización, consciente de esta necesidad, asume la responsabilidad que a este respecto le atañe, de tal manera que la tendencia hoy día es que la empresa se preocupe porque sus empleados reciban la formación necesaria para estar al día en los temas relacionados con sus funciones.

Los principales **objetivos** que pretende alcanzar la formación continua son (Chacón, Holgado, López y Sanduvete, 2006):

- **Promoción interna:** aumentar los conocimientos de los empleados para fomentar el cambio a puestos de trabajo con mayor responsabilidad y mejores condiciones laborales dentro de la organización.
- **Actualización:** poner al día los conocimientos de los empleados en función de la evolución que el puesto que desempeña haya experimentado.
- **Prevención:** formar para solventar posibles necesidades que puedan surgir en un futuro inmediato.

En función del objetivo que prime, la **formación será percibida** por la organización de manera distinta tal y como se muestra, de manera resumida, en la tabla 1.1 (Instituto de Formación y Estudios Sociales, 2002):

1. GASTO NECESARIO “Aprender para trabajar”	2. BENEFICIO SOCIAL “Aprender para saber”
3. SISTEMA INDIRECTO DE RETRIBUCIÓN “Aprender como premio”	4. INVERSIÓN EMPRESARIAL “Aprender para desarrollar”

Tabla 1.1. Resumen de las percepciones sobre la formación (Instituto de Formación y Estudios Sociales, 2002).

1. **Gasto necesario:** la formación se percibe como un gasto necesario cuando todas las acciones van dirigidas a mejorar el desempeño en el puesto de trabajo, sin perseguir el desarrollo personal del empleado.
2. **Beneficio social:** la formación es percibida como un beneficio social cuando el objetivo principal que consigue alcanzar es el desarrollo del individuo.
3. **Sistema indirecto de retribución:** la formación se percibe como un sistema de retribuciones cuando se utiliza como premio a los trabajadores por actuaciones consideradas beneficiosas para alcanzar los objetivos de la organización.

- 4. Inversión empresarial:** la formación es considerada una inversión cuando se percibe como trabajo necesario en la actualidad para recoger un fruto en el futuro.

La realidad es que, en general, los cuatro tipos de percepción de la formación se dan en mayor o menor medida en todas las organizaciones.

Y para la consecución de los objetivos planteados en las acciones formativas, el **ponente** o **formador** adquiere un papel fundamental. Contraria a la visión tradicional donde esta figura tenía prácticamente todo el peso de la enseñanza y el alumnado quedaba relegado a un segundo plano, la idea preponderante actual es que, coordinando y guiando, consiga ceder el papel principal al alumnado, haciéndoles así partícipes directos de su propio aprendizaje. Algunas cualidades que ha de tener un buen formador son las siguientes (di Paolo y Patterson, 1983):

- **Conocimiento de la teoría y la práctica:** el ponente no sólo ha de transmitir conceptos teóricos sino que, siendo consciente de que se trata de lograr la transferencia de lo aprendido al puesto de trabajo, ha de dar una visión eminentemente aplicada a todo lo que imparte.
- **Planificación de la instrucción:** es función del ponente el diseño de la acción formativa, consistente en la definición de los objetivos a alcanzar, los contenidos a impartir y su temporalización, el estudio del perfil del alumnado y la adaptación de los contenidos a sus características, y la evaluación de dicha formación.
- **Guía de la instrucción:** otra función del formador es, durante la realización de la acción formativa, mantener la atención del alumnado sobre el contenido a tratar y las actividades a realizar, además de anticipar soluciones a posibles imprevistos referidos a factores ambientales como, por ejemplo, que las mesas y sillas estén ordenadas en fila cuando se pretendía realizar actividades dinámicas, que no haya pizarra o algo similar para escribir, etc.
- **Técnicas de presentación:** el primer paso a realizar en el contexto formativo es la presentación tanto del formador como del alumnado, con lo que se conseguirá romper el hielo (lo cual facilitará la participación de todos) y tener una idea global de la posible diversidad existente respecto al perfil de los distintos participantes.
- **Motivación:** un punto esencial para la consecución de los objetivos es que el formador sea capaz de motivar a su alumnado. Algunas posibles técnicas para ello son hacer actividades lo más cercanas posible a lo que será después la realización de la tarea en el puesto de trabajo, fomentar la participación creando un ambiente distendido entre los participantes, planificar y ejecutar ejercicios donde el mayor peso de la actividad recaiga sobre el alumnado o usar ejemplos sencillos y clarificadores.
- **Estrategias de instrucción:** existen numerosos métodos para conseguir que las presentaciones no sean meramente magistrales; aparte de la clásica charla de

experto se pueden realizar, entre otras técnicas, juegos, simulaciones de posibles circunstancias que se pueden dar en el puesto de trabajo, role-play o debates entre los participantes.

- **Comunicación:** una de las tareas principales del ponente es la transmisión de conocimientos. Por ello, se hace fundamental que se exprese con sencillez y fluidez, que detecte el posible feedback que el alumnado le haga llegar, que tenga habilidades para crear y guiar el trabajo en grupos, que tenga en cuenta el uso de la comunicación no verbal, que sea capaz de promover la participación planteando cuestiones apropiadas, etc.

El hecho de que los programas de formación que aquí se tratan se califiquen de **continuos** se debe a que la intervención no se concibe como algo puntual, sino con permanencia a lo largo del tiempo, como parte importante de la organización. Es por ello que la formación no se considera una meta en sí misma, sino un **instrumento** que se utiliza para la mejora de la calidad del servicio prestado por la organización a los usuarios (Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999; 2001; Chacón, Pérez, Holgado y Lara, 2001b).

2. LA EVALUACIÓN COMO MEJORA DE LOS PROGRAMAS.

Una vez definido en qué consiste un programa de formación y cuáles son sus objetivos, comenzaremos a centrarnos en la evaluación de estos programas como instrumento de mejora continua a partir de un método científico que pretende ser lo más objetivo posible para que, de este modo, los programas resulten resistentes de cara a posibles manipulaciones que puedan intentar realizar los distintos implicados.

Se realizó un estudio bibliográfico donde se encontraron 121 trabajos referentes a la evaluación de la formación continua (la metodología utilizada en este estudio se describe más pormenorizadamente en el capítulo 3).

A niveles generales puede decirse que, entre estos trabajos, se encontró un elevado número de modelos teóricos presentados de manera algo caótica y desorganizada, y tan generales que resultaba difícil aplicarlos en los distintos contextos de intervención.

Por ello, para facilitar la comprensión de estos modelos y la utilidad de la exposición, se presentará cada uno de ellos siguiendo dos esquemas: por un lado, se marcarán los objetivos operacionalizados, las fases de evaluación y su utilidad; y en segundo lugar, se describirán los elementos del acrónimo “u t o s t_i” (Chacón y Shadish, en prensa, adaptado de Shadish, Cook y Campbell, 2002a). Se determinarán por tanto los usuarios a los que el modelo puede adaptarse (units), la intervención que se propone (treatment), los resultados que se recogen (output), el contexto (setting) y el momento de medida (time).

Concretamente, en primer lugar se definirá la **evaluación a nivel general**, útil para todos los campos de intervención, y se presentará una compilación de los modelos teóricos más importantes.

En segundo lugar, centrándonos en el ámbito concreto de la **evaluación de los programas de formación continua** se mostrarán también, entre otros aspectos conceptuales, los modelos teóricos más utilizados.

2.1. Evaluación de programas: concepto y modelos teóricos.

2.1.1. Definición de evaluación y principales características.

El término **evaluación** ha sido definido de maneras muy diversas en función de en qué aspecto se haya centrado cada uno de los autores (Anguera y Chacón, en prensa, a). A pesar de esta pluralidad, existe consenso en incluir en la definición algunos aspectos:

- La evaluación supone una emisión de **juicios de valor** (Scriven, 1967).
- Estos juicios de valor han de basarse en un **estudio sistemático que recoja datos válidos** acerca de los resultados obtenidos por el programa (Stufflebeam y Shrinkfield, 1987; Fernández-Ballesteros, 1995).
- Los resultados de la evaluación han de ser **útiles** a los distintos implicados (Weiss, 1973; Wholey, 1983; Alvira, 1991). Según Stufflebeam y Shrinkfield (1987), se pueden distinguir principalmente tres usos:
 - **Recapitulación:** se centra en el estudio de los programas ya terminados. Consiste en el estudio de los resultados y la difusión de esta información entre los distintos implicados.
 - **Perfeccionamiento:** los datos que la evaluación recoge tienen como función la mejora de la calidad del servicio. Por tanto, se centra no sólo en los resultados, sino también en el proceso.
 - **Ejemplificación:** consiste en el uso de la evaluación no sólo de manera práctica, sino también para tratar cuestiones teóricas.

Es deseable que una buena evaluación cumpla las siguientes **condiciones**:

- **Utilidad:** se ha de intentar devolver la información a los distintos implicados a través de informes en los que se explique con claridad lo obtenido con la evaluación.
- **Factibilidad:** se han de planificar procesos de evaluación que sean viables; que puedan llevarse a cabo en la realidad. De nada servirá diseñar una evaluación si, por ejemplo, la falta de recursos disponibles impedirá que se ejecute.
- **Ética:** los resultados de una evaluación no han de ser manipulados. Además, han de respetarse los derechos de los distintos implicados.
- **Exactitud:** Se ha de dejar claro cuál es el objetivo que se pretende con la

evaluación. Los resultados obtenidos han de responder de manera fiable y válida a tal objetivo.

En este sentido, Cook (1997) propone algunos puntos importantes a tener en cuenta como **guía** para las evaluaciones que se quieran llevar a cabo:

- La evaluación ha de entenderse de manera **ecléctica**; es decir, existe una gran diversidad de metodologías que serán más o menos adecuadas en función del objetivo que persigue la evaluación y que, además, pueden ser complementarias como, por ejemplo, la recogida y posterior análisis de datos cualitativos y cuantitativos.
- Es recomendable establecer los **canales de comunicación** adecuados con los que difundir los resultados de la evaluación a todos los implicados. La desinformación a algunos implicados puede ser interpretada como un intento de manipulación en busca de unos intereses concretos.
- Para fomentar el uso de la evaluación es indispensable, entre otras cuestiones, dotar de **credibilidad** tanto al proceso como a los resultados. Para lograrlo, el primer paso es la transparencia en todo momento (explicar, por ejemplo, cómo se llegaron a los resultados y por qué se tomaron esas decisiones en concreto y no otras).
- Es aconsejable que las evaluaciones también sean **evaluadas**, para estar totalmente seguros de que los resultados son fiables y por si pudiera mejorarse el proceso de alguna manera; es decir, los profesionales de la evaluación tendrían que cuestionar de manera continua su propio trabajo en términos de objetividad y sensibilidad (Chelimsky, 1997).

En contraposición, éstas son las **características que en cualquier evaluación se tendrían que evitar** (Anguera y Chacón, en prensa, a):

- **Pseudoevaluación:** se da cuando a) no se lleva a cabo realmente una evaluación debido a que no se especificó el objetivo que se pretendía alcanzar o la manera en que se recogerían los datos; o b) se deciden de antemano los resultados que han de encontrarse, con lo que todos los esfuerzos durante el proceso se centran en que finalmente se encuentre lo que se esperaba.
- **Identificación de medida con evaluación:** tener en cuenta para evaluar únicamente los datos de manera aislada es insuficiente; habría que tener en cuenta el contexto, determinante para llegar a una comprensión más exacta del tema a evaluar pues, como dijera Ortega y Gasset (cit. en Bria, Doltra, Moreno, Pedrals, Juan y Boldú, 2004), “Yo soy yo y mis circunstancias”.
- **Sectarismo o parcialidad:** se da cuando, por alguna circunstancia, la evaluación responde a algún tipo de prejuicio o interés. Puede surgir sectarismo o parcialidad en distintos momentos de la evaluación: a) en la **selección** de la información, cuando sólo se tiene en cuenta aquello que puede casar con el resultado esperado de la evaluación; b) en la **obtención** de información cuando, por ejemplo, únicamente se pregunta acerca del tema a un grupo concreto de

implicados y a otros no; y c) en el **momento de difusión** como, por ejemplo, cuando la oposición saca los resultados de algún estudio que afecta negativamente al partido que gobierna, justo antes de las elecciones.

- **Subjetivismo:** hay más riesgo de no actuar de manera objetiva en la evaluación cuanto menos fundamentados estén los datos. Por ello, es recomendable que la recogida de datos sea rigurosa y el análisis de éstos se realice con métodos estadísticos, para que objetivamente sirvan de argumentación a la hora de tomar decisiones.
- **Confusión de niveles:** es interesante conocer quién solicita qué información y para qué la quiere, porque esto puede dar respuesta a qué intereses ocultos podrían estar en juego. Ante una misma situación, en función del demandante de la evaluación, los intereses pueden ser totalmente diferentes. Así, por ejemplo, a los **políticos** les podría interesar averiguar si son adecuadas las políticas de intervención que se están llevando a cabo o si sería necesario modificar la financiación; los **directivos** de una empresa podrían preguntarse si el programa consiguió sus metas y qué mejoras podrían hacerse; quizá los **profesionales que han llevado a cabo directamente la intervención** quisieran conocer si su trabajo ha sido efectivo y cómo podría mejorarse; a los **usuarios** y los **allegados** quizá les interesaría saber si el programa les podría ayudar a cumplir sus expectativas; finalmente, los **potenciales usuarios** podrían preguntarse si sería interesante participar en el programa.
- **Burocratización:** ocurre muy a menudo que la rutina y los trámites administrativos pueden llegar a entorpecer y entorpecer el proceso de evaluación, pues el mínimo paso que se quiera dar o la mínima innovación que se pretenda implantar puede, por ejemplo, precisar de permisos sin los cuales no existe modo de seguir adelante.

Existen muchas clasificaciones de tipos de **evaluación**; una de las más comunes y que aún hoy día se utiliza a pesar de su antigüedad es la presentada por Scriven (1967) al diferenciar entre la función **formativa** y la **sumativa** de la evaluación: la función formativa se lleva a cabo paralelamente a la intervención, a lo largo de todo el programa y, al obtener información durante todo el proceso, da la posibilidad de detectar in situ posibles aspectos a mejorar y cambiarlos en el momento; la función sumativa se da una vez finalizado el programa que se ha ido reajustando con la función formativa y la finalidad es la de servir de ayuda para tomar decisiones acerca de la continuidad de éste. En resumen, podría decirse que la función sumativa se encarga de emitir juicios sobre los resultados obtenidos, mientras que la formativa se centra en el proceso. En la tabla 1.2 que a continuación se presenta, aparecen las diferencias más importantes entre la evaluación sumativa y la formativa (adaptada de Brown y Gerhardt, 2002).

	EVALUACIÓN FORMATIVA	EVALUACIÓN SUMATIVA
Propósito	Detectar debilidades concretas para mejorarlas en el momento.	Determinar la eficacia de la intervención para mantenerla o eliminarla.
Momento de medida	A través de todo el proceso.	Al final de la intervención.
Resultado	Unas directrices para revisar la intervención y mejorarla.	Un informe sobre los resultados finales y recomendaciones acerca de la continuidad del programa.

Tabla 1.2. Diferencias principales entre la evaluación formativa y sumativa (adaptada de Brown y Gerhardt, 2002).

Otra clasificación muy común va en función de **quién** realiza la evaluación, distinguiéndose estos tipos:

- **Evaluación externa:** se da cuando los evaluadores no son las mismas personas que ejecutan el programa y no forman parte de la organización evaluada. Esta forma de evaluar tiene algunas **ventajas**, como el hecho de que los **resultados** son más **creíbles** debido a que son personas ajenas y, por tanto, no tienen intereses especiales a favor del programa que puedan sesgar los resultados; o que **maximiza el uso de estándares**, con lo que se consigue una evaluación comparable con la de otros programas, se facilita la creación de instrumentos pues pueden venir ya elaborados y, por tanto, estar avalados por estudios de validez de constructo, por ejemplo, ya realizados. Sin embargo, la evaluación externa también tiene **inconvenientes**, como que **maximiza la reactividad**, ya que los usuarios perciben a los evaluadores como entes ajenos al programa, lo que puede modificar su conducta con respecto a circunstancias más naturales; o que es **más costosa** porque personal ajeno al programa ha de ser contratado y retribuido aparte.
- **Evaluación interna:** en este caso, los evaluadores pertenecen a la organización que ejecuta el programa, aunque no son directamente responsables de su implementación. Lo que en la evaluación externa eran puntos débiles, se convierte en la interna en **puntos fuertes**, como la **minimización de la reactividad**, ya que los evaluadores son personas conocidas por los usuarios e implementadores; o el **abaratamiento de costes**, debido a que los evaluadores forman parte de la organización que evalúa, por lo que no reciben retribución extra. Del mismo modo, lo que en la evaluación externa eran puntos fuertes, en la interna se convierte en **puntos débiles**, como el hecho de que **minimiza la credibilidad** debido a que puede haber interés en que la evaluación resulte positiva ya que es la propia organización del evaluador la que lleva a cabo la intervención; o la **minimización del uso de estándares**, con lo que hay más dificultad para obtener datos comparables con los de otros programas, la realización de los instrumentos es más costosa por ser personalizada y adaptada a la situación concreta, y no suelen poseer estudios de validez.

Respecto a estos dos tipos de evaluación ya mencionados, la relación entre puntos fuertes y débiles puede apreciarse en el cuadro 1.3 que a continuación se presenta:

	VENTAJAS	INCONVENIENTES
Evaluación interna	Minimiza la reactividad Menos costosa	Minimiza la credibilidad Minimiza el uso de estándares
Evaluación externa	Maximiza la credibilidad Maximiza el uso de estándares	Maximiza la reactividad Más costosa

Tabla 1.3. Ventajas e inconvenientes de la evaluación interna y externa (Anguera y Chacón, en prensa, a).

- **Evaluación mixta:** este tipo de evaluación combina los dos anteriores, de tal manera que los evaluadores externos realizan su trabajo en estrecho contacto y con la participación de miembros de la organización que llevan a cabo el programa. Se puede considerar la manera más conveniente de evaluar, ya que solventa los puntos débiles de las evaluaciones externa e interna por separado, a la vez que mantiene los puntos fuertes de cada una de ellas.
- **Autoevaluación:** se da cuando la evaluación es realizada por los directamente implicados en la ejecución del programa. Valorando sus acciones, determinan si se están cumpliendo las metas propuestas; por ejemplo, puede ser que un usuario que pretende dejar de fumar y esté participando en un programa para alcanzar este objetivo, cuente a diario el número de cigarrillos que se fuma; también puede consistir en que los profesionales que implementan un programa valoren su grado de competencia al desempeñar su profesión.
- **Evaluación participativa:** consiste en minimizar la posible distancia entre el evaluador y los beneficiarios. Persigue crear un buen ambiente en que todos los implicados disfruten con la intervención, hecho que aumenta la probabilidad de conseguir los objetivos. Es la tendencia que hoy día se está buscando en intervención comunitaria, pues supone la participación activa de quienes anteriormente sólo se consideraban necesitados que recibían ayuda. Con el aumento del denominado “empowerment”, el programa ayuda a aumentar la autoeficacia y la capacidad de resolver sus propios problemas, pensando que llegará un momento en que el programa se acabe y se tendrán que valer por sí mismos, sin apoyo profesional.

Sin más preámbulos, se presentan a continuación algunos modelos tradicionales en evaluación de programas.

2.1.2. Principales modelos tradicionales en evaluación de programas.

A continuación, se presenta un resumen de algunos de los modelos teóricos clásicos de mayor repercusión en la evaluación actual, de la evolución de algunos de éstos hoy día y de otros actuales. La adecuación de uno u otro modelo dependerá del objeto de estudio y del contexto de evaluación. Además, hay que tener en cuenta que, lejos de ser contrarios, en muchas ocasiones estos modelos pueden utilizarse de manera complementaria.

2.1.2.1. Modelo de planificación de la evaluación.

Los **objetivos** de este modelo creado por Cronbach (1988) son principalmente dos:

- Alcanzar un alto grado de validez, tanto interna como externa. La **validez interna** se define como la capacidad de generalizar los resultados desde la muestra a la población de la que proviene; la **validez externa** consiste en la posibilidad de generalización desde la muestra a la población diferencial.
- Extrapolar los resultados encontrados (Stufflebeam y Shrinkfield, 1987).

La **metodología** que propone para alcanzar estos objetivos es multidisciplinar, donde coexisten desde el control experimental hasta la información cualitativa.

Los pasos a seguir son básicamente dos:

1. Delimitar los siguientes elementos:

- **Unidades de evaluación (u: units):** puede hacer referencia a una persona de manera individual o a un grupo, si éste es medido como unidad de intervención.
- **Tratamientos (t: treatments):** son las distintas intervenciones en las que participan las unidades (ya sean personas o grupos).
- **Observaciones o resultados (o: outcomes):** se refiere a cómo se van a registrar los datos.
- **Contextos de actuación (s: settings):** es el marco en el que se desarrolla la evaluación.

Al combinar estos cuatro elementos, se conforma el acrónimo “utos”, que puede hacer referencia a tres niveles:

- **utos:** hace referencia a la muestra.
 - **UTOS:** se refiere a la población representativa de la muestra tomada.
 - ***UTOS:** hace referencia a cualquier población diferente a aquella de la que se extrajo la muestra del estudio.
2. Una vez delimitados los distintos elementos en los tres niveles, se pasa propiamente al estudio de validez; concretamente, la **validez interna** se considera la generalización de los “utos” a los “UTOS”, mientras que la **validez externa** hace referencia a la generalización de los “utos” a los “*UTOS” (ver figura 1.1).

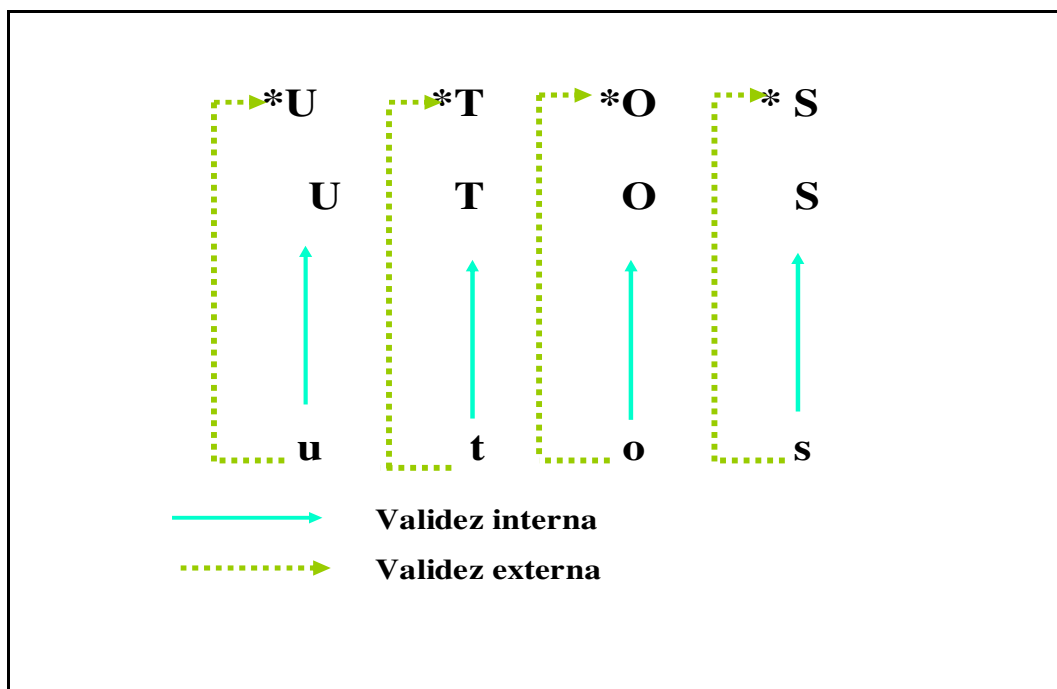


Figura 1.1. Validez interna y externa según Cronbach, adaptado de Cook, Campbell y Peracchio (1990).

Según Cronbach (1982), la **utilidad** principal de una evaluación es la mejora del programa existente.

Para ello, un aspecto a tener en cuenta es que existen dos tipos de **implicados** que pueden presentar intereses distintos: los responsables del programa (desde el nivel político al técnico) y los usuarios.

2.1.2.2. Evolución del modelo de planificación de la evaluación.

Tras varios años de evolución, Shadish, Cook y Campbell (2002a) introdujeron tanto cambios conceptuales como nuevos elementos en el modelo, de tal manera que hoy día puede quedar representado como se muestra en la figura 1.2:

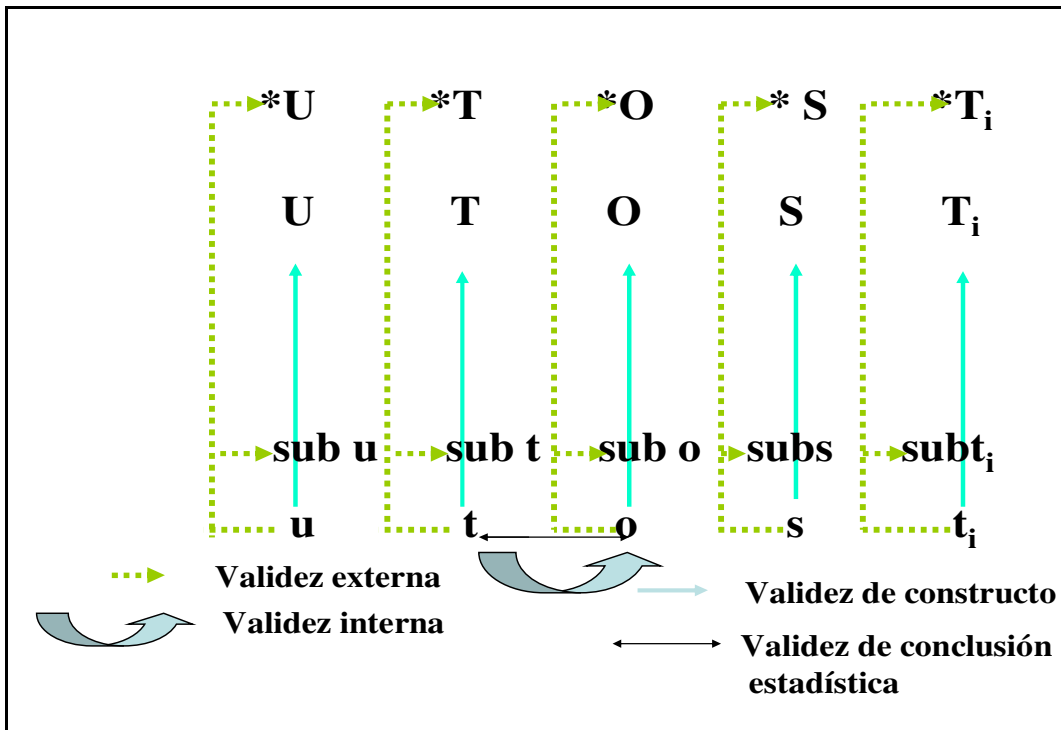


Figura 1.2. Tipos de validez según Chacón y Shadish (en prensa), adaptado de Shadish, Cook y Campbell (2002a).

Tras comparar este modelo de Shadish, Cook y Campbell con el de Cronbach, que sirvió de punto de partida, cabe destacar tres grandes cambios:

Inclusión de **un elemento más**: además de las unidades (u), los tratamientos o intervenciones (t), las observaciones o resultados (o) y el contexto (s), Shadish, Cook y Campbell consideran que es necesario incluir un elemento más, el **tiempo**, representado por **t_i (time)**. Hace referencia al momento de la recogida del dato.

Inclusión de **un nivel más**: se distingue la composición de las muestras por submuestras (representado en la figura 1.2. por “sub”). Con esto, se incluye la idea de que las comparaciones, además de hacerse entre distintas muestras de una población, pueden hacerse también entre los distintos subgrupos de una muestra.

Inclusión de **otros tipos de validez**: el **objetivo** sigue siendo el estudio de validez y la extrapolación de resultados pero, además de la interna y externa (cuyas definiciones varían), también se tiene en cuenta la de constructo y la estadística:

- **Validez de conclusión estadística**: se refiere a la validez de las inferencias realizadas acerca de la correlación entre “t” (tratamiento) y “o” (resultados); en otras palabras, sería el estudio de en qué medida y cómo el tratamiento o programa de intervención correlaciona con los resultados medidos.
- **Validez interna**: la definición de este tipo de validez dada por Shadish, Cook y Campbell (2002a) varía sustancialmente de la dada por Cronbach (cit. en Cook, Campbell y Peracchio, 1990). Se refiere al grado en que la covariación entre “t” (tratamiento) y “o” (resultados) muestran una relación causal; es decir, se estudia

hasta qué punto los resultados obtenidos son realmente debidos a las variables manipuladas en el tratamiento o no se habrían encontrado también sin esta intervención.

- **Validez de constructo:** es el grado de representatividad de la muestra ($utost_i$) en relación con la población de referencia ($UTOST_i$). Es lo que Cronbach denominaba “validez interna”.
- **Validez externa:** se refiere a la medida en que los distintos componentes de la muestra ($utost_i$) podrían generalizarse o extrapolarse a poblaciones distintas a la de referencia ($U*T*O*S*T_i*$) y a otras submuestras (representados en la figura 1.2. por “sub u”, “sub t”, “sub o”, “sub s” y “sub t_i ”). Exceptuando la inclusión de las posibles submuestras, coincide con la idea de “validez externa” de Cronbach.

Este acrónimo será utilizado para exponer los demás modelos con claridad (véase las tablas 1.4 y 1.5).

2.1.2.3. Acreditación.

Con origen a finales del siglo XIX en Estados Unidos, el **objetivo** que persigue este modelo es el reconocimiento, a modo de garantía, de la calidad del objeto de estudio.

El primer **paso** es el establecimiento de manera clara de unos estándares de rendimiento acordes con la temática a tratar; y, a continuación, un grupo de evaluadores externos valoran el servicio en función del baremo, y otorgan o rechazan la correspondiente acreditación.

La **utilidad** principal es la acreditación de los programas educativos e instituciones. De hecho, hoy día existen agencias dedicadas exclusivamente a la acreditación de productos y servicios como, por ejemplo, la Asociación Española para la normalización (AENOR) (Holgado, 2002).

2.1.2.4. Modelo dirigido a la consecución de objetivos.

Propuesto por Tyler (1942), el **objetivo** de esta evaluación es decidir si los objetivos propuestos por el programa realmente se cumplieron.

Las **fases** de evaluación son las siguientes:

- Primero, concreción clara y operativa de los objetivos que se pretenden alcanzar con la evaluación y los instrumentos que se van a utilizar para ello.
- Después, recogida de los datos.
- Finalmente, análisis para determinar la congruencia entre los objetivos propuestos y los alcanzados realmente.

La **metodología** es sistemática. Los **instrumentos** más utilizados son los tests de

rendimiento y determinadas técnicas conductistas.

Este modelo de evaluación está centrado en los **resultados**, por lo que no es útil para la mejora continua de los programas de intervención (Stufflebeam y Webster, 1988).

El **momento** de recogida de información se hace al final de la intervención (Stufflebeam y Shrinkfield, 1987).

Uno de los **contextos** en los que más ha sido utilizado este modelo es el educacional.

2.1.2.5. Modelo orientado hacia el consumidor (de meta libre).

Según Scriven (1967), el **objetivo** de la evaluación consiste en determinar hasta qué punto las necesidades que tienen los usuarios del programa son cubiertas por éste, entendiéndose como necesidad todo aquello que sea esencial para alcanzar un nivel satisfactorio de calidad de vida. Por tanto, el criterio que sirve de guía para evaluar el programa es hasta qué punto éste contribuye al bienestar del usuario.

Las **fases de evaluación** según este modelo son la recogida de información, el análisis de datos y la posterior valoración justificada. Su mayor peculiaridad es que se pretende minimizar el sesgo que se puede producir cuando los evaluadores conocen cuáles son los resultados esperados; para ello, se usa la técnica del ciego, consistente en ocultar a los evaluadores la meta que pretende alcanzar el programa a evaluar. De este modo, además, aumentan las probabilidades de encontrar objetivos obtenidos de manera inesperada, sin que estuvieran planificados en principio (lo que es denominado evaluación de la efectividad).

El **resultado** con este modelo es la búsqueda por parte de los evaluadores de juicios de valor justificados acerca del programa en cuestión.

2.1.2.6. Modelo experimentalista de evaluación.

En este modelo de evaluación propuesto por Campbell y Stanley (1963) y Suchman (1967), el **objetivo** es generalizar los programas sociales exitosos.

Las **fases** de evaluación son:

a) **Planificación del diseño** experimental.

- La **muestra** es escogida al azar de entre toda la población.
- La **asignación de las personas a las distintas condiciones del programa** también se realiza al azar.
- Se **manipulan variables** para observar los efectos que se consiguen.
- Se **recogen datos** que sirven para llegar a conclusiones acerca de qué podría modificarse para la mejora de los resultados.

A pesar de que el uso del diseño experimental tiene grandes ventajas como es el control de variables extrañas gracias a la asignación aleatoria, en el caso concreto de los programas sociales también conlleva aspectos controvertidos desde el punto de vista de la ética porque este tipo de asignación posibilita que personas que tienen problemas y realmente necesitan participar en la intervención no sean elegidos para ello, siendo así excluidos (Airasian, 1988).

- b) **Evaluación del proceso:** se estudia de qué manera se llevó a cabo el programa.
- c) **Evaluación de resultados:** responde a si se consiguieron los objetivos.

Además, para mejorar el programa y su evaluación, se estudian las posibles **amenazas a la validez** (se trata de averiguar las posibles razones por las que una inferencia puede resultar errónea).

2.1.2.7. Modelo de evaluación centrado en el cliente.

Este modelo, desarrollado por Stake (1978), tiene como **objetivo** responder a las necesidades de los clientes que encargan la evaluación. Se concibe la evaluación como una parte de un sistema más complejo y pluralista y se centra en los distintos tipos de recogida de información y análisis de datos.

En cuanto al **tipo de información** necesaria, identifica las siguientes:

- Los **antecedentes:** son los datos más relevantes acerca de la historia del objeto de evaluación.
- Las **transacciones:** son la información acerca de las actividades propias del objeto de evaluación y las interacciones entre los implicados.
- Los **resultados:** incluyen la consecución de los objetivos deseados y de los no esperados.

Respecto al **análisis de datos**, en cada uno de los tipos de información, se comparan dos dimensiones:

- Las **intenciones** son los antecedentes, las transacciones y los resultados deseados; por tanto, son los propósitos del programa.
- Las **observaciones** son los antecedentes, las transacciones y los resultados que realmente se han producido.

Con esta información se realizan dos **tipos de análisis:**

- **Análisis de la congruencia:** se trata de analizar si el programa logró los objetivos establecidos previamente.
- **Análisis de las contingencias:** se trata de poner en relación resultados

específicos con antecedentes y transacciones concretos, para averiguar a qué factores se debieron dichos resultados encontrados.

Como **resultado**, cuanto más se acerquen las observaciones a las intenciones, menos discrepancias se habrán dado entre lo esperado y lo encontrado en la realidad, con lo que mejor habrá resultado el desarrollo del programa. Finalmente, se pasa a la emisión de un **juicio de valor** desde la perspectiva de los distintos implicados, al no tener todos las mismas necesidades.

2.1.2.8. Modelo CIPP o evaluación orientada hacia la toma de decisiones.

Para Stufflebeam (1971), el **objetivo** de una evaluación es la mejora de las intervenciones.

Se parte de un diseño de **metodología** multidisciplinar en el que se considera de gran importancia la comunicación continua entre los implicados y los evaluadores. Las **categorías** del diseño, que conforman el acrónimo CIPP, son las siguientes:

- **Evaluación del contexto (C):** se trata de la delimitación del marco donde se va a desenvolver el estudio de evaluación.
- **Evaluación de entrada (I: inputs):** se trata de delimitar aquellas novedades que introduce el programa en la cotidianidad de los participantes. Se estudian los recursos disponibles para la intervención.
- **Evaluación del proceso (P):** proporciona retroalimentación continua acerca de las actividades que se están llevando a cabo, en función de las necesidades que van surgiendo. Además, se realiza un informe en el que se expone cómo se está desarrollando el programa.
- **Evaluación del producto (P):** supone la emisión de juicios de valor acerca de los objetivos obtenidos con el programa, tanto esperados como inesperados. Para ello, se toma información de todos los implicados. Además, se realiza un informe que sirve para modificar el programa, en busca de la obtención de los resultados que se pretendían.

Distintos tipos de diseño se adaptan sin problemas a este modelo, como pueden ser el experimental, cuasi-experimental u observacional.

En este modelo, pueden destacarse tres **utilidades**:

- Servir de **guía** para la toma de decisiones.
- Proporcionar **datos** en los que basarse para emitir posteriormente juicios de valor.
- Promover la comprensión de los **factores implicados**.

Este modelo, según el propio autor, está orientado hacia los **ámbitos** de la educación y los servicios sociales.

2.1.2.9. Modelo de evaluación basado en el contexto político.

Weiss (1972), propuso como **objetivo** de su modelo que los resultados de las evaluaciones fueran **útiles** en contextos caracterizados por las implicaciones políticas.

Las **fases** de evaluación son las siguientes:

- a) Búsqueda de **metas de interés político**.
- b) Traducción de estas metas a **indicadores** medibles.
- c) **Recogida y análisis** de la información.
- d) Realización del **informe de evaluación** y presentación de los resultados de manera adaptada a cada uno de los implicados.

Este modelo tiene presente la influencia del contexto político en los **resultados** de la evaluación. Se diferencia por ello entre el uso instrumental y el iluminativo:

- El **uso instrumental** de la evaluación implica que los políticos usan estos resultados como criterio único a la hora de la toma de decisiones.
- El **iluminativo o conceptual** implica que, aunque la evaluación no sirva como base fundamental de la toma de decisiones política, al menos sí se tiene en cuenta e influye a medio o largo plazo (Shadish, Cook y Levinton, 1991).

2.1.2.10. Fases de evaluación: el proceso de evaluación como un continuo.

Este modelo concibe el **objetivo** de la evaluación como la emisión de juicios de valor acerca de la valía de un programa (Anguera y Chacón, en prensa, a; Alvira, 1991; Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla, 1999).

El **proceso** de evaluación se lleva a cabo con la participación de los **implicados** para obtener información útil para cada uno de los grupos, aumentar la probabilidad de utilizar los resultados de la evaluación, facilitar la implementación de lo diseñado y potenciar la validez del proceso de evaluación.

Las **fases** de evaluación se presentan como evaluación formativa (ver figura 1.3), en contra de la lineal o dirigida a los resultados (Cascio, 1989; Goldstein, 1989). Desde este planteamiento, programación y evaluación se entienden como dos caras de la misma moneda, de tal manera que el programa, gracias a la evaluación continua a la que es sometido, es capaz de autorregularse y de aportar información útil a todos los grupos de implicados a lo largo de todo el proceso de intervención (Chacón, López y Sanduvete, 2004; Chacón, López, Campos y Sanduvete, 2003).

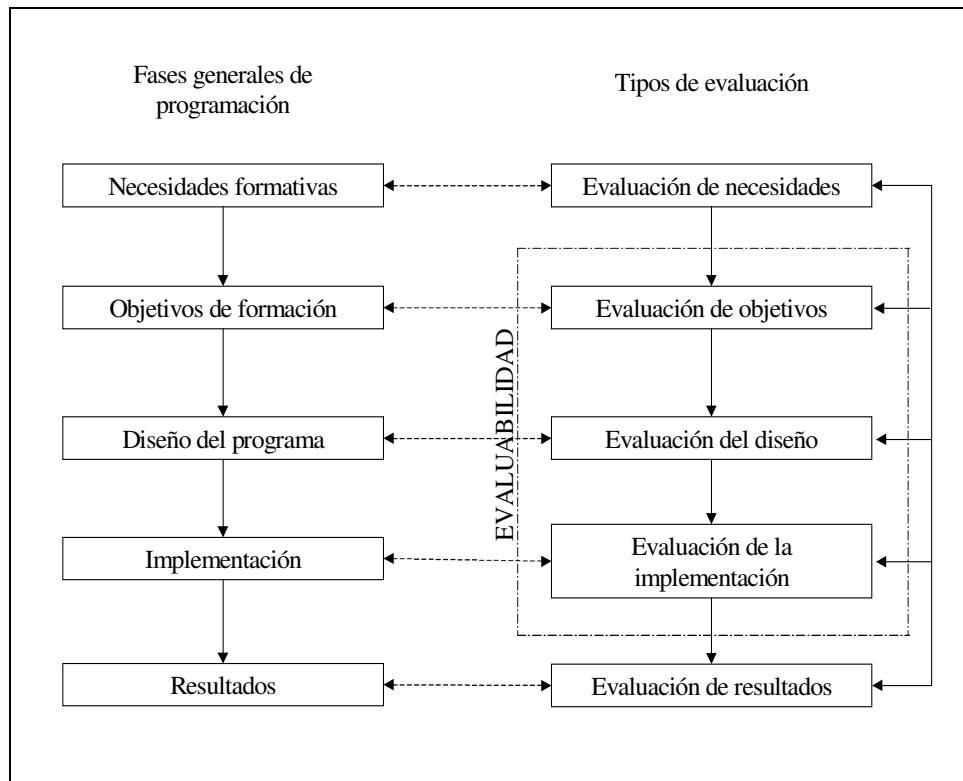


Figura 1.3. Esquema no-lineal de evaluación (Anguera y Chacón, en prensa b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999).

Tal como se muestra en esta figura, se presentan las distintas fases divididas en tres momentos: **antes** de la intervención, **durante** ésta y **al final** (Anguera y Chacón, en prensa, b Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999). El primer momento se subdivide en **detección de necesidades, objetivos y diseño**, el segundo hace referencia a la **implementación** y el tercero a los **resultados**. Para cada una de las fases, se define brevemente en qué consiste y se pasará, respecto a su evaluación, a explicitar tres puntos importantes: **delimitación de objetivos, metodología/diseño y valoración**.

- Antes de la intervención:
 1. La **detección de necesidades** es un procedimiento sistemático de identificación de necesidades, priorización y toma de decisiones de cara a la planificación de un programa para la superación de dichas necesidades (Witkin y Altschuld, 2000; Chacón, Lara y Pérez, 2002c; Hernández, Pozo y Alonso, 2004). En general, una necesidad es considerada como una discrepancia entre el estado actual de un grupo de personas y su estado deseado (Kaufman, 1988). La detección de necesidades consiste en el análisis de estas discrepancias y una posterior priorización de las acciones futuras, de tal manera que se solucionen antes las necesidades más preocupantes, básicas y urgentes.

La evaluación de esta fase se concreta en los siguientes aspectos:

- *Delimitación de objetivos*: la evaluación de necesidades busca analizar las características esenciales del problema de intervención a solucionar. En función de la fuente de información, encontramos distintos tipos de necesidad (Anguera y Chacón, en prensa, b):
 - a) **normativa**: es aquella que el experto o profesional define como necesidad en una situación determinada, basándose en unos criterios estándar o una normativa. El especialista establece un nivel “deseable” y compara con el nivel real. Si un individuo o grupo no alcanza este nivel, se considera que presenta la necesidad. Una definición normativa de la necesidad no es absoluta; quizá usando otras dimensiones, no se tendría concordancia con la necesidad normativa; también dependerá del momento temporal (por ejemplo, los derechos reconocidos a la infancia no son los mismos actualmente que en el siglo pasado).
 - b) **Experimentada, sentida o percibida**: en este caso, necesidad equivale a carencia subjetiva, dato obtenido a través de un estímulo provocador de respuesta (una pregunta o un cuestionario, por ejemplo). Al evaluar una determinada necesidad de un servicio se pregunta a los individuos si creen que lo necesitan, es decir, si lo perciben como tal necesidad. La necesidad experimentada o sentida, en sí misma, es una medida inadecuada de la necesidad real, pues es fluctuante y se ve modulada por la situación actual del individuo, tanto por defecto como por exceso.
 - c) **Necesidad expresada o demandada**: es la propia de las personas que demandan un servicio. Se diferencia de la necesidad percibida en que, en este caso, la persona que siente la necesidad la expresa por iniciativa propia, sin necesidad de que se le pregunte o se le brinde la oportunidad explícitamente.
 - d) **Necesidad comparada**: se establece a partir de la diferencia entre los servicios que existen en un área y los servicios de otra, aunque estas diferencias no tienen por qué corresponder a una necesidad ya que, a lo mejor, en el área en cuestión no se necesita dicho servicio; por otro lado, no hay seguridad de que todas las necesidades en el área de referencia estén satisfechas.
- *Metodología/diseño*: algunos instrumentos comúnmente utilizados en esta fase para la recogida de información son los **datos secundarios** (el análisis de documentos), **encuestas** a los potenciales usuarios, **entrevistas** en profundidad a implicados y expertos y **técnicas de grupo**.
- *Valoración*: Una vez recogidos y analizados los datos, se ha de llegar a conclusiones acerca de tres aspectos fundamentalmente:
 - La **pertinencia** del programa: si realmente es necesaria la implementación, si es cierto que existen necesidades que se han de

cubrir.

- La **priorización** de necesidades: supone atender en primer lugar a aquellas necesidades más urgentes en función de la conjunción de distintos criterios previamente explicitados como pueden ser la gravedad en sus consecuencias o el número de implicados.
- Los **criterios**: es necesario explicitar las razones concretas que guiarán la decisión de por qué unos potenciales beneficiarios participarán en el programa y otros no lo harán.

2. Los **objetivos** son aquellos aspectos que se quieren conseguir con la intervención.

Respecto a la evaluación de esta fase, a continuación se comentan los tres puntos fundamentales:

- *Delimitación de objetivos*: se especifica una **definición** de los objetivos a conseguir que ha de ser **concreta, medible, operativa** y con **temporalización**.
 - *Metodología/diseño*: por una parte, se requiere en este apartado una **justificación teórica**, consistente en el estudio de la literatura con el que se detecten los objetivos de programas similares y se estudien los posibles problemas que puedan surgir, sus causas y posibles soluciones a éstos. En definitiva, se pretende explicar por qué se toman esos objetivos concretos y no otros. Sería aconsejable, si fuera posible, realizar **simulaciones** para estimar unos posibles resultados antes de la implementación del programa.
 - *Valoración*: En función del contexto de intervención y de la bibliografía existente acerca de la temática, en esta fase se han de tomar decisiones sobre la **suficiencia** del programa; es decir, si éste responde a todas las necesidades priorizadas, sobre las que se ha decidido que se va a intervenir.
3. El **diseño** es el plan de trabajo que posteriormente se va a llevar a cabo. Ha de responder a los objetivos previstos y ha de ser tan específico que otra persona que no sea quien lo ha planificado pueda ponerlo en práctica. Se considera una cualidad positiva la flexibilidad del diseño pues, a la hora de implementarlo, podrían surgir algunos problemas a los que habría que responder reaccionando rápidamente y modificando lo planeado.

Respecto a la evaluación de esta fase, se han de tener en cuenta los siguientes aspectos:

- *Delimitación de objetivos*: en esta fase se ha de concretar al detalle la manera en que se va a ejecutar posteriormente el programa. Se presta especial atención a si existe **coherencia interna**, lo cual supone que no quedan cabos sueltos: que todas las necesidades priorizadas van a ser cubiertas por los objetivos previstos, que al menos una actividad va encaminada a cada uno de estos objetivos, que no queda actividad sin recursos humanos y materiales, etc.

- *Metodología/diseño:* como instrumentos para evaluar el diseño de los programas de intervención, es común el uso del **análisis de la bibliografía** existente respecto al tema y la recogida de información a través de **entrevistas** y **cuestionarios** respondidos por los implicados.
 - *Valoración:* en esta fase, tal y como se comentó en el apartado de “delimitación de objetivos” se han de tomar decisiones acerca del grado de **coherencia interna** que presenta el programa. Concretamente, es necesario explicitar las **acciones** que se pretenden llevar a cabo para conseguir los objetivos y ligar cada una de ellas a unos **recursos humanos y materiales** y una **temporalización y calendarización** concretos. Además, se concretarán cuáles serán los **usuarios** del programa, en qué **momentos se registrarán** los datos para la evaluación y cualquier **otra particularidad** de la intervención que pueda resultar de interés.
 - **Durante la intervención:**
4. En este momento, la fase que se realiza es la denominada **implementación**, consistente en la puesta en práctica del programa previamente diseñado. Respecto a su evaluación, a continuación se especifican las cuestiones más importantes:
- *Delimitación de objetivos:* el objetivo de la evaluación de esta fase es determinar en qué medida la aplicación del programa se ajusta a lo diseñado.
 - *Metodología/diseño:* los instrumentos más utilizados para la recogida de información en esta fase son las **entrevistas** y **cuestionarios** realizados a los implicados, los **registros de observación**, los **grupos de discusión** y el **sistema de registro de datos estandarizados** como, por ejemplo, los sistemas de indicadores, las fichas, los autoinformes o los historiales.
 - *Valoración:* tras evaluar esta fase, se concluye acerca del **grado de ajuste** entre lo que se diseñó en la fase anterior y lo que realmente se ha implementado. Estas posibles discrepancias y sus causas son buscadas en todos los componentes del programa: actividades desempeñadas, temporalización y calendarización de éstas, medios humanos y materiales utilizados, usuarios del programa y momentos de registro, entre otros. Cuando existen discrepancias, se pone a prueba la flexibilidad del diseño realizado porque es de suponer que, ante la necesidad forzosa de incluir alguna modificación, la planificación ha de estar preparada y por ello se sabe cómo responder ante la nueva situación.
- **Después de la intervención:**
5. Los **resultados** suponen la medición de lo que se ha obtenido con la intervención. Es la forma de determinar empíricamente si se lograron los objetivos marcados, si se alcanzaron metas no previstas y/o con qué relación coste-beneficio. Los puntos a destacar son:
- *Delimitación de objetivos:* el objetivo que se persigue en esta fase es analizar

los efectos del programa en los usuarios.

- **Metodología/diseño:** en general, se propone un diseño en el que se estudia el efecto a través de comparaciones entre un grupo control (con características similares a la población diana y que no participa en la intervención) y un grupo en el que sí se interviene. Concretamente, a continuación se presentan algunos de los posibles diseños a utilizar:

Los **diseños experimentales** requieren ciertas condiciones: a) el experimentador manipula alguna/s variable/s independiente/s; b) con esto, se pretende encontrar efecto en una variable dependiente; c) los sujetos se asignan aleatoriamente (al azar) a las distintas condiciones experimentales. Desde un planteamiento exclusivamente metodológico, se insiste generalmente en las ventajas de que se utilicen criterios completamente conocidos, preferiblemente el azar, para la asignación de las personas a los distintos grupos, con lo que se potencia una estimación insesgada del efecto.

Sin embargo, porque los grupos ya vienen a veces conformados de manera natural o por cuestiones éticas, no podemos estar seguros de que este muestreo “equiprobable” garantice la “equidad”, entendida ésta como la dotación de los recursos disponibles en función de las necesidades concretas de cada usuario, en lugar del reparto igualitario. Utilizando el azar, quizá quedaran fuera del programa algunas personas que necesitaran con mayor premura participar en el programa. ¿Se puede entonces considerar justa esta asignación? Quizá no; tal vez fuera preferible haber usado otros métodos.

Así pues, para los casos en los que no es posible o no se considera ético utilizar la aleatorización como forma de asignación de las personas a los distintos grupos, se proponen los **diseños cuasiexperimentales**, en los que también hay manipulación de variable independiente y medición de dependiente, pero la asignación de las personas a los grupos no se realiza al azar. Existen distintas alternativas para la conformación de grupos lo más similares posibles entre sí. Concretamente, éstas son algunas propuestas de actuación (Shadish, Cook y Campbell, 2002b):

- **Diseño con grupo control no equivalente y prueba previa y posterior** (ver figura 1.4). Utilizando, por ejemplo, el **emparejamiento previo**, consistiría en hacer grupos en función de las características de las personas y, una vez conformados, asignar usuarios de características similares a los distintos grupos, de tal manera que así se está seguro de que éstos no quedan descompensados, al menos, respecto a las características estudiadas. Así, si existieran diferencias entre ambos grupos en la medida de la variable dependiente después de la intervención, las características emparejadas quedarían descartadas como variables influyentes. La comparativa se realizaría, una vez conformados los dos grupos similares, tomando al menos una medida de la variable dependiente en los dos grupos en el mismo momento antes de la intervención (O_1 según la figura 1.4) y otra después de ésta (O_2). La X representa la intervención y la línea discontinua muestra que los grupos que se comparan son no equivalentes. Este diseño presenta ciertos problemas como, por ejemplo, que los cambios se deban a la maduración de

las personas por el paso del tiempo en lugar de a la intervención en sí. Este y otros problemas se solventan con el uso del diseño que a continuación se presenta.

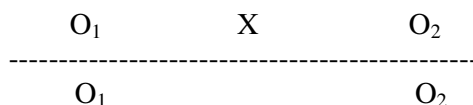


Figura 1.4. Diseño con grupo control no equivalente y prueba previa y posterior.

- **Diseño de cohortes básico** (ver figura 1.5): consiste en que, a pesar de que no son las mismas personas a lo largo del tiempo, el efecto logrado es semejante al encontrado en los estudios longitudinales (tomando datos de las mismas personas a lo largo del tiempo), pero en un periodo de tiempo más breve, pues los grupos están formados por personas que, aun teniendo características similares y viviendo circunstancias muy parecidas, difieren en la edad (denominados grupos de cohortes). En la figura 1.5, la línea que separa la observación 1 de la 2 representa el uso de grupos de cohortes.

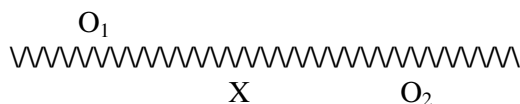


Figura 1.5. Diseño de cohortes básico.

- **Diseño de discontinuidad en la regresión:** consiste en **explicitar un criterio de inclusión** a uno u otro grupo, que es denominado **punto de corte**. Este valor concreto de la variable dependiente servirá de referencia, de tal manera que todas las personas que sobrepasen este valor comenzarán a participar en el programa. Es el diseño considerado de mayor calidad de entre los diseños cuasiexperimentales, ya que comparte una característica con los aleatorios, y es que se consigue la estimación insesgada del efecto del programa. Para que se pueda llevar a cabo este diseño, es necesario que se cumplan ciertas características, por lo que su uso queda restringido: ha de existir **relaciones lineales** entre variables y se ha de asumir que la línea de regresión **se prolongaría más allá** del punto de corte si no se produjese la intervención. Además, presenta algunos problemas: uno de ellos es la dificultad de generalizar de unas situaciones a otras por lo específico que es cada diseño; el otro es la dificultad de disponer de un punto de corte adecuado: ¿por qué tomar un valor y no otro? Probablemente, dependiendo de las características de la sociedad en la que llevemos a cabo la intervención, este punto variaría. Para llegar a un consenso, una opción interesante sería oír la opinión de diversos expertos y sólo entonces marcar el punto concreto. En la figura 1.6, se muestra la representación gráfica de este diseño, donde la C simboliza el punto de corte, las O son las distintas observaciones y la X, la implementación del programa.

C	O ₁	X	O ₂
C	O ₁		O ₂

Figura. 1.6. Diseño de discontinuidad en la regresión.

Independientemente del tipo de asignación de las personas a los distintos grupos que se lleve a cabo, es fundamental no variar la asignación posteriormente o excluir a personas que previamente fueron asignadas al grupo de intervención porque esta forma de actuar, además de suponer la pérdida de la equivalencia probabilística intergrupos, implicaría importantes amenazas a la validez del estudio, como la mortalidad diferencial o la igualación compensatoria.

- **Valoración:** como conclusión en esta fase, se trata de determinar en qué grado se cumplieron los objetivos del programa. Se pueden medir tres conceptos:

- **Eficacia:** responde a la cuestión de si se cumplieron los objetivos propuestos.
- **Efectividad:** con este concepto se estudia si, además de los objetivos previstos, se lograron algunos otros objetivos inesperados, independientemente de que sean positivos o negativos para las personas que participan en el programa.
- **Eficiencia:** es el estudio de la relación coste-beneficio. Para determinar si realmente se pudieron conseguir los mismos objetivos con menos coste o lograr mayor grado de resultados positivos al mismo precio, se hace necesario comparar entre programas similares.

- **A lo largo de todo el proceso:**

6. La **evaluabilidad** quizá no sea una fase de evaluación-intervención, sino que es un estudio que se realiza a lo largo de todo el proceso y trata de determinar si un programa realmente se puede evaluar. Para ello, ha de estar planificado sistemáticamente e implementado tal y como se ha diseñado.

Una duda que podría surgir llegados a este punto es: ¿qué hacemos si consideramos que el programa, tal como se está llevando a cabo, no se puede evaluar? Como respuesta a esta pregunta, lo primero que se ha de tener en cuenta es que cualquier intervención, por más déficit que presente, siempre será mejor que la no intervención. Por ello, si se detecta que hay mejoras que realizar, habría que informar a quienes participan en el programa (tanto usuarios como profesionales) acerca de qué se puede hacer para avanzar, siempre partiendo del trabajo que ya se está haciendo.

La máxima **utilidad** de este modelo es poder detectar inmediatamente aspectos a mejorar y tratar de solventarlos al momento.

Estas fases del proceso de evaluación son las que se van a seguir para estructurar la parte aplicada de este trabajo. Así, en el capítulo 5, se trata la **evaluación de necesidades**; en el 6, la evaluación **durante** la realización de las acciones formativas; y en los 7 y 8, la **evaluación de resultados** (satisfacción, conocimiento y transferencia).

Pero antes de pasar a estas cuestiones, a continuación se explican los modelos de evaluación más utilizados en el **ámbito concreto de la formación continua**.

2.2. Evaluación de programas de formación continua.

A continuación, se presentan algunos modelos de evaluación de la formación que suelen tomarse como referencia habitualmente. Para su exposición, se sigue el mismo esquema que en el apartado anterior referido a los modelos de evaluación de programas en cualquier contexto de intervención (ver tabla 1.5).

2.2.1. El modelo de Parker.

En el modelo de Parker (1973) se diferencian cuatro **fases**:

- **Satisfacción de los participantes:** se basa en la valoración de los participantes sobre el programa de formación en el que han participado. Entre otros aspectos, la satisfacción suele hacer referencia a los objetivos y contenidos del programa, a la metodología docente o a la utilidad.
- **Conocimientos adquiridos por los participantes:** mide el cambio en conocimientos, habilidades o actitudes que han experimentado los participantes tras su paso por el programa de formación.
- **Desempeño en el trabajo:** hace referencia al grado en que la formación ha contribuido a mejorar el rendimiento en el trabajo de los participantes en las acciones formativas.
- **Desempeño del grupo:** pretende valorar el impacto del programa de formación sobre el grupo dentro del cual el participante desempeña su función, o sobre la organización en su globalidad. Es un aspecto difícil de evaluar porque hay muchos factores además de la formación que pueden afectar al desempeño del grupo.

Según Parker (1973), la mayoría de las evaluaciones se centran en la valoración de la satisfacción y de los conocimientos. La evaluación del desempeño en el puesto de trabajo y del desempeño del grupo es mucho menos frecuente. Como **resultado**, los programas de formación suelen ser evaluados en términos de cuántas personas han sido formadas, si les ha gustado la formación recibida y qué nuevos aspectos han aprendido.

2.2.2. El modelo de “Bell System”.

El modelo de “Bell System” desarrollado por Gossman a principios de los setenta (cit. en Chacón, Holgado, López y Sanduvete, 2006), divide la evaluación de resultados en cuatro **fases**:

- **Resultados de reacción:** son las opiniones de los participantes principalmente sobre el programa de formación con respecto al contenido, metodología, materiales o actividades.
- **Resultados de capacitación:** se refieren a los conocimientos, habilidades y actitudes adquiridos por los participantes una vez finalizada la acción formativa.
- **Resultados de aplicación:** implica cuál ha sido el grado en que los participantes han aplicado en sus puestos de trabajo aquellos conocimientos o habilidades que habían recibido en la formación.
- **Resultados de valor:** su objetivo es mostrar el valor de la formación en relación con su coste. Representa el grado en el que una organización se beneficia de la formación en términos económicos, de tiempo, esfuerzos o recursos invertidos.

2.2.3. El modelo de etapas.

Dick y Carey (cit. en Brown y Gerhardt, 2002) proponen un modelo de evaluación formativa de tres **fases**: “**de uno en uno**”, en **grupos reducidos** y con **test generalizado**.

- **Evaluación “de uno en uno”:** en esta primera etapa el **objetivo** es eliminar posibles errores que aparezcan en la acción formativa y verificar que las influencias logradas con el programa son las que realmente esperaba el formador.

El **procedimiento** a seguir sería tomar de entre tres a cinco participantes y, de manera individual, hacerles preguntas acerca de si las enseñanzas transmitidas y los materiales utilizados están claros. El evaluador examinaría el impacto de lo aprendido en los participantes utilizando para ello pruebas posteriores a la intervención y valoraría los puntos positivos y negativos existentes en el contexto de aprendizaje.

Los **resultados** presentarían la emisión de un juicio acerca de la claridad, la disponibilidad y el impacto de los materiales.

- **Evaluación en grupos reducidos:** en esta segunda etapa, el **objetivo** es determinar la calidad de la acción formativa.

Los **criterios** utilizados para valorar son la cantidad de tiempo invertido y los resultados obtenidos en una prueba posterior.

El **procedimiento** consistiría en elaborar una prueba que sirviera como pretest y postest, y la realizarían un grupo de participantes de entre 8 y 20 personas. Los evaluadores medirían el tiempo empleado para la realización de la prueba y los conocimientos mostrados tras la acción formativa en comparación con el nivel del que partían.

- **Evaluación con test generalizado:** en esta última fase, el **objetivo** sería llevar a

cabo una revisión final para tratar de determinar si los cambios incluidos tras las etapas anteriores surtieron el efecto esperado.

El **procedimiento** consistiría en recoger la opinión de entre 30 y 50 participantes (incluidos tanto alumnos como ponentes) acerca de su satisfacción con la acción formativa. El evaluador determinaría si se alcanzaron los niveles de conocimiento esperados.

2.2.4. El modelo de los componentes.

Weston, McAlpine y Bordonaro (cit. en Brown y Gerhardt, 2002) proponen como **objetivo** de la evaluación determinar si se adquirieron conocimientos y si éstos se llevaron al puesto de trabajo.

Se puede **utilizar** en una amplia gama de acciones formativas puesto que se tienen en cuenta las posibles variaciones respecto a cuatro componentes: **participantes, roles, métodos y situación.**

Por ejemplo, respecto a los **usuarios**, hay que tener en cuenta su nivel de conocimiento, distinguiéndose entre noveles y expertos. Los **noveles** son definidos como aquéllos que no tienen formación o experiencia en la materia a tratar, mientras que los **expertos** sí muestran conocimiento, al menos, en un campo concreto relacionado con los contenidos del curso.

Además, respecto a los **roles o tareas**, afirman que es fundamental distinguir claramente entre el **evaluador**, quien desarrolla y conduce la evaluación formativa; el **alumnado**, que da sentido a la formación por ser quien ha de adquirir los conocimientos para la mejora del desempeño en sus puestos de trabajo; el **crítico**, quien verifica la validez del proceso de evaluación y emite los juicios valorativos; y el **revisor**, quien finalmente decide y lleva a cabo los cambios pertinentes.

Respecto al **método**, estos autores no concretan uno posible, aunque recomiendan que se explicita.

Por último, en referencia al **contexto** se comenta que, en la recogida de información, el número de participantes define la situación. Los autores consideran que cuanto menor número de personas participen en la evaluación, más detallada va a ser la información, pero menos generalizable; a más participantes, por el contrario, la información será menos detallada pero más susceptible de ser generalizada.

2.2.5. Modelo integrador de la evaluación de la formación.

Brown y Gerhardt (2002) proponen un modelo de evaluación en el que se integran las etapas y componentes de los dos modelos anteriormente planteados (el de Dick y Carey y el de Weston, McAlpine y Bordonaro, cit. en Brown y Gerhardt, 2002).

Según estos autores, ambos modelos tienen puntos fuertes, pero también poseen limitaciones. Tras integrarlos a ambos, consideran que se crea un modelo más completo.

El **objetivo** es evaluar a lo largo del tiempo diversos componentes o productos

para así tratar de mejorar las acciones formativas continuamente.

Concretamente, se obtienen **resultados** cuantitativos tras la evaluación de los siguientes aspectos:

- El **concepto**: delimita aquello que los participantes deberían aprender y está compuesto por todo aquel material que se necesita para cubrir los objetivos a alcanzar con la acción formativa. Se evalúan dos criterios:
 - **Unidad estratégica de negocios**: trata de determinar si los objetivos de la formación son acordes con los objetivos más relevantes de la organización, sus estrategias, los problemas que les surge y su desarrollo.
 - **Contexto de trabajo en grupo**: trata de evaluar en qué medida el ambiente de trabajo apoyará la aplicación de lo aprendido en el puesto de trabajo.
- El **diseño**: es un documento en el que aparecen desarrolladas las actividades a realizar para lograr los objetivos previstos explicitados en el concepto (el producto anteriormente mencionado). Dos criterios son evaluados:
 - **Tareas dirigidas hacia el objetivo**: se resuelve en qué grado las actividades a realizar en la acción formativa son acordes con los objetivos que se pretenden conseguir.
 - **Instrucciones acordes con el objetivo**: se estudia si la metodología con la que se lleva a cabo las actividades son consistentes con los objetivos que se pretenden conseguir.
- Un **prototipo**: es un clarificador ejemplo del programa final. Los criterios que se evalúan son:
 - **Utilidad**: se define como el grado en que se necesitan los diversos materiales para lograr los objetivos de la acción formativa; en otras palabras, en qué medida los materiales aportados apoyan y facilitan el aprendizaje.
 - **Satisfacción del alumnado**: consiste en evaluar el grado de aceptación que la acción formativa ha tenido entre los participantes.
- El **piloto**: es un boceto completo de todos los materiales que serán usados en la versión final de la acción formativa. Dos criterios son evaluados acerca de este producto:
 - **Viabilidad**: hace referencia a si los recursos disponibles son suficientes para alcanzar el nivel deseado de aprendizaje.
 - **Compromiso del participante**: es un criterio general que hace referencia al grado de motivación y atención presentadas por el alumnado.

La **utilidad** de este modelo radica en que de este modo se determina de manera concreta qué aspectos han de mejorarse para optimizar la consecución de los objetivos

previstos.

2.2.6. Directrices generales sobre cómo llevar a cabo la evaluación de planes de formación según Connolly.

Según Connolly (1983), el **objetivo** de la evaluación de la formación es medir la eficacia del programa.

Las **fases** por las que pasaría una evaluación son:

- Elaboración de una **propuesta de diseño de evaluación** clara y concisa que contenga las siguientes secciones: **descripción** breve del **programa**, de los **propósitos de la evaluación** y de la **utilidad** que puede darse a los resultados encontrados; **procedimiento**, en el que se describa el proceso de recogida de los datos y los instrumentos que se utilizarán (tendrían que contener cuestiones acerca de, al menos, objetivos, metodología e implementación de la acción formativa); estimación del **coste** final; y **calendarización** de las actividades concretas a realizar para llevar a cabo dicha evaluación.
- **Recogida de información** en diferentes momentos:
 - **Antes del inicio de la acción formativa:** supone realizar una detección de necesidades en la que se trata de recoger información acerca del punto de partida, de las razones por las que se solicita la evaluación, de los recursos de los que se dispone y de las limitaciones existentes.
 - **Después de finalizada la acción formativa:** consistiría en preguntar a los participantes qué aprendieron y qué conocimientos pudieron llevar a la práctica en su puesto de trabajo.
- **Análisis de la información recogida y valoración.**
- **Difusión de los resultados entre los participantes** a través de, por ejemplo, cartas.

2.2.7. Recomendaciones para evaluar planes de formación según Komras.

Aunque no es un modelo propiamente dicho, Komras (1985) aporta posibles recomendaciones para llevar a cabo la evaluación de un programa de formación.

Según esta autora, los **objetivos** de la evaluación son determinar si la formación alcanzó los objetivos previstos; recoger información de los participantes con los instrumentos en uso; determinar la validez de dichos instrumentos; estudiar la validez de cada ítem y cada ejercicio de manera individual; localizar posibles errores; y obtener recomendaciones y sugerencias de mejora de los participantes.

Por ejemplo, se aportan recomendaciones para detectar posibles errores en cuatro categorías: el lenguaje, el procedimiento de medida, el diseño y el nivel de interés:

- En ocasiones, los formadores pueden encontrarse con una audiencia cuyo nivel

de conocimientos en **lenguaje** (vocabulario, estructuras gramaticales, etc.) no es muy elevado. Para evitar que esto provoque problemas de comunicación, se recomienda a los formadores que usen vocabulario sencillo, frases cortas y que expliquen los términos técnicos que puedan ser difíciles de entender. Para estar seguros de que se está entendiendo, se sugiere prestar especial atención al feedback que los participantes aportan.

- Respecto al **instrumento de medida** de conocimientos, se comenta que se ha de evitar el uso de cuestiones excesivamente difíciles y excesivamente fáciles. Con las preguntas que prácticamente nadie respondió correctamente, se sugiere que se pregunte a los participantes dónde radicó la dificultad, para intentar cambiarlo. Se propone el uso de ítems de respuesta múltiple en lugar de aquellos que se responden escribiendo una palabra en un hueco porque este último formato lleva a más confusión.

Además, se recomienda en este sentido que los participantes sean informados del uso que se le va a dar a la información que se está recogiendo; por ejemplo, que se expliciten las ocasiones en las que se esté midiendo validez de un instrumento, para que sean conscientes de que no están siendo evaluados y evitar así estrés innecesario.

- Los problemas de **diseño** hacen referencia al material didáctico. Para que sea fácil de comprender, se sugiere una estructura clara y la presentación escueta de cada punto, en lugar de una redacción homogénea y extensa.
- La última categoría se refiere al **nivel de interés** por la evaluación mostrado por los participantes. Se recomienda que los instrumentos usados sean lustrosos, con colores, gráficas y no muy engorrosos de rellenar. Se podría hacer un estudio piloto con una muestra pequeña de personas que dieran su opinión acerca de dichos instrumentos, cambiar aquello que se considera negativo y, una vez mejorado, extrapolar su uso al grupo completo.

2.2.8. Evaluación de la Formación continua de recursos humanos.

Basándose principalmente en Campbell (del que ya se habló en el apartado de modelos de evaluación generales) y Goldstein, del que se dará más detalle en el siguiente punto, Latham (1988) realizó un estudio acerca de la evaluación de la formación a partir de un análisis histórico, tomando como puntos centrales el papel del líder y, teniendo en cuenta que su investigación se produjo principalmente en Estados Unidos, trata la generalización del método a otras culturas. Tras este estudio, propone algunas recomendaciones:

El **objetivo** de la evaluación es determinar la eficacia de la formación y tratar de delimitar aquellas variables que influyen en estos resultados para proponer mejoras para próximas ocasiones.

Con una metodología sistemática y rigurosa, las **fases** propuestas para llevar a cabo una evaluación son las siguientes:

- **Detección de necesidades** realista y detallada.

- **Planificación y diseño** de la evaluación.
- **Ejecución** del plan: evaluación del proceso y de los resultados obtenidos. Medición con instrumentos con índice adecuado de fiabilidad y validez.
- **Seguimiento** de varias medidas posteriores para estar seguros de que los cambios obtenidos se mantienen a lo largo del tiempo.
- Entrega a los interesados de un **informe final** de evaluación donde, entre otra información, se aporte el índice de fiabilidad de la medida.

2.2.9. Evaluación de la formación según Goldstein.

Según Goldstein (1989), el **objetivo** de la evaluación es aportar información suficientemente detallada como para detectar qué **aspectos concretos** habrían de cambiarse para mejorar la eficacia del programa de formación.

Las **fases** de evaluación propuestas son las siguientes:

- Un detallado y cuidadoso estudio de las **necesidades** que muestra cada organización concreta. El programa de formación es parte de la organización, por lo que su programación ha de tener en cuenta el contexto organizacional y su idiosincrasia.
- Evaluación de **resultados** a partir de índices objetivos. Siguiendo esta línea, para determinar el **cambio** producido por la formación, Arvei y Cole (1989) proponen el ANOVA o el ANCOVA como posibles análisis comparativos entre los resultados encontrados antes y después de la acción formativa, aunque no son los únicos, ya que se pueden utilizar diversos métodos más sofisticados para detectar el cambio, como los modelos lineales jerárquicos.
- Evaluación de la **utilidad** a partir de índices objetivos. Acorde con este planteamiento, Cascio (1989) define la **utilidad** como el análisis de los resultados logrados por el programa en términos económicos. Presenta la siguiente fórmula para el cálculo concreto:

$$VAN = - C_0 + \sum B_t (1/1 + i)^t > 0,$$

donde “VAN” es el valor actual neto, “C₀” es el coste del programa, “B_t” son los beneficios del programa, en el periodo de tiempo “t”, e “i” es la tasa de descuento. El valor obtenido debería dar mayor de “0”, lo que significaría que los beneficios han superado al coste.

2.2.10. Evaluación de la formación en el sector público.

Este modelo propuesto por Sims (1993) se basa principalmente en el modelo de los cuatro niveles de Kirkpatrick (1999) que se comentará en el próximo apartado.

Se plantea la evaluación como un proceso obligatorio en el sector público,

puesto que se trata de un gasto de dinero de los ciudadanos que ha de ser justificado. Los **objetivos** concretos a alcanzar con una evaluación de la formación continua son (Swierczek y Carmichael, cit. en Sims, 1993):

- Mejorar el programa de formación.
- Proporcionar feedback a los distintos interesados: los diseñadores del programa de formación, los implementadores, los participantes, etc.
- Determinar el nivel de conocimiento y habilidad de los empleados.

Las **fases** que se proponen son las siguientes:

- Evaluación **antes** de la realización de la acción formativa. Concretamente, se proponen los siguientes aspectos:
 - Evaluación de necesidades, donde se estudien las carencias de formación de los potenciales usuarios y sus expectativas, el estilo de aprendizaje y la cultura de formación (cómo se percibe la relación ponente-alumnado, en qué medida el participante se ha de esforzar, etc.).
 - Planificación: es necesario, antes de empezar la evaluación, determinar entre otras cuestiones los objetivos que se pretenden cumplir, el uso que se le va a dar prestando atención en que sea apropiado, los instrumentos que se van a utilizar, el procedimiento que se va a seguir, el horario, los materiales, los destinatarios y los recursos de los que se van a disponer (en cuestiones de financiación, tiempo y recursos humanos, entre otros).
 - Estudio del clima organizacional: anticipar si el contexto actuará favorablemente hacia la evaluación o si por el contrario la dificultará, será de gran ayuda para anticipar problemas y solventarlos con mayor soltura.
 - Elaboración de instrumentos apropiados para la recogida de información: entrevistas, cuestionarios, observación, etc.
 - Entrenamiento a los evaluadores en las principales técnicas y en el uso de los instrumentos de evaluación para tratar de evitar errores de medida como son los siguientes:
 - La **tendencia a valores centrales**: muchas personas tienden a hacer una interpretación mediocre de los hechos que observan, de tal manera que los puntúa tendiendo al centro y evitando las posiciones extremas.
 - El **efecto halo**: consiste en obviar detalles discrepantes en pro de una visión general positiva o negativa; por tanto, en ocasiones se recurre a prejuicios y falsas valoraciones.
 - Los efectos de **primacía y recencia**: consisten en dar mayor importancia a los primeros hechos que ocurren (primacía) o a los últimos (recencia), simplemente porque destacan por su posición y se recuerdan más.

- Evaluación **durante** el desarrollo de las acciones formativas. Entre otros aspectos, se sugiere centrar la atención en los siguientes:
 - Feedback. Es imprescindible recoger información del alumnado durante el curso para detectar qué contenidos no se comprenden y profundizar más en ellos.
 - Flexibilidad del diseño. Es necesario asegurarse de que la estructura del programa permite la detección de cuestiones que habrían de ser cambiadas durante el proceso y que el diseño es lo suficientemente maleable como para admitir dichas modificaciones sin cambiar con ello los objetivos a lograr.
 - Evaluación de la implementación: entre otras cuestiones, responde al grado de concordancia entre objetivos, método y ejecución. Así, una posible discordancia sería que un formador estuviera impartiendo un curso que requiriera mucha práctica con un número excesivamente elevado de alumnos para hacerlo con soltura. También se estudia si la calendarización está siendo realista y se está pudiendo seguir en la práctica.
- Evaluación de **resultados**. Concretamente se propone la evaluación a diversos niveles:
 - Evaluación del esfuerzo: responde a preguntas tales como qué se hizo y cómo se hizo.
 - Eficacia o evaluación del desempeño: si se consiguieron los objetivos previstos.
 - Eficiencia: hace referencia a la posibilidad de lograr los mismos resultados, pero utilizando menos tiempo, menos dinero y menos recursos humanos y/o materiales.
 - Transferencia: si las mejoras llegaron a reflejarse en el puesto de trabajo.
- **Toma de decisiones** en base a criterios previamente explicitados a partir de los resultados relevantes obtenidos, ignorando aquellos que sólo distraen. Por ejemplo, un error que habría que evitar en este momento de la evaluación es asumir que un solo dato referido a un elemento concreto del sistema de evaluación es suficiente para garantizar un juicio válido.
- **Difusión de los resultados** a todas las audiencias de manera comprensible.
- **Generalización** de los resultados exitosos en la medida de lo posible.

Esta evaluación centra su **utilidad** en proponer mejoras para el futuro.

En relación a los **usuarios**, es importante destacar la necesidad de conocer el perfil de los interesados en la evaluación, el número de interesados y sus intereses.

Respecto a la **metodología** propuesta, se remarca la necesidad de obtener evidencias objetivas y datos cuantitativos. Para ello, se propone la selección de evaluadores externos al programa, su entrenamiento y el diseño de instrumentos apropiados para la recogida de información. Se comenta que enriquece el uso de diversos métodos de recogida de información como la escucha, la interacción y la discusión.

En referencia al diseño de evaluación, este autor comenta que, de menor a mayor complejidad y también de menor a mayor rigor, se encontrarían el postest, el pre-post y, finalmente, el pre-post con grupo control. En definitiva, lo que se pretende al elegir el diseño de evaluación es que permita determinar con cierta seguridad que:

- Un cambio ocurrió (por ejemplo, el incremento de productividad de los empleados o el descenso de accidentes laborales).
- El cambio puede atribuirse al programa de formación.
- Se espera un cambio similar si el programa de formación fuera de nuevo implementado con otros participantes (generalización).

2.2.11. El modelo de los cuatro niveles.

El modelo de los cuatro niveles desarrollado por Kirkpatrick (1999) tiene como **objetivo** la evaluación de los resultados obtenidos en las acciones formativas. Concretamente, desarrolla un sistema de trabajo que recoge información que pretende responder a cuestiones referidas sobre: a) la satisfacción de los participantes con el programa; b) los conocimientos, habilidades o actitudes aprendidas mediante la formación; c) si los participantes han cambiado su conducta en el puesto de trabajo en función de la formación; y d) si dichos cambios han afectado positivamente a la organización.

De este modo, la evaluación se estructura en cuatro **fases** que a continuación se enumeran y seguidamente se pasarán a comentar:

- a) Nivel 1: **Reacción** de los participantes.
- b) Nivel 2: Grado de **aprendizaje**.
- c) Nivel 3: **Transferencia** al puesto de trabajo.
- d) Nivel 4: **Impacto** de la formación en la organización.

Nivel 1: **Reacción** de los participantes.

El **objetivo** de esta fase es valorar el nivel de **satisfacción** de los participantes con la formación recibida; en otras palabras, analizar si el programa les ha gustado o no.

La **utilidad** a este nivel es recabar información que permita mejorar futuras programaciones (Phillips, 1990).

Respecto a la **metodología**, en la mayoría de los casos se suelen utilizar las encuestas de satisfacción, en las que los participantes aportan la información en un tiempo relativamente corto sobre aspectos tales como los materiales empleados, los formadores, la metodología o los contenidos de la acción formativa (García, 1994). Se recomienda su anonimato para que el participante pueda responder a los ítems sin ningún miedo a recibir algún tipo de represalia (Kirkpatrick, 1977).

El problema con este tipo de evaluación es el que se refleja con la expresión “hojas de felicidad”, término con el que en ocasiones se hace referencia a los cuestionarios de satisfacción; es decir, se plantea la escasa sensibilidad de esta medida para detectar aspectos a mejorar, a no ser que se trate de elementos valorados de manera muy saliente (Thayer, 1991). Además, por ser la más fácil de llevar a cabo es la que se realiza con mayor frecuencia, prácticamente en el 100% de las ocasiones (Phillips y Phillips, 2001) y, precisamente por ello, es necesario cuestionarse la calidad métrica con la que se está realizando esta medición (Ventosa, 1998; Holgado, Chacón, Losada, Pérez y López, 2003; Pérez, Barbero, Holgado, Chacón, Ruiz, Campos, López y Suárez, 2003; Holgado, Chacón, Barbero y Sanduvete, 2006).

Otro de los aspectos negativos de este nivel de evaluación hace referencia a los **resultados** que se recogen: son subjetivos. A modo de ejemplo, puede darse la situación de que el formador que consigue que el alumnado aprenda más no es el mejor valorado por no ser quizás, a veces, tan “simpático” como otros. En este sentido, Alliger, Tannenbaum, Bennett, Trave y Shotland (cit. en Phillips y Phillips, 2001) distinguieron entre las cuestiones dirigidas a determinar si la acción formativa había resultado divertida de las dirigidas a conocer si había sido percibida como útil, concluyendo que este último aspecto correlacionaba más con la medición de otros niveles que a continuación se describirán.

Nivel 2: Grado de **aprendizaje**.

La evaluación del grado de **aprendizaje** tiene como **objetivo** proporcionar evidencias empíricas sobre la medida en que los participantes han asimilado los conocimientos, habilidades, actitudes o valores que el programa de formación pretendía (Basarab y Root, 1992).

Su **utilidad** radica en que permite detectar si el programa de formación ha estado bien planteado en relación con los objetivos didácticos que perseguía (Ventosa, 1998; Grafinger, 1998).

La evaluación del nivel dos es más compleja que la del anterior; se cree que sólo el 60% de los programas de formación realizados son evaluados a este nivel (Phillips y Phillips, 2001), aunque también hay que tener en consideración que no todas las acciones formativas pueden ser evaluadas en este sentido ya que existen algunas cuyo objetivo es meramente informar, como por ejemplo la exposición de una nueva política organizacional.

A pesar de que, en la actualidad, sigan siendo pocas las organizaciones que evalúan el grado de aprendizaje, existe una tendencia cada vez más fuerte por llevarla a cabo, ya que la certificación de conocimientos es una actividad cada vez más extendida

entre las organizaciones (Pérez, Barbero, Holgado, Chacón, Ruiz, Campos, López y Suárez, 2003; Holgado, Chacón, Ruiz, López, y Barbero, en elaboración).

Respecto a la **metodología** utilizada, cabe destacar que algunos posibles **instrumentos** de evaluación son los siguientes (Phillips, 1990; Kirkpatrick, 1977):

- **Pruebas de opción múltiple:** midiendo cuál era la situación original de los participantes antes de implementar el programa de formación y cuál es su estado tras el mismo, se pueden realizar comparaciones respecto al **conocimiento** adquirido a nivel estadístico.
- **Pruebas situacionales:** es el método de evaluación más utilizado cuando lo que se quieren detectar son habilidades y actitudes requeridas para el puesto de trabajo. En ellas, se trata de simular una situación de trabajo asimilable a la realidad para estudiar la reacción del participante.
- **Autoevaluación y evaluación en parejas:** consiste en la cumplimentación de unas hojas de evaluación en las que se responde acerca de los conocimientos que se adquirieron. Existen dos posibilidades: que uno mismo responda a su propia evaluación o que sea una segunda persona quien, a través de observación, responda acerca de los conocimientos que el primero muestra haber obtenido.

En relación al posible **diseño** a utilizar, es destacable que comparando pretest y postest, se puede estudiar si existen diferencias pero no se obtienen pruebas de que dichas diferencias se deban a la formación recibida; por ello, se recomienda el uso del diseño pre-post con grupo control. En este diseño, si los cambios son debidos realmente a la intervención, el grupo que recibió la intervención ha de mostrar diferencias al comparar ambos momentos, mientras que el grupo control, que no recibió formación, ha de obtener una diferencia al comparar ambos momentos estadísticamente menor que el grupo que sí la recibió (Kirkpatrick, 1977).

Nivel 3: **Transferencia** al puesto de trabajo.

El **objetivo** de esta fase es medir el grado en que los participantes han transferido lo aprendido durante la formación a sus puestos de trabajo. En concreto, habría que valorar cuáles de las habilidades transmitidas durante la formación están siendo usadas en el trabajo, cuáles no, el éxito de la formación en establecer estas habilidades y qué otras necesidades no formaron parte de la formación.

Esta información es **útil** para asesorar la toma de decisiones en relación a la modificación o identificación de nuevas necesidades para futuros programas de formación (Broad y Newstrom, 1992).

La complejidad de la evaluación de este nivel y su coste es mayor que en los casos anteriores ya que precisa de un detallado análisis del puesto de trabajo, lo que conlleva salir fuera del aula de formación e implicar a un rango mayor de audiencias en el proceso de evaluación tales como directivos, compañeros o clientes, por ejemplo (Amat, 1992). Así, se estima que sólo el 30% de los programas de formación son evaluados a este nivel (Phillips y Phillips, 2001).

Respecto a la **metodología**, cabe destacar que algunos **instrumentos** de evaluación de este nivel son los cuestionarios y encuestas de seguimiento, las observaciones en el puesto de trabajo, las entrevistas y grupos focales de seguimiento, las acciones y planes de mejora, las sesiones de seguimiento y la monitorización (Phillips y Phillips, 2001).

Aunque difícil y costoso, el **diseño** que se sugiere utilizar es el pretest-postest con grupo control (Kirkpatrick, 1977). Para ello, habría que seguir tres pasos:

1. Medir la conducta de los futuros participantes en la acción formativa y de un grupo de personas que no van a participar en la formación y con características similares al otro grupo. Sólo realizar este paso ya es altamente complicado por dos problemas principalmente: a) porque requiere tomar mediciones a un grupo de personas que puede que no quiera ser estudiado debido a que, a cambio, no reciben ningún tipo de beneficio; y b) porque se ha de tomar la medida con bastante antelación para que sea la conducta que las personas realmente suelen mostrar. Además, lo ideal sería tomar más de una medida previa para poder conformar una línea base y así asegurarnos de que los resultados tomados previamente a la intervención no sobrevaloran o infravaloran la conducta real.
2. Este segundo paso es aún más difícil. Consistiría en medir la misma conducta que se midió en el momento previo, pero transcurrido un tiempo considerable ya que hay que esperar a que el empleado asimile los conocimientos adquiridos y los pueda poner en práctica con naturalidad y no debido a una ilusión pasajera por haber realizado un curso no hace mucho. Lo ideal sería poder medir en varias ocasiones (al mes, a los tres meses y a los seis meses) para asegurarnos de que el cambio conductual es permanente.
3. El último paso consistiría en comparar las diferencias obtenidas en las medidas previas y posteriores del grupo que participó en la intervención, con las diferencias encontradas en ambos momentos en el grupo control. Como ya se comentó para el nivel de aprendizaje, es una manera de estar más seguros de que las diferencias obtenidas en ambos momentos se deben realmente a la formación y no a otras variables.

Nivel 4: **Impacto** de la formación en la organización.

En este cuarto nivel de evaluación, el **objetivo** es relacionar los resultados del programa de formación con la mejora de la organización. Implica obtener datos antes y después de implementar el programa y analizar la mejora respecto a ventas, productividad, coste, calidad, movimiento de personal, etc.; y contribuye, entre otras cosas, a determinar hasta qué punto se consiguen resultados similares en distintos departamentos de la misma institución y si las habilidades o conocimientos transmitidos tienen repercusiones positivas en el servicio que se ofrece (Gaines y Robinson, 1998).

En la actualidad, la **utilidad** que se le da a los resultados obtenidos en esta fase es prácticamente el análisis de la contribución de la formación para la mejora de la calidad.

Este nivel de evaluación es el menos utilizado por ser el más difícil de realizar,

de tal manera que se estima que en sólo entre el 10 y el 20% de los planes de formación se evalúa impacto (Phillips y Phillips, 2001).

En síntesis, en este modelo observamos tres tendencias (Newstrom, 1978):

- El valor de la información, en cuanto a riqueza de matices, es mayor conforme evaluamos desde el nivel uno hasta el cuatro; es decir, la evaluación del impacto es la de mayor valor para la organización o institución.
- La evaluación de la satisfacción es la más frecuente ya que, en comparación con el impacto, es más fácil de medir.
- Por el punto anterior, la dificultad en la ejecución de la evaluación aumenta conforme avanzamos de nivel.

La figura 1.7 representa bipolarmente estas dimensiones que caracterizan los cuatro niveles de evaluación descritos.

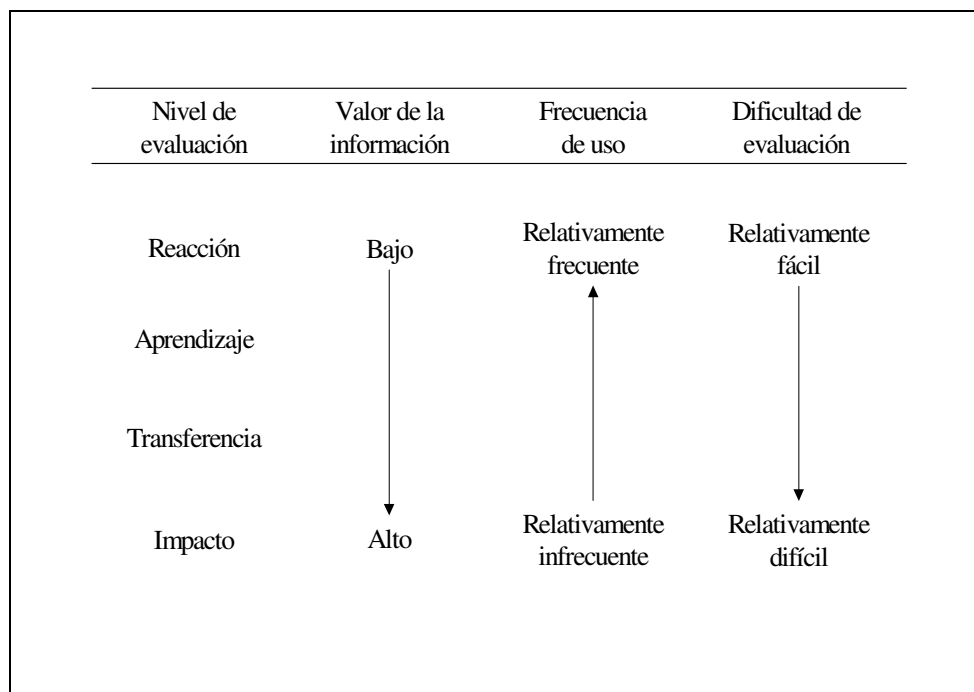


Figura 1.7. Características del modelo de los cuatro niveles de evaluación de Kirkpatrick (1999), adaptada de Phillips (1990).

2.2.12. El quinto nivel: retorno de la inversión.

Phillips y Phillips (2001), partiendo del modelo de Kirkpatrick anteriormente expuesto, añaden una fase más a la que denominan “return of investment” (ROI), cuya traducción es “**retorno de la inversión**”.

El **objetivo** es comparar los beneficios netos monetarios obtenidos gracias a la

formación con los costes generados (Cascio, 1989; Pineda, 2000).

No tiene **utilidad** a no ser que se hayan evaluado los otros cuatro niveles propuestos por Kirkpatrick. Al ser una fase más compleja de evaluar, se estima que sólo se lleva a cabo en aproximadamente el 5 ó 10% de las ocasiones (Phillips y Phillips, 2001).

La **metodología** consiste en aislar el efecto logrado por la formación de aquél conseguido por otras variables (Munoz y Munoz, 1999). El **diseño** usado para lograr esta diferenciación es el **grupo control**: el grupo de intervención participaría en la acción formativa, mientras que el grupo control no. Ambos grupos habrían de ser comparables, equivalentes, respecto a las características demográficas y a las influencias ambientales recibidas. Se recomienda, si fuera posible, la asignación aleatoria de los participantes a los distintos grupos.

Este nivel de evaluación no estaría completo hasta que los **resultados** fueran transformados a valores económicos y se compararan con los costes del programa de formación (Munoz y Munoz, 1999); es decir, habría que calcularse un índice denominado ROI ratio (Phillips y Phillips, 2001). Esta comparación mostraría la verdadera contribución de la formación. Los cinco pasos a seguir para convertir los resultados en valores monetarios serían: 1. centrar la atención en una unidad aislada; 2. determinar el valor de cada unidad; 3. calcular el cambio en rendimiento; 4. obtener una cantidad anual; 5. determinar el valor anual de mejora.

Estos autores van más allá y esbozan la existencia de un sexto nivel referido a los **beneficios intangibles** (Phillips y Phillips, 2001), donde se estudian los posibles beneficios logrados con la formación que no pueden ser convertidos a términos económicos.

3. CONCLUSIONES.

A continuación, se muestran dos tablas donde se tratan de resumir los modelos de evaluación de programas que en el apartado 2.1 de este capítulo se han desarrollado (ver tabla 1.4) y aquellos referidos concretamente a la evaluación de la formación continua, desarrollados en el apartado 2.2 (ver tabla 1.5). Para cada uno de los modelos, se especifica un nombre para su distinción, el/los autor/es que lo plantea/n, el año en que se publica una primera referencia a dicho modelo, los objetivos, las fases que los conforman (que no necesariamente muestran ordenación cronológica ya que, en ocasiones, se solapan o forman un bucle continuo), la utilidad, los usuarios a los que va dirigida, el tratamiento (entendiendo la intervención como un proceso estrechamente relacionado y en interrelación con la evaluación, ya que al evaluar a su vez se está interviniendo), los resultados, el contexto y el momento en que se realiza.

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Modelo de planificación de evaluación	Cronbach	1988	- Estudio de validez. - Extrapolación de resultados.	1º. Delimitación de las unidades de evaluación, tratamientos, resultados y contextos, a nivel muestral, poblacional y de poblaciones alternativas a la de estudio. 2º. Estudio de la validez.	En cualquier programa.	Diversos implicados: - responsables técnicos. - responsables políticos. - participantes.	Multidisciplinar: desde control experimental hasta información cualitativa.	General, cualitativo y cuantitativo.	General.
Evolución del modelo de planificación de la evaluación	Chacón y Shadish (adaptado de Shadish, Cook y Campbell)	2002	- Estudio de validez (concepto más amplio). - Extrapolación de resultados.	1º. Además de los elementos delimitados por Cronbach, se incluye el tiempo, estudiado a un nivel más: el de las submuestras. 2º. Estudio de validez (concepto más amplio).	En cualquier programa.	Diversos implicados: - responsables técnicos. - responsables políticos. - participantes.	General.	General.	General.
Acreditación	País de origen: EEUU	Final del siglo XIX	Reconocimiento, a modo de garantía, de la calidad del objeto de estudio.	1º. Establecimiento de estándares de rendimiento. 2º. Evaluadores externos valoran el servicio. 3º. Otorgamiento/rechazo de la acreditación.	Acreditación de organizaciones.	Organizaciones.	Estándares.	Datos cuantitativos.	General.
Modelo dirigido a la consecución de objetivos	Tyler	1942	Decidir acerca de si los objetivos propuestos realmente se cumplieron.	1º. Diseño/planificación de la evaluación. 2º. Recogida de datos. 3º. Congruencia entre los objetivos propuestos y los alcanzados realmente.	Evalúa resultados (no proceso)	Alumnado	- Sistemático. - Tests de rendimiento. - Técnicas conductistas.	- Evaluación de resultados.	- Ámbito educativo. - Evaluación al final de la intervención.

Tabla 1.4. Resumen de los modelos generales de evaluación presentados.

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Modelo orientado hacia el consumidor (de meta libre)	Scriven	1967	Determinar hasta qué punto las necesidades de los usuarios son cubiertas por el programa.	Sin conocer los resultados esperados: 1°. Recogida de datos. 2°. Análisis. 3°. Valoración justificada.	Especialmente adecuado para la evaluación de la efectividad.	---	Juicios de valor justificados.	Objetividad.	A lo largo de todo el proceso.
Modelo experimentalista de evaluación	Campbell y Stanley; Suchman	1963; 1967	- Evaluación de resultados y del proceso. - Cómo mejorar los programas.	1° Planificación/diseño experimental. 2° Evaluación del proceso. 3° Evaluación de resultados. 4° Estudio de las amenazas a la validez. 5° Generalización: detección de las variables comunes en los programas exitosos.	Adecuado para cuando es posible hacer un diseño experimental.	Personas con problemas sociales.	Diseño experimental.	Conclusiones con validez.	A lo largo de todo el proceso. Contexto social.
Modelo de evaluación centrado en el cliente	Stake	1978	Responder a las necesidades de los clientes que encargan la evaluación.	1°. Análisis de antecedentes, transacciones y resultados. 2°. Comparación de lo deseado y la realidad en los tres aspectos. 3°. Análisis de a qué se debieron los resultados. 4°. Emisión de juicios de valor.	Uso general.	Distintos implicados con diferentes necesidades.	Análisis de las congruencias y las contingencia.	Juicios de valor desde la perspectiva de cada implicado.	A lo largo de todo el proceso. Ámbito general.

Tabla 1.4. Resumen de los modelos generales de evaluación presentados (continuación).

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Modelo CIPP o evaluación orientada hacia la toma de decisiones	Stufflebeam	1971	Mejorar las intervenciones.	1º. Evaluación del contexto. 2º. Evaluación de entrada (inputs). 3º. Evaluación del proceso. 4º. Evaluación del producto.	- Servir de guía para la toma de decisiones. - Obtener datos para emitir juicios de valor. - Promover la comprensión de los factores implicados.	Distintos implicados.	- Diseños diversos. - Comunicación continua entre evaluadores e implicados.	Informe con posibles mejoras.	- Ámbito de la educación y los servicios sociales. - A lo largo de todo el tiempo.
Modelo de evaluación basado en el contexto político	Weiss	1972	Obtener resultados útiles teniendo en cuenta el contexto político.	1º. Búsqueda de metas de interés político. 2º. Traducción de estas metas a indicadores medibles. 3º. Recogida y análisis de la información. 4º. Realización de informe de evaluación.	Uso instrumental e iluminativo.	Diversos.	Indicadores.	Informe adaptado a cada implicado.	Contexto influido por la política.
Fases de evaluación: el proceso de evaluación como un continuo	Anguera y Chacón; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla; Alvira	En prensa; 1999; 1991	Emitir juicios de valor tras obtener por el método científico resultados cuantificables.	1º. Antes de la intervención: - Detección de necesidades. - Objetivos. - Diseño. 2º. Durante la intervención: implementación. 3º. Después de la intervención: resultados.	Detectar posibles mejoras y ejecutarlas al momento.	Participantes activos en la evaluación.	Método científico: definición de objetivo, metodología y valoración en cada fase.	Juicios de valor con base científica.	Contexto variado. A lo largo del tiempo.

Tabla 1.4. Resumen de los modelos generales de evaluación presentados (continuación).

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Modelo de Parker	Parker	1973	Determinar si se cumplieron los objetivos.	1º. Evaluación de la satisfacción de los participantes. 2º. Evaluación de los conocimientos adquiridos por los participantes. 3º. Desempeño en el trabajo. 4º. Desempeño del grupo.	Determinar si el curso gustó, si se aprendió y si se aplicó al puesto de trabajo a nivel individual y grupal.	Trabajadores.	---	Informe.	Formación Continua. Al final de la acción formativa.
Modelo de "Bell System"	Gossman	1971	Determinar si se cumplieron los objetivos.	1º. Evaluación de resultados de reacción. 2º. Evaluación de resultados de capacitación. 3º. Evaluación de resultados de aplicación. 4º. Evaluación de resultados de valor.	Determinar si el curso gustó, si se aprendió, si se aplicó al puesto de trabajo y si provocó beneficios económicos.	Trabajadores.	---	Informe.	Formación Continua. Al final de la acción formativa.
Modelo de etapas	Dick y Carey	1996	Evaluar posibles errores y la calidad de la acción formativa.	1º. Evaluación "de uno en uno": eliminar posibles errores y verificar que los resultados se corresponden con los objetivos planeados. 2º. Evaluación en grupos reducidos: determinar la calidad de la acción formativa. 3º. Evaluación con test generalizado: seguimiento.	Recoge información acerca del curso y del aprendizaje adquirido.	Trabajadores.	1. Entrevistas. 2. Contabilización del tiempo invertido. 3. Prueba de conocimiento.	Determinar si la evaluación mejoró los resultados de la acción formativa.	Ámbito de aprendizaje. Durante la acción formativa y al final.

Tabla 1.5. Resumen de los modelos de evaluación de la formación continua presentados.

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Modelo de los componentes	Weston, McAlpine y Bordonaro	1995	Evaluar prestando atención a cuatro componentes: participantes, expertos, método y situación.	1º. Delimitación de: - Perfil de los participantes. - Roles: evaluador, alumnado, revisor y crítico. - Método: técnicas e instrumentos. - Situación: contexto. 2º. Recogida de información. 3º. Toma de decisiones.	Evaluación adaptada a las características de la acción formativa.	- Número reducido: mucho detalle, pero imposibilidad de generalizar. - Número amplio: poco detallada pero fácil de generalizar.	Técnicas e instrumentos explicitadas.	Información cualitativa.	Formación continua. Al final de la acción formativa.
Modelo integrador	Brown y Gerhardt	2002	Evaluar contenidos, diseño, utilidad, viabilidad y resultados.	1º. Delimitación del contenido y diseño; prueba. 2º. Recogida de información. 3º. Toma de decisiones.	Apto para evaluación de diversos componentes a lo largo del proceso.	---	Puntuación de objetivos, contexto de trabajo, tareas, instrucciones, utilidad, satisfacción, viabilidad y compromiso del participante.	Detección de puntos débiles y propuestas de mejora.	Formación continua. A lo largo del proceso.
Directrices generales	Connolly	1983	Evaluar la eficacia a través del estudio del diseño.	1º. Diseño: Descripción del programa, de los propósitos, de la utilidad, del procedimiento, estimación del coste final y calendarización de las actividades. 2º. Ejecución. 3º. Valoración final.	Propuesta de un buen diseño, por lo que la evaluación se facilita.	---	- Elaboración de diseño y ejecución acorde a éste. - Entrevistas.	Distribución entre todos los implicados.	Formación continua. A lo largo del tiempo.

Tabla 1.5. Resumen de los modelos de evaluación de la formación continua presentados (continuación).

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Recomendaciones	Komras	1985	Evaluar la consecución de los objetivos; recoger información; determinar la validez de los instrumentos, de cada ítem y cada ejercicio; localizar errores; y obtener recomendaciones de mejora.	1°. Descripción de objetivos. 2°. Recogida de información. 3°. Estudio de validez de los instrumentos, ítems por separado y ejercicios utilizados. 4°. Evaluación de resultados: responder a si se cumplieron los objetivos previstos. 5°. Recogida de propuestas de mejora.	Evaluación de componentes concretos del proceso.	---	Recogida de información con instrumentos validados.	---	Formación continua. A lo largo del tiempo.
Evaluación de la formación continua de recursos humanos	Latham	1988	Eficacia de la acción formativa y estudio de variables influyentes.	1°. Detección de necesidades. 2°. Planificación y diseño de evaluación. 3°. Recogida de información con instrumentos fiables y validados. 4°. Seguimiento. 5°. Informe.	Recomendaciones para mejorar el proceso, lo que redundará en el resultado.	---	Instrumentos con estudio de fiabilidad y validez.	Informes.	Formación continua, recursos humanos. A lo largo de todo el proceso.
Evaluación de la formación continua	Goldstein y Cascio	1989	Detectar aspectos concretos que mejorar a lo largo de todo el proceso.	1°. Estudio de necesidades. 2°. Evaluación del proceso. 3°. Evaluación de resultados y utilidad. 4°. Inclusión de posibles mejoras a partir de los datos encontrados en la evaluación.	Análisis de resultados en términos económicos.	---	- Fórmulas de índices económicos. - Análisis de datos cuantitativos: ANOVA, ANCOVA, modelos lineales jerárquicos.	- Detección de puntos débiles. - Propuestas de mejora.	Formación continua. A lo largo de todo el proceso.

Tabla 1.5. Resumen de los modelos de evaluación de la formación continua presentados (continuación).

Nombre	Autor/es	Año	Objetivos	Fases	Utilidad	Usuarios	Tratamiento	Resultados	Contexto/ Tiempo
Evaluación de la formación en el sector público	Sims	1993	Mejorar el programa de formación, proporcionar feedback a los interesados y determinar el nivel de conocimiento y habilidad de los empleados.	<p>1°. Diseño de la evaluación: objetivos a cumplir, actividades a llevar a cabo, recursos disponibles, calendarización, criterios a partir de los cuales se tomarán las decisiones.</p> <p>2°. Recogida de información a lo largo de todo el proceso.</p> <p>3°. Análisis de los datos.</p> <p>4°. Elaboración de informes.</p> <p>5°. Difusión de la información a los participantes.</p>	Toma de decisiones a partir de datos empíricos.	Son diversos. Es necesario conocer sus necesidades e intereses.	Mediciones válidas.	<ul style="list-style-type: none"> - Análisis de datos. - Toma de decisiones. 	Formación continua. A lo largo de todo el proceso.
El modelo de los cuatro niveles	Kirkpatrick	1977	Evaluar resultados en cuatro niveles: satisfacción, aprendizaje, transferencia e impacto.	<p>1°. Evaluación de la satisfacción.</p> <p>2°. Evaluación de conocimientos, habilidades, actitudes y valores.</p> <p>3°. Evaluación de la transferencia de lo aprendido al puesto de trabajo.</p> <p>4°. Evaluación del impacto.</p>	Recabar información para mejorar futuras acciones.	Alumnado.	Encuestas, pruebas de conocimiento, entrevistas, observación.	Objetivo y subjetivo.	Formación continua. Al final.
El quinto nivel: retorno de la inversión	Phillips y Phillips	2001	Evaluar el “retorno de la inversión”.	<p>1°. Centrar la atención en una unidad aislada.</p> <p>2°. Determinar el valor de cada unidad.</p> <p>3°. Calcular el cambio en rendimiento.</p> <p>4°. Obtener una cantidad anual.</p> <p>5°. Determinar el valor anual de mejora.</p>	Adecuado para realizar análisis económico coste-beneficio.	---	Dos grupos, uno participa en la acción formativa y el control no recibe intervención.	ROI ratio.	Formación continua. Al final de la intervención.

Tabla 1.5. Resumen de los modelos de evaluación de la formación continua presentados (continuación).

A pesar de que parecen ser muchos modelos distintos y, por tanto, podrían esperarse grandes novedades entre planteamientos, realmente no son muy diferentes los unos con respecto a los otros. Así, por ejemplo, los modelos de Kirkpatrick (1977), Bell System de 1971 (cit. en Chacón, Holgado, López y Sanduvete, 2006) y Parker (1973) presentan prácticamente la misma estructura, proponiendo una evaluación de resultados en distintos niveles, la mayoría de ellos coincidentes.

Estos parecidos pueden deberse, en parte, a que muchos surgen en la misma época y, por tanto, las características contextuales permitieron un desarrollo paralelo de las teorías. Otra de las razones por las que algunos modelos pueden ser tan semejantes es porque en ocasiones surgen a partir de otros, para cubrir deficiencias; así, por ejemplo, Phillips y Phillips (2001) presentan un quinto nivel a partir de los cuatro de Kirkpatrick (1977); o Brown y Gerhardt (2002) presentan un modelo integrador a partir de los modelos de Dick y Carey (1996) y Weston, McAlpine y Bordonaro de 1995 (ambos cit. en Brown y Gerhardt, 2002).

Las conclusiones a las que se puede llegar tras este análisis de la bibliografía son las siguientes:

1. A pesar de que se han encontrado modelos que proponen una evaluación formativa como es el caso de los presentados por Anguera y Chacón (en prensa, b), Alvira (1991) y el Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios (1999); Stake (1978); o Stufflebeam (1971), generalmente la evaluación de la formación se aborda desde una **perspectiva lineal**, donde prima la medida de los resultados, como ocurre con los modelos de Kirkpatrick (1999), Phillips y Phillips (2001) o Weston, McAlpine y Bordonaro en 1995 (cit. en Brown y Gerhardt, 2002). En la evaluación de la formación exclusivamente de resultados, se podría llegar a emitir juicios de valor y tomar decisiones sustentadas en la información proporcionada por una serie de datos finales “ciegos”, sesgados, parciales, ya que no se tiene ningún referente del proceso (Goldstein, 1989). Esta primacía de la evaluación sumativa sobre la formativa puede deberse a que hoy día se busca generalmente de manera primordial la demostración del beneficio final logrado por la formación y, a nivel político, la consecución de los objetivos de manera rápida y llamativa.
2. Como referente de trabajo, se suele adoptar el **modelo de evaluación de los cuatro niveles de Kirkpatrick** (Phillips y Phillips, 2001) o alguno semejante como el propuesto por Bell System en 1971 (cit. en Chacón, Holgado, López y Sanduvete, 2006) o Parker (1973).
3. La evaluación de la **satisfacción o reacción de los participantes** es la más **frecuente**, seguida de la de conocimientos. Son escasas las organizaciones que se plantean la evaluación de la transferencia y del impacto, argumentado para ello falta de cultura y de conocimientos sobre evaluación.
4. Aunque todos los modelos explicitan los objetivos de la evaluación, en pocas ocasiones las variables en juego son definidas operativamente. Se trata de modelos mayoritariamente genéricos cuyas variables son definidas a nivel conceptual, pero con un grado de **operativización** y especificación bastante bajo, lo que dificulta su aplicación y adaptación a los diversos contextos de intervención.

5. Como **instrumentos** de evaluación destacan las **encuestas de satisfacción** con escala de valoración y los **tests de rendimiento** de elección múltiple, siendo llamativo el hecho de que, casi en la totalidad de los casos, dichos instrumentos se aplican **sin controlar ni evaluar sus propiedades métricas**.

Algunas mejoras para solventar los puntos débiles encontrados en la literatura son las que a continuación se presentan:

1. En referencia al uso de modelos de evaluación sumativa lineales se propone como alternativa la utilización de **modelos de evaluación formativa** para que dicho proceso de evaluación permita proponer una mejora continua de la calidad y de este modo se responda a las necesidades y directrices de la institución, así como a las demandas de las audiencias implicadas en la formación. Algunos de los modelos encontrados que proponían una evaluación formativa son los que a continuación se enumeran:

- En lo que a la evaluación de programas en general respecta, el modelo experimentalista de evaluación de Campbell y Stanley (1963) y Suchman (1967); el modelo de evaluación centrado en el cliente de Stake (1978); el modelo CIPP o evaluación orientada hacia la toma de decisiones de Stufflebeam (1971); o las fases de evaluación presentadas por Anguera y Chacón (en prensa, b), Alvira (1991) y el Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla (1999).
- En el ámbito concreto de la evaluación de programas de formación continua, el modelo de etapas de Dick y Carey de 1996 (cit. en Brown y Gerhardt, 2002); el modelo integrador de Brown y Gerhardt (2002); las directrices generales propuestas por Connolly (1983); las recomendaciones de Komras (1985); la evaluación de la formación continua en recursos humanos de Latham (1988); la evaluación de Goldstein (1989) y Cascio (1989); o la evaluación de la formación en el sector público propuesta por Sims (1993).

2. Respecto a la **operativización** de los modelos de evaluación, este hecho se considera algo imprescindible para el avance de la ciencia; teoría y práctica han de ser como uña y carne, totalmente complementarios y no en pugna lo uno contra lo otro porque tan necesario es partir de un modelo como tener instrumentos concretos que ligen lo abstracto con lo concreto. Por ello, se recomienda la operativización de los constructos y variables que entren en juego en la evaluación.

3. Respecto a los instrumentos de evaluación, se propone la realización de estudios de sus **índices métricos** y de las posibles **amenazas a la validez** pues es necesario recoger información para poder evaluar, pero cerciorarse de que el dato tiene cierto nivel de calidad también es fundamental. De los modelos estudiados en este capítulo, algunos que dan relevancia al estudio de las cualidades métricas se presentan a continuación:

- Como evaluación de programas en general, el modelo de acreditación; el modelo experimentalista de evaluación de Campbell y Stanley (1963); el modelo de planificación de la evaluación de Cronbach (1988); o el modelo

presentado por Shadish, Cook y Campbell (2002a).

- En el ámbito concreto de la evaluación en formación continua, las recomendaciones de Komras (1985); o la evaluación de la formación continua en recursos humanos propuesta por Latham (1988).

En este trabajo, se pretende concretar propuestas operativas y prácticas tomando como base las mejoras previamente enumeradas, para tratar de avanzar, aunque sea discretamente, hacia la optimización de la evaluación en formación continua.

Así, tal y como se describió en la introducción, tratando de operativizar las variables influyentes en la evaluación y en busca de datos de calidad, en el **capítulo 2** se presentan, en forma de escala, posibles variables concretas que pueden potenciar el grado de calidad de una acción formativa o un conjunto de éstas.

CAPÍTULO 2. ELABORACIÓN DE UNA ESCALA PARA MEDIR LA CALIDAD DE ESTUDIOS PRIMARIOS

1. Introducción	77
2. Primera fase: compilación de ítems existentes en la literatura sobre calidad en estudios primarios y aplicación exploratoria.....	79
2.1. Compilación de ítems sobre calidad en estudios primarios	79
2.1.1. Método.....	79
2.1.1.1. Muestra	79
2.1.1.2. Instrumentos	80
2.1.1.3. Procedimiento.....	80
2.1.2. Resultados.....	80
2.2. Aplicación exploratoria.....	81
2.2.1. Método.....	81
2.2.1.1. Muestra	81
2.2.1.2. Instrumentos	82
2.2.1.3. Procedimiento.....	82
2.2.2. Resultados.....	83
3. Segunda fase: estudio de validez de contenido	83
3.1. Método	84
3.1.1. Muestra.....	84
3.1.2. Instrumentos	84
3.1.3. Procedimiento	85
3.2. Resultados	85
4. Tercera fase: elaboración de una versión depurada de la escala y aplicación exploratoria	90
4.1. Elaboración de la escala (versión depurada).....	90
4.1.1. Método.....	90
4.1.1.1. Muestra	90
4.1.1.2. Instrumentos	91
4.1.1.3. Procedimiento.....	91
4.1.2. Resultados.....	91
4.2. Aplicación exploratoria.....	94
4.2.1. Método.....	94
4.2.1.1. Muestra	94
4.2.1.2. Instrumentos	94
4.2.1.3. Procedimiento.....	95
4.2.2. Resultados.....	96

5. Doble proceso deductivo-inductivo: elaboración de una versión integradora de la escala	97
5.1. Método	97
5.1.1. Muestra	97
5.1.2. Instrumentos	97
5.1.3. Procedimiento	97
5.2. Resultados	97
6. Discusión.....	114
7. Conclusiones	133

CAPÍTULO 2. ELABORACIÓN DE UNA ESCALA PARA MEDIR LA CALIDAD DE ESTUDIOS PRIMARIOS.

1. INTRODUCCIÓN.

Tal como se concluyó en el capítulo anterior puede decirse que, en general, la evaluación de la formación continua no presenta estudios de validez y fiabilidad; generalmente, consiste en la evaluación de los resultados, principalmente al nivel más inicial de los propuestos por Kirkpatrick (1999), referido al estudio de la satisfacción; y suele basarse en modelos poco operativos y metodología con un bajo nivel de estandarización. Estas conclusiones son acordes a las halladas por otros autores que han estudiado la evaluación de la formación en distintos contextos tales como, por ejemplo, Trevisan (2004) en el ámbito de la formación de adultos, Follete y Beitz (2003) en el contexto de formación en psicología clínica y Moresky, Eliades, Bhimani, Bunney y VanRooyen (2001) en la formación para trabajadores de organizaciones no gubernamentales.

Teniendo en cuenta este punto de partida, el **objetivo** en este capítulo fue determinar las variables a considerar para el estudio de la **calidad** de un programa, dando especial relevancia a las cuestiones metodológicas, pero también considerando algunas sustantivas relativas a nuestro ámbito de intervención de acuerdo con la literatura al uso.

Se elaboró una **escala** para medir la calidad de los estudios primarios que actualmente no es considerada como un instrumento cerrado, sino en fase de prueba. La utilidad de esta escala es la posibilidad de ser aplicada en un estudio específico o en un conjunto de estudios de temática común, y la obtención de datos empíricos que ayuden a detectar puntos débiles y a proponer posibles mejoras para aumentar su calidad metodológica. En el **capítulo 3** se aplicará la escala resultante a los programas de formación continua obtenidos tras una búsqueda bibliográfica. Se intenta de este modo obtener una evidencia más para corroborar o rechazar algunos aspectos encontrados en la revisión de modelos como la falta de estandarización y operacionalización en los estudios en este ámbito, intentando detectar otras posibles flaquezas y proponiendo posteriormente posibles mejoras al respecto.

El hecho de que los programas posean un diseño y evaluación de calidad metodológica es de interés pues es un modo de dar credibilidad a los resultados obtenidos ya que, en principio, podría partirse de la hipótesis de que los sesgos de los estudios disminuirán al aumentar la calidad de su diseño.

A este respecto, habría que acabar con el mito de que todo diseño experimental es de calidad y otros diseños carecen de ella porque no todo estudio tiene por objetivo establecer una relación causal y, a su vez, no todo diseño aleatorio tiene necesariamente que ser mejor que los no aleatorios para aportar evidencias empíricas de inferencias causales válidas; así, posiblemente, un diseño no aleatorio con un control de covariante fiable aporte resultados más válidos que un diseño experimental con mortalidad diferencial (Shadish y Myers, 2004; Shadish y Ragsdale, 1996), lo que a su vez lleva implícito que calidad metodológica se suele ligar a efecto causal.

La calidad del diseño de los programas es un concepto complejo y multidisciplinar. En la literatura, existen distintas aproximaciones al estudio de esta calidad; su operacionalización depende del constructo “calidad de intervención” que se tome como referente (validez interna, validez externa, calidad del informe, adecuación de análisis estadísticos realizados, implicaciones éticas, repercusiones para el área de intervención, etc.). Básicamente, existen dos planteamientos para medir la calidad sistemáticamente: estudiar un único componente de la intervención como indicador de calidad (por ejemplo, tipo de asignación de usuarios, tamaño de la muestra, etc.) u obtener un índice global de calidad a partir de las puntuaciones asignadas a una serie de ítems ponderados sobre la calidad de cada intervención particular (Chacón, 2004; Chacón, Sánchez-Meca, Sanduvete y Alarcón, en elaboración).

En este caso, se ha optado por estudiar la calidad mediante la opción del índice global por los siguientes motivos principales:

- Si se mide adecuadamente, permite el estudio de sus **índices métricos** (índices de fiabilidad y validez).
- Es el instrumento de medición **más utilizado**, por lo que resulta fácil encontrar ítems de diversas fuentes a partir de los cuales crear una nueva escala (Jüni, Altman y Egger, 2001).
- Posibilita no sólo el estudio de la calidad a partir de un **índice global cuantitativo** de fácil interpretación sino que, además, aporta información detallada de distintos aspectos a partir de los diferentes ítems. En este trabajo en concreto, interesa más esta última posibilidad, pues se pretende buscar variables (de los aspectos del diseño y la evaluación) que modulan los resultados de los programas, aunque también se calculará el índice global y se estudiará si existe alguna relación entre éste y otras variables de interés.

Una posible duda que puede surgir es que, si las escalas son los instrumentos más utilizados para medir la calidad de los estudios primarios, ¿qué necesidad hay de elaborar una más? Se intentarán aportar las siguientes novedades con respecto a lo ya existente:

- **Sistematizar indicadores útiles:** no sólo se pretende calcular un índice general de calidad; también se pretende obtener información con cada uno de los ítems acerca de cómo mejorar los programas, principalmente desde un punto de vista metodológico.
- Lograr medir con dicha escala estudios primarios con **cualquier tipo de diseño**. De este modo, se intenta cubrir la falta de atención en los diseños no experimentales que comúnmente se da en la literatura.

A continuación, se pasa a describir el proceso de elaboración de la escala. En primer lugar, se recogieron distintos ítems que posteriormente pasarían a formar parte de un cuestionario exploratorio de medición de la calidad en estudios primarios; además, se realizó un estudio exploratorio de la calidad en el diseño de programas psicológicos, sociales y de la educación con los ítems considerados relevantes; en

segundo lugar, se estudió la validez de contenido de dicho cuestionario; a partir de los resultados obtenidos, en tercer lugar se depuraron estos ítems y volvió a hacerse un estudio exploratorio, esta vez con los programas referidos a la formación continua; finalmente, en base al compendio de todos los resultados previos, se obtuvo una versión integradora de la escala, tratando de lograr un instrumento que midiera de manera homóloga la calidad de los estudios primarios independientemente del diseño que presentaran y una mayor concreción y operatividad de sus ítems. Este proceso deductivo-inductivo aún no ha llegado a su fin. El último paso hasta ahora dado ha sido el estudio exploratorio que se presentará en el próximo capítulo.

2. PRIMERA FASE: COMPILACIÓN DE ÍTEMS EXISTENTES EN LA LITERATURA SOBRE CALIDAD EN ESTUDIOS PRIMARIOS Y APLICACIÓN EXPLORATORIA.

La primera fase de elaboración de la escala para medir la calidad de los estudios primarios obtuvo como resultado la recopilación exhaustiva de todos los ítems utilizados para tal fin en la literatura. Posteriormente, se realizó una aplicación exploratoria a estudios de ámbitos social, educacional y psicológico con los ítems considerados más relevantes para ir analizando el funcionamiento de éstos y para obtener una aproximación de las características que generalmente presentaban las intervenciones en estos contextos de actuación. A continuación, se detallan el método utilizado y los resultados obtenidos en esta primera fase.

2.1. Compilación de ítems sobre calidad en estudios primarios.

2.1.1. Método.

2.1.1.1. Muestra.

Se recogieron los **artículos** referidos a la medición de la calidad de los estudios publicados, disponibles en las 11 bases de datos informatizadas a las que se tuvo acceso a través de la Universidad de Sevilla y que resultaron de interés, con la finalidad de detectar aquellos ítems que usualmente se consideraban relevantes. Concretamente, las bases de datos utilizadas fueron EBSCO Online, Medline, Serfila, CABHealth, CINAHL, Econlit, MathSci, Current Contents, Humanities Index, ERIC y PsycINFO. Otros artículos relacionados con la temática fueron obtenidos gracias a la colaboración de la Unidad de Meta-análisis de la Universidad de Murcia.

Concretamente, los **artículos** estudiados fueron los siguientes **27**: Sánchez-Meca y Ato (1989); O'Rourke y Detsky (1989); Weisz, Hawley, Pilkonis, Woody y Follette (2000); Tritchler (1999); Jüni, Altman y Egger (2001); Sánchez-Meca (1997); Sutton, Abrams, Jones, Sheldon y Sing (2000); Moher, Jones y Lepage (2001); Moher, Schulz y Altman (2001); Begg, Cho, Eastwood, Horton, Moher, Olkin, Pitkin, Rennie, Schulz, Simel y Stroup (1996); Moher, Jadad, Nichol, Penman, Tugwell y Walsh (1995); Moher, Pham, Jones, Cook, Jadad, Moher, Tugwell y Klassen (1998); Des Jarlais, Lyles, Crepaz y "the TREND group" (2004); Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Lijmer, Moher, Rennie y de Vet (2003); Bossuyt, Reitsma, Bruns,

Gatsonis, Glasziou, Irwig, Moher, Rennie, de Vet, y Lijmer (2003); Olivares, Rosa y Sánchez-Meca (2000); Education Group for Guidelines on Evaluation (1999); Campbell, Elbourne y Altman (2004); Bosch, Guardiola y grupo de trabajo Foundation Workshop 2002 (2003); Altman, Schulz, Moher, Egger, Davidoff, Elbourne, Gotzsche y Lang (2001); Brown (1991); Jüni, Witschi, Bloch y Egger (1999); McGuire, Bates, Dretzke, McGivern, Rembold, Seabold, Turpin y Levin (1985); Emerson, Burdick, Hoaglin, Mosteller y Chalmers (1990); Moher, Jadad y Tugwell (1996); MacPherson, White, Comings, Jobst, Rose y Niemzow (2002); y Greenland (1994).

2.1.1.2. Instrumentos.

Para la recopilación definitiva de dichos artículos, se utilizó el software **Procite-5** con el que, entre otras funciones, se pudo eliminar del listado de resultados encontrados en las distintas bases de datos aquellos que ya estaban incluidos, con lo que se evitaban las repeticiones.

2.1.1.3. Procedimiento.

Los pasos para la compilación de los ítems de medición de la calidad de estudios primarios fueron tres:

1º. Búsqueda de artículos: en primer lugar, se llevó a cabo una revisión bibliográfica y se recopilieron todos los artículos encontrados que hicieran referencia a la medición de la calidad de estudios primarios (el listado de artículos encontrados se presenta en el apartado referido a la muestra). Las **palabras clave** introducidas para la realización de dicha búsqueda fueron “quality” (calidad) y “primary studies” (estudios primarios).

2º. Recogida de los ítems concretos: de los artículos previamente recogidos, tomamos todos los ítems encontrados que se habían elaborado con la idea de medir la calidad. No se trató de que fueran mutuamente excluyentes sino que fueran exhaustivos por lo que, aunque algunos pueden parecer redundantes, a partir de los datos se dio la posibilidad de depurarlos a posteriori y así definir los constructos con mayor precisión. En el anexo III (pág. xix) se relaciona la procedencia de los ítems incluidos en el cuestionario exploratorio.

3º. Estructuración en dominios y subdominios: tomando como base la estructuración propuesta por Sánchez-Meca (1997) en relación a las variables moderadoras de un meta-análisis, todos los ítems fueron organizados en diferentes dominios y subdominios.

2.1.2. Resultados.

Se compilieron todos los ítems encontrados en la documentación recopilada que hacían referencia a estudios de calidad del diseño. La mayoría de estos ítems provinieron de las escalas de medición de la calidad de los estudios primarios que más frecuentemente aparecen en la literatura (Chacón, Sánchez-Meca, Sanduverte y Alarcón, en elaboración). Concretamente son: a) CONSORT (Consolidated Standards of Reporting Randomized Trials) (Moher, Schulz, y Altman, 2001); b) TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) (Des Jarlais, Lyles, Crepez y “the TREND Group”, 2004); c) STRICTA (Standards for Reporting

Interventions in Controlled Trials of Acupuncture) (MacPerson, White, Commings, Jobst, Rose y Niemzow, 2002); d) STARD (Standards for Reporting of Diagnostic Accuracy) (Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Moher, Rennie, de Vet y Lijmer, 2003); y e) Guía para evaluar los artículos referentes a la intervención en educación (Education Group for Guidelines on Evaluation, 1999). En la tabla 2.1 se presentan las principales características de cada una de estas escalas:

	CONSORT	TREND	STRICTA	STARD	EDUCACIÓN
Año de publicación	1996 (revisado en 2001)	2004	2002	2003	1999
Número de ítems	22	22	6	25	17
Adecuado para medir...	Estudios aleatorios	Estudios no aleatorios	Acupuntura	Diagnóstico médico	Intervenciones en educación

Tabla 2.1. Características de las escalas de medición de la calidad más comunes.

En definitiva, 43 ítems fueron compilados, todos aquellos que se consideraron útiles para la medida de la calidad del diseño (se eliminaron aquellos que hacían referencia exclusivamente a la calidad del reporte). Se pretendió ser exhaustivos, sin analizar el posible solapamiento parcial entre los ítems encontrados. En el anexo II (pág. vii), se muestra el listado de ítems resultante, tanto en español como en inglés; en el anexo III (pág. xix), se especifica la procedencia de cada uno de estos 43 ítems. Tal y como puede observarse, los ítems incluidos en el cuestionario de validez de contenido sobre la escala para medir la calidad de los estudios primarios fueron divididos en tres dominios (Sánchez-Meca, 1997):

- Características **extrínsecas**, referidas al contexto en el que se escribió cada estudio.
- Características **sustantivas**, referidas al contenido de cada intervención. Se distinguieron tres subdominios:
 - **Muestra**: descripción de las personas participantes.
 - **Contexto**: aquello que queda ajeno al núcleo de intervención pero que podría influir en los resultados obtenidos.
 - **Tratamiento**: la intervención en sí y el modo en que se realizó el estudio.
- Características **metodológicas**, referidas a aspectos tales como el diseño y el análisis de datos.

Con estos ítems, en la segunda fase se realizó el estudio de validez de contenido que se describe en el próximo apartado.

2.2. Aplicación exploratoria.

2.2.1. Método.

2.2.1.1. Muestra.

La búsqueda se realizó en las **bases de datos** informatizadas disponibles en la Universidad de Sevilla, de las que se extrajeron los artículos que interesaban.

Se compilaron **2087** resúmenes de los que se estudiaron finalmente **1899** tras eliminar aquellos trabajos en los que faltaban datos, cuya intervención se llevaba a cabo con sujetos no humanos y/o que eran la replicación de otro estudio previamente incluido.

2.2.1.2. Instrumentos.

El software “**Procite-5**” se utilizó para el tratamiento de la información de los resúmenes; el programa estadístico “**SPSS 12.0**” para la codificación de los datos y su posterior análisis; y, finalmente, un **sistema de categorías** compuesto por 19 ítems, elegidos de entre todos los encontrados por su utilidad a la hora de determinar la calidad de los estudios primarios (Chacón, Sánchez-Meca, Alarcón, & Marín, 2002; Chacón, García, Alarcón y Sanduvete, 2003).

2.2.1.3. Procedimiento.

Se realizó una búsqueda en todas las bases de datos a las que se tenía acceso en la Universidad de Sevilla, que guardaran relación con el tema a estudiar (se excluyeron aquellas específicas sobre algunos temas que no interesaban como, por ejemplo, de economía). Se recogieron artículos publicados hasta julio de 2005.

Las **palabras clave** utilizadas fueron random (aleatorio), non-random (no aleatorio), effect size (tamaño de efecto), quasi-experimental (cuasi-experimental), experimental, meta-analysis (meta-análisis), intervention program (programa de intervención), evaluation (evaluación), assessment (detección), social y education (educación). Se introdujeron en inglés porque las bases de datos utilizadas estaban en este idioma. Las razones por las que se tomaron estas palabras clave y no otras fueron las siguientes:

- Por un lado, se pretendía encontrar intervenciones que presentaran cualquier tipo de diseño:
 - Diseños experimentales, caracterizados porque, además de que el experimentador manipula alguna/s variable/s independiente/s y pretende encontrar efecto en una variable dependiente, los sujetos se asignan aleatoriamente a las distintas condiciones experimentales. Para encontrar este tipo de diseño, se usaron las palabras “aleatorio” y “experimental”.
 - Diseños cuasi-experimentales, que se diferencian de los anteriores principalmente en que los sujetos no son asignados aleatoriamente a las distintas condiciones del tratamiento. Es por ello que se tomaron como palabras clave “no aleatorio” y “quasi-experimental”.
- Se intentaban detectar intervenciones planificadas, sistematizadas, temporalizadas y con recursos humanos y materiales especificados; es decir, que mostraran cierto control; que siguieran aquellas pautas de carácter científico que permiten la posibilidad de que estos programas sean evaluados de manera fidedigna; en

definitiva, se intentaban detectar intervenciones que presentaran “evaluabilidad”. Para encontrar intervenciones con estas características, se utilizaron las palabras “programa de intervención”, “evaluación” y “detección”.

- Se intentaba encontrar estudios en los que se explicitaran los resultados encontrados con la suficiente precisión como para que, si ello fuera de interés, se pudieran incluir en un estudio meta-analítico. Por ello, se tomaron las palabras clave “tamaño de efecto” y “meta-análisis”.
- Finalmente, se pretendía encontrar intervenciones en distintos contextos; en diferentes ámbitos. Es por ello que se incluyeron, entre las palabras clave utilizadas, “social” (con la intención de encontrar programas en el contexto comunitario) y “educacional” (en busca de intervenciones en el ámbito educativo).

De los ítems recopilados respecto a medición de la calidad en estudios primarios, se eligieron aquellos considerados de mayor interés en función de su utilidad y se incluyeron como sistema de categorías para codificar los estudios encontrados.

Tres codificadores independientes categorizaron dichos estudios, obteniéndose una fiabilidad intercodificadores adecuada, con un índice de correlación intraclase de 0.85.

2.2.2. Resultados.

Respecto al **análisis bibliográfico** exploratorio referido a la calidad que suelen presentar los estudios primarios en los ámbitos psicológico, social y educacional, los resultados se muestran en el anexo IV (pág. xxiii) (Chacón, García, Alarcón y Sanduvete, 2003).

A grandes rasgos, puede decirse que los estudios mostraron dos características:

- Un **bajo grado de especificación**. Así, la orientación teórica no fue explicitada en el 67.2% de las ocasiones; la edad no fue especificada en el 82.2% de los casos; y no solió presentarse tamaño de efecto ni datos suficientes para calcularlo.
- Un **grado medio de control y estandarización**: en el 63.4% de las ocasiones, no hubo aleatorización en la asignación de las personas a los grupos, pero sí se dio control de alguna variable extraña; la mayoría de los estudios presentaron un diseño pre-experimental o cuasiexperimental; en la mayoría de las ocasiones, sólo se daba medida posterior; en la mayoría de los casos, sólo algunas variables eran medidas en todos los momentos; al menos una variable dependiente fue semiestándar en el 71.9% de los casos; y se dieron usualmente técnicas de control, pero no el ciego.

Tras este análisis exploratorio, se pasó a tratar de determinar con más precisión y certeza los aspectos que, según los expertos en la materia, influyen en el nivel de calidad metodológica de un estudio.

3. SEGUNDA FASE: ESTUDIO DE VALIDEZ DE CONTENIDO.

Todos los ítems recogidos en la primera fase, una vez estructurados en las distintas dimensiones, sirvieron como cuestionario exploratorio del que se realizó un estudio de validez de contenido para tratar de determinar, según la opinión de expertos en la materia, qué ítems eran viables, útiles para medir calidad y representativos de su dimensión. A continuación, se describe el método utilizado y los resultados obtenidos en dicho estudio de validez de contenido.

3.1. Método.

3.1.1. Muestra.

La muestra para la realización del estudio de validez de contenido de la escala elaborada estuvo compuesta por **30 expertos** en cuestiones de diseño y evaluación de programas y de medición de la calidad (miembros del “Methods Group de la Campbell Collaboration” y miembros de la “Asociación Española de Metodología de las Ciencias del Comportamiento”). Concretamente, fueron 12 mujeres y 18 hombres, de procedencia europea y norteamericana (el 50% de los participantes fueron españoles por cuestiones de accesibilidad). La media de edad fue de 47 años y 14 la media de años de experiencia en estas cuestiones.

3.1.2. Instrumentos.

Se utilizó un **cuestionario** que contenía los 43 ítems recopilados en la fase anterior (ver anexo II, pág vii). Se utilizó una escala de 3 puntos (siendo -1 la puntuación más baja; 0 la intermedia y 1, la más alta) para medir, en cada uno de los ítems, los tres constructos más frecuentemente utilizados en los estudios de validez de contenido (Chacón, Pérez, Holgado y Lara, 2001a):

1. **Representatividad**, que hacía referencia al grado en que cada ítem representaba el dominio al que había sido asignado.
2. **Utilidad**, referido a la medida en que cada ítem específico era útil para evaluar la calidad de los estudios con respecto al dominio donde fue asignado.
3. **Viabilidad del dato**, referido a si las circunstancias posibilitaban la existencia del dato y su recogida.

Los ítems fueron diferenciados en tres dimensiones; concretamente, 6 formaron parte de las características **extrínsecas**, 14 de las características **sustantivas** (5 respecto a la muestra, 3 en relación al contexto y 6 en el tratamiento) y 23 de las características metodológicas. En cada dominio y subdominio, además, se dejó un espacio por si algún participante quería especificar alguna posibilidad no incluida.

Otro instrumento utilizado fue **internet** para la distribución de los cuestionarios y su posterior recogida.

Por último, se utilizó el software **Microsoft Excel** para codificar y analizar los datos.

3.1.3. Procedimiento.

Se llevó a cabo un estudio de validez de contenido para determinar qué ítems eran considerados los más aptos para medir cada uno de los dominios referidos a la calidad. El procedimiento sistemático puede resumirse en tres pasos:

1. **Selección de la muestra:** Se obtuvo la colaboración de 30 personas expertas en Meta-análisis, evaluación de la calidad y diseño (una descripción más detallada se presentó en el apartado 3.1.1 referido a la muestra).
2. **Distribución del instrumento:** el cuestionario fue repartido a cada experto a través de correo electrónico o en persona, en diversos encuentros de divulgación científica, en tres ocasiones (Ritter y Sue, 2007); concretamente, la primera solicitud se realizó en el *V Annual Campbell Collaboration Colloquium* celebrado en Lisboa en febrero de 2005; la segunda, en el *IX Congreso de Metodología de las Ciencias Sociales y de la Salud*, celebrado en Granada, en septiembre de 2005; y la tercera, a través de correo electrónico a quienes aún no habían respondido.
3. **Análisis de datos:** una vez recogidas las respuestas, se aplicó el índice de Osterlind (1998) con la finalidad de determinar el grado de acuerdo entre expertos respecto a las puntuaciones otorgadas a cada uno de los ítems dentro de su dominio y respecto a los tres constructos: representatividad, utilidad y viabilidad del dato. Concretamente, la fórmula del índice de Osterlind es la siguiente:

$$I_{ik} = \frac{(N-1) \sum_{j=1}^n X_{ijk} + N \sum_{j=1}^n X_{ijk} - \sum_{j=1}^n X_{ijk}}{2(N-1)n}$$

Siendo:

- N = número de dominios.
- X_{ijk} = el valor que cada juez otorga a cada ítem.
- n = número de jueces.

Este índice podía oscilar entre -1 y 1, ya que la escala utilizada fue de tres puntos (-1, 0 y 1), siguiendo las recomendaciones del autor. Obtener el valor extremo negativo en un ítem (-1), se interpretaría como que todos los expertos puntuaron a ese ítem con el valor más negativo; del mismo modo, el valor extremo positivo (1), implicaría que todos los expertos estuvieron de acuerdo en que el ítem descrito merecía en el aspecto estudiado la puntuación máxima. Siguiendo también las recomendaciones del autor, se consideraron valores aceptables aquellos que se encontraron por encima del 0.5 (Osterlind, 1998).

3.2. Resultados.

A continuación, en la tabla 2.2 se presenta cada ítem del cuestionario con sus respectivas puntuaciones obtenidas en función del índice de Osterlind (1998) en los tres aspectos estudiados: **representatividad (columna R)**, **utilidad (U)** y **viabilidad del dato (V)**. En negrita se marcan los valores iguales o superiores a 0.5; con negrita y

cursiva se destacan aquellos ítems que obtuvieron una puntuación mayor a ésta en los tres aspectos.

CARACTERÍSTICAS EXTRÍNSECAS (N=30)	R	U	V
1. Tipo de publicación	0.429	0.571	0.714
2. Año de publicación	-0.071	0.214	0.857
3. Índice de impacto de la revista	-0.143	0.071	0.286
4. Base de datos en que se encontró	-0.214	0.357	0.357
5. Entrenamiento de los investigadores	0.071	0.5	0
6. Estructura recomendada por la APA	-0.143	0	0.143
CARACTERÍSTICAS SUSTANTIVAS (N=30)			
De la muestra			
7. <i>Rango de edad</i>	0.571	0.5	0.571
8. <i>Media de edad</i>	0.786	0.786	0.714
9. Desviación típica de la edad	0.429	0.143	0.429
10. Origen cultural	0.143	0.214	0.286
11. Nivel socioeconómico	-0.071	0.071	-0.286
Del contexto			
12. Contexto de intervención	-0.214	0.071	0
13. Campo de intervención	0.5	0.357	0.857
14. País	0.357	0.429	0.714
Del tratamiento			
15. Orientación teórica	0.286	0.8	0
16. Evidencia empírica previa	0.143	0.286	0.071
17. <i>Periodo de tratamiento</i>	0.786	0.929	0.643
18. <i>Grado de intensidad del tratamiento</i>	0.786	0.929	0.786
19. <i>Unidades (en grupo o individual)</i>	1	0.929	0.929
20. Los puntos fuertes y débiles son discutidos	0.429	0.143	0
CARACTERÍSTICAS METODOLÓGICAS (N=30)			
21. <i>Criterios de inclusión y exclusión de las unidades de la muestra especificados</i>	0.643	0.929	0.5
22. <i>Asignación aleatoria de las unidades a los grupos</i>	0.929	1	0.643
23. <i>Tipo de metodología/ diseño</i>	0.857	0.929	0.643
24. <i>Tamaño de la muestra</i>	0.786	0.857	1
25. Estadístico para calcular el tamaño de la muestra	0.429	0.5	0.286
26. Mortalidad experimental	0.714	0.857	0.143
27. Sin mortalidad	0.571	0.5	0.429
28. Mortalidad experimental entre grupos	0.714	0.857	0.071
29. Exclusiones posteriores a la asignación aleatoria	0.643	0.643	0.214
30. Periodo de línea base	0.071	0.214	0
31. Periodo de seguimiento	0.571	0.714	0.286
32. <i>Momentos de medida</i>	0.929	0.929	1
33. Las medidas del pre-test aparecen en el pos-test	0.786	0.857	0.429
34. Variables dependientes normalizadas	0.571	0.643	0.357
35. Homogeneidad de la intervención	0.571	0.357	-0.143
36. Técnicas de control	0.714	0.857	0.214
37. Definición del constructo	0.857	0.714	-0.071
38. Métodos estadísticos para inferir los valores perdidos	0.643	0.571	0.214
39. Especificación de los intervalos de confianza en los análisis estadísticos	0.143	0.214	0.5
40. <i>Tamaño de efecto y valor</i>	0.714	0.786	0.571
41. Otros datos aparte de los objetivos marcados	0.143	0.214	0.357
42. Interpretación de los resultados	0.143	0.071	0.214
43. Interpretación de los sesgos de los resultados	0.429	0.214	0.071

Tabla 2.2. Índices de representatividad, utilidad y viabilidad del dato obtenidos en cada uno de los ítems de la escala de calidad depurada.

Como puede verse en la tabla 2.2, los ítems que obtuvieron valoraciones consideradas positivas (por encima de 0.5) en los tres aspectos medidos son los siguientes:

- Respecto a las **características sustantivas**, concretamente referidas a las de la **muestra** (de las personas participantes en la intervención), dos ítems fueron considerados representativos de la dimensión, útiles y viables:
 - El **rango de edad** (ítem 7, con valores de 0.57, 0.5 y 0.57 en representatividad, utilidad y viabilidad respectivamente).
 - La **media de edad** (ítem 8), con valores aún más positivos; concretamente, 0.79, 0.79 y 0.71 respectivamente.
- Aún haciendo referencia a las **características sustantivas**, pero esta vez centrándonos en el **tratamiento**, los ítems con valoraciones consideradas positivas en los tres aspectos estudiados fueron:
 - El **periodo de tratamiento** (ítem 17), con unas valoraciones de 0.79, 0.93 y 0.64 en representatividad, utilidad y viabilidad del dato respectivamente.
 - El **grado de intensidad del tratamiento** (ítem 18), con valores concretos de 0.79, 0.93 y 0.79.
 - La **unidad** (ítem 19), con valoraciones muy positivas; concretamente 1 en representatividad y 0.93 tanto en utilidad como en viabilidad del dato.
- Los ítems que a partir de ahora se muestran hacen referencia a **características metodológicas**:
 - **Criterios de inclusión y exclusión de las unidades de la muestra especificados** (ítem 21). Sus valores obtenidos fueron 0.64 en representatividad, 0.92 en utilidad y 0.5 en viabilidad del dato. Por tanto, aunque los valores de los tres aspectos estudiados coincidieron o estuvieron por encima del valor considerado crítico (0.5), el que destacó fue el de utilidad.
 - **Asignación aleatoria de las unidades a los grupos** (ítem 22). Este ítem obtuvo valores muy positivos en representatividad de su dimensión y utilidad (0.93 y 1 respectivamente) y un valor no tan positivo, aunque sí por encima de 0.5 en viabilidad del dato (0.64, concretamente).
 - **Tipo de metodología/ diseño** (ítem 23). Siguiendo la línea del ítem anterior, éste obtuvo valoraciones positivas en los tres aspectos recogidos, pero especialmente en los referidos a representatividad y utilidad; concretamente, obtuvo 0.86 en representatividad, 0.93 en utilidad y 0.64 en viabilidad del dato.

- **Tamaño de la muestra** (ítem 24). Obtuvo altas puntuaciones en representatividad y utilidad (0.79 y 0.86 respectivamente). En viabilidad del dato, obtuvo la máxima puntuación (1).
 - **Momentos de medida** (ítem 32). Este ítem obtuvo valoraciones muy positivas en los tres aspectos estudiados (0.93 en representatividad y utilidad y 1 en viabilidad del dato).
 - **Tamaño de efecto y valor** (ítem 40). Este ítem obtuvo 0.71 en representatividad, 0.79 en utilidad y 0.57 en viabilidad del dato.
- Numerosos ítems obtuvieron valoraciones consideradas positivas en dos de los tres aspectos estudiados. A continuación, se presentan aquéllos que obtuvieron valores por encima de 0.5 en representatividad y utilidad, pero no en viabilidad del dato, todos ellos dentro del dominio de **características metodológicas**:
 - **Mortalidad experimental** (ítem 26). Este ítem obtuvo unas valoraciones muy positivas en representatividad y utilidad (0.71 y 0.86 respectivamente) y, sin embargo, obtuvo una puntuación muy baja en viabilidad del dato (0.14).
 - **Sin mortalidad** (ítem 27). Este ítem obtuvo índices superiores a 0.5, aunque no muy elevados; concretamente, recibió 0.57 en representatividad y 0.5 en utilidad.
 - **Mortalidad experimental entre grupos** (ítem 28). Sus índices, bastante elevados tanto en representatividad como en utilidad, fueron 0.71 y 0.86.
 - **Exclusiones posteriores a la asignación aleatoria** (ítem 29). Los índices obtenidos fueron 0.64 tanto en representatividad como en utilidad.
 - **Periodo de seguimiento** (ítem 31). Este ítem obtuvo índices considerados como positivos tanto en representatividad como en utilidad, siendo el valor en éste último aspecto algo más alto (concretamente, obtuvo 0.57 y 0.71 respectivamente).
 - **Las medidas del pre-test aparecen en el post-test** (ítem 33). Obtuvo índices muy positivos tanto en representatividad como en utilidad (concretamente, 0.79 y 0.86 respectivamente).
 - **Variables dependientes normalizadas** (ítem 34). Los expertos puntuaron este aspecto como representativo de su dimensión y útil, aunque no obtuvo índices muy elevados (0.57 y 0.64 respectivamente).
 - **Técnicas de control** (ítem 36). Este aspecto obtuvo índices bastante positivos tanto en representatividad (0.71) como en utilidad (0.86).
 - **Definición del constructo** (ítem 37). Este ítem fue considerado por los expertos como sumamente representativo (0.86) y muy útil (0.71). Llama la

atención el hecho de que su viabilidad fue considerada tan baja que el índice alcanzó un valor negativo (concretamente, -0.07).

- **Métodos estadísticos para inferir los valores perdidos** (ítem 38). Aun no obteniendo valoraciones muy elevadas, este ítem fue considerado tanto representativo (obteniendo un índice de 0.64) como útil (0.57).
- Un solo ítem, concretamente el número 1 dentro del dominio de **características extrínsecas**, referido al **tipo de publicación**, obtuvo valoraciones por encima del 0.5 en utilidad y viabilidad del dato (aproximadamente 0.57 y 0.71 respectivamente) y por debajo de este valor, aunque muy cercano a él, en representatividad (aproximadamente, 0.43).
- Por último, también un solo ítem, concretamente el referido al **campo de intervención** (ítem 13, dentro del subdominio de las **características sustantivas del contexto**) obtuvo índices con valoraciones iguales o por encima de 0.5 en representatividad (0.5) y viabilidad del dato (0.86), pero no en utilidad (0.36).

La información sobre estos resultados hasta ahora descrita aparece esquematizada de manera general en el anexo V (pág. xxvii), donde se muestran los ítems que obtuvieron índices iguales o mayores a 0.5 combinando de manera distinta los aspectos valorados; de esta forma, la línea A se refiere a aquellos que obtuvieron estas valoraciones en representatividad, utilidad y viabilidad del dato (R+U+V); la B a los que obtuvieron índices con estos valores en representatividad y utilidad, pero no en viabilidad (R+U); la C a los que los obtuvieron en utilidad y viabilidad (U+V); y la D a los que los obtuvieron en representatividad y viabilidad (R+V).

La tabla 2.3 que a continuación se presenta muestra el porcentaje de ítems que obtuvieron índices considerados adecuados respecto a los aspectos valorados, tanto de manera global (columna de “escala total”), como en función de los dominios/dimensiones a los que fueron asignados (características extrínsecas, sustantivas y metodológicas). La primera columna explicita las distintas combinaciones de los aspectos valorados: representatividad (R), utilidad (U) y/o viabilidad del dato (V). Entre paréntesis se muestra la frecuencia (el número de ítems) correspondiente a cada porcentaje.

ASPECTOS	CARACT. EXTRÍNSECAS (6)	CARACT. SUSTANTIVAS (14)	CARACT. METODOLOGICAS (23)	ESCALA TOTAL (43)
R, U, V	0% (0)	35.71% (5)	26.9% (6)	25.58% (11)
R, U	0% (0)	35.71% (5)	69.59% (16)	48.84% (21)
U, V	16.67% (1)	35.71% (5)	26.09% (6)	27.91% (12)
R, V	0% (0)	42.86% (6)	26.09% (6)	27.91% (12)

Tabla 2.3. Porcentaje de ítems con índices iguales o superiores a 0.5.

Como puede observarse, los ítems incluidos en la dimensión “**características extrínsecas**” obtuvieron índices bastante bajos, ya que sólo uno superó el 0.5 en utilidad y viabilidad. Con esto se puede concluir que esta dimensión no fue valorada como importante para determinar la calidad del diseño. No existen diferencias en los datos independientemente de que se tengan en cuenta los tres aspectos (representatividad, utilidad y viabilidad) o sólo dos.

Aunque tampoco existen diferencias importantes en la distribución de frecuencia de ítems con índices mayores o iguales que 0.5 (en torno al 36%) entre los aspectos valorados en las **características sustantivas**, sí se podría decir que, al menos, fueron considerados, en su conjunto, más importantes para determinar la calidad de un diseño que las características extrínsecas.

Respecto a las **características metodológicas**, cabe destacar que diez ítems (lo que supone más del 25% del total y casi el 50% de los ítems referidos a este dominio) obtuvieron valoraciones positivas en representatividad y utilidad, pero no en viabilidad; concretamente, obtuvieron índices positivos el 69.59% de los ítems incluidos en esta dimensión. Si, sin embargo, se tiene en cuenta el aspecto de viabilidad del dato, el porcentaje de ítems con índices positivos desciende a un 26.09%. Esto nos lleva a deducir que, a pesar de que los aspectos metodológicos son considerados importantes para determinar la calidad del diseño de las intervenciones, no se suelen facilitar.

Esto es corroborado atendiendo a la última columna, donde se estudia el porcentaje de ítems con índices óptimos sin diferenciar en función del dominio en el que estén incluidos: el porcentaje más elevado se encuentra cuando se obvian los resultados obtenidos en viabilidad del dato y se atiende únicamente a la **representatividad** y la **utilidad**.

4. TERCERA FASE: ELABORACIÓN DE UNA VERSIÓN DEPURADA DE LA ESCALA Y APLICACIÓN EXPLORATORIA.

En esta tercera fase se llegó a la elaboración de una versión depurada de la escala para medir la calidad de estudios primarios tomando como base los resultados encontrados en el estudio de validez de contenido previamente descrito y otros ítems propuestos por expertos en medición de la calidad y meta-análisis (principalmente, por la Unidad de Meta-análisis de la Universidad de Murcia).

Por otro lado, se realizó un nuevo estudio exploratorio para probar el funcionamiento de dicha escala y conocer los resultados preliminares acerca de las características relacionadas con la calidad del diseño que suelen presentar los estudios primarios, esta vez en el ámbito concreto de la formación continua.

4.1. Elaboración de la escala (versión depurada).

4.1.1. Método.

4.1.1.1. Muestra.

Para la **elaboración de la versión depurada** de la escala para medir la calidad de los estudios primarios se utilizaron como muestra los 43 ítems que se presentaron en la fase anterior (ver anexo II, pág. vii) junto con los resultados obtenidos en el estudio de validez de contenido y otros ítems propuestos por expertos.

4.1.1.2. Instrumentos.

Se tomó, de los 43 ítems iniciales, los 23 que cumplieron el criterio de inclusión (al menos un valor de índice de Osterlind de 0.5 en dos de los tres conceptos estudiados). Hay que tener en cuenta que los ítems 26 y 27 (“mortalidad experimental” y “sin mortalidad”) se recogieron en sólo uno (el nuevo ítem 6 “mortalidad global”) y que el ítem 36 “técnicas de control” se desglosó en tres: los nuevos ítems 13, 14 y 15, referidos al “enmascaramiento del evaluador, del usuario y del profesional que realiza la intervención”, respectivamente. Por otro lado, se incluyeron otros ítems propuestos por expertos a través del cuestionario de validez de contenido y en reuniones.

4.1.1.3. Procedimiento.

El procedimiento seguido fue el siguiente:

1. **Recopilación de ítems que cumplieron el criterio de inclusión:** en primer lugar, se recogieron aquellos ítems que en el estudio de validez de contenido obtuvieron un índice de un valor de 0.5 ó mayor en al menos dos de los tres conceptos estudiados (representatividad, utilidad y viabilidad).
2. **Recogida de otros ítems propuestos por expertos:** a continuación, se recogieron aquellos ítems que fueron propuestos por los expertos participantes en reuniones realizadas o a través de la prueba de validez de contenido (concretamente, en los apartados de respuesta abierta).
3. **Comparación:** finalmente, se tomaron como base los ítems resultantes del estudio de validez de contenido y se añadieron nuevos ítems propuestos por expertos; en ocasiones, en ítems ya presentes, se crearon nuevas categorías. Se diferenció además entre los ítems que medían la calidad del reporte (referidos a si el artículo contenía todos los apartados propuestos por la APA, si provenía de una revista con alto índice de impacto, etc.) y los que se dirigían a la calidad del estudio en sí, tomándose únicamente los incluidos en este último caso.

4.1.2. Resultados.

En la tabla 2.4 que a continuación se presenta, se muestran los 33 ítems que conformaron finalmente dicho instrumento con sus diferentes opciones (columna de la izquierda) y las razones por las que se incluyó (columna de la derecha).

1. Grupo control: 0-inactivo; 1-activo	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999)
2. Criterios de selección de la muestra (inclusión/exclusión) (Assignment criteria; en Shadish, Cook y Campbell, 2002b, p.323): 0-no especificados; 1-especificados	Estudio validez de contenido (ítem 21; R, U y V \geq 0.5)
3. Azar: 0- pre-experimentales y cuasiexperimentales (sin control de vvee); 0.5-cuasiexperimentales con control de vvee (balanceo, bloqueo, estratificación); 1-experimentales (asignación aleatoria de las unidades a grupos)	Recomendación de expertos y estudio validez de contenido (ítem 22; R, U y V \geq 0.5)
4. Tipo de metodología/ diseño: 0-pre-experimental; 0.5-cuasi-experimental; 0.75-serie temporal (obs pre \geq 30 y post \geq 30) y discontinuidad en regresión; 1-experimental-aleatorio	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio

	validez de contenido (ítem 23; R, U y V \geq 0.5)
5. Muestra: 0 - n<12; 0.5 - n=[12-40]; 1- n>40	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 24; R, U y V \geq 0.5)
6. Mortalidad global: 0- \geq 20%; 0.5 - %=]0-20[; 1- 0%	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítems 26 y 27; R y U \geq 0.5)
7. Mortalidad diferencial: 0- \geq 20%; 0.5 - %=]0-20[; 1- 0%	Recomendación de expertos y estudio validez de contenido (ítem 28; R y U \geq 0.5)
8. Exclusiones posteriores a la asignación aleatoria (post-assignment attrition; en Shadish, Cook y Campbell, 2002b, p.323): 0- \geq 20%; 0.5 - %=]0-20[; 1- 0%	Estudio validez de contenido (ítem 29; R y U \geq 0.5)
9. Seguimiento: 0 – no se da seguimiento; 0.3< 6 meses; 0.6 – meses = [6-11]; 1- \geq 12 meses	Recomendación de expertos y estudio validez de contenido (ítem 31; R y U \geq 0.5)
10. Momentos de medida: 0-posterior; 1- previo y posterior	Recomendación de expertos y estudio validez de contenido (ítem 32; R, U y V \geq 0.5)
11. Las medidas del pre-test aparecen en el pos-test: 0- falta más de uno; 0.5- falta 1; 1- todas aparecen en todos los momentos	Recomendación de expertos y estudio validez de contenido (ítem 33; R y U \geq 0.5)
12. Variables dependientes normalizadas: 0- sólo autoinformes sin estandarizar; 0.5-ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado; 1-al menos un instrumento es objetivo o normalizado (p.ej. registro fisiológico, pruebas baremadas)	Recomendación de expertos y estudio validez de contenido (ítem 34; R y U \geq 0.5)
13. Enmascaramiento del evaluador: 0-no; 1-sí	Recomendación de expertos y estudio validez de contenido (ítem 36 desglosado; R y U \geq 0.5)
14. Enmascaramiento del usuario: 0-no; 1-sí	Recomendación de expertos y estudio validez de contenido (ítem 36 desglosado; R y U \geq 0.5)
15. Enmascaramiento del profesional que realiza la intervención: 0-no; 1-sí	Recomendación de expertos y estudio validez de contenido (ítem 36 desglosado; R y U \geq 0.5)
16. Homogeneidad de la intervención: 0- no todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales; 1- todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales	Recomendación de expertos y estudio validez de contenido (ítem 35, no valorado)
17. Definición del constructo: 0-no; 0.5-no operativa; 1-operativa	Estudio validez de contenido (ítem 37; R y U \geq 0.5)
18. Métodos estadísticos para inferir los valores perdidos: 0-no; 1-sí	Estudio validez de contenido (ítem 38; R y U \geq 0.5)
19. Tamaño de efecto y valor: 0-no; 1-sí	Estudio validez de contenido (ítem 40; R, U y V \geq 0.5)
20. Índice de calidad metodológica: suma de puntuaciones: 0-19	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999)
21. Índice estadístico calculado	Descriptivo, de interés para

	cálculo posible meta-análisis
22. Los resultados mostraron diferencias estadísticamente significativas (0-no; 1-sí)	Descriptivo, de interés para concluir acerca de la eficacia del programa
23. Índice de variabilidad facilitado	Descriptivo, de interés para posible meta-análisis
24. Número de participantes en cada grupo	Descriptivo, de interés para posible meta-análisis
25. Número de grupos en el estudio	Descriptivo, de interés para posible meta-análisis
26. Exclusiones tras medidas posteriores: 0-no; 1-sí	Descriptivo, de interés para posible meta-análisis
27. Rango de edad especificado: 0-no; 1-sí	Estudio validez de contenido (ítem 7; R, U y V \geq 0.5)
28. Media de edad (valor concreto)	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 8; R, U y V \geq 0.5)
29. Periodo de tratamiento: 0.5- \leq 6 meses; 1->6 meses	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 17; R, U y V \geq 0.5)
30. Intensidad del tratamiento: número de sesiones a la semana	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 18; R, U y V \geq 0.5)
31. Unidades de intervención: 0-individual; 1-grupal	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 19; R, U y V \geq 0.5)
32. Área formativa	Descriptivos, de interés para posible meta-análisis en el área concreta de la formación
33. Campo de intervención (destinatarios)	Usado en estudios meta-analíticos (Sánchez –Meca, Rosa y Olivares, 1999) y estudio validez de contenido (ítem 13; R y V \geq 0.5)
34. Tipo de publicación (1-Revista; 2-Libro; 3-Tesis; 4-Congreso; 5-Otros)	Estudio validez de contenido (ítem 1; U y V \geq 0.5)

Tabla 2.4. Proveniencia de cada uno de los ítems que compusieron la versión depurada de la escala de calidad.

La razón preponderante de la inclusión de los ítems en esta versión depurada de la escala fue la obtención en el estudio de validez de contenido de índices considerados positivos en al menos dos de los tres aspectos estudiados (representatividad, utilidad y viabilidad) y recomendados por expertos en diversas reuniones convocadas para tratar de hacer más exhaustivos el conjunto de ítems recogidos. En muchos casos, además, eran ítems utilizados frecuentemente en estudios meta-analíticos (por ejemplo, en Sánchez Meca, Rosa y Olivares, 1999). Los ítems 1, 16, del 20 al 26 y el 32 se incluyeron por otras razones:

- El **ítem 1** se incluyó por haber sido utilizado frecuentemente en los estudios meta-analíticos publicados.
- El **ítem 16** que, a pesar de haber sido incluido en el estudio de validez de contenido (ver el ítem 35 de la tabla 2.2), sólo superó el valor 0.5 en representatividad, fue incluido por recomendación de expertos en la materia.
- El **ítem 20** hace referencia a un índice global de calidad que se obtendría sumando los ítems del 1 al 19, que son aquellos que realmente se consideran indicadores de calidad metodológica de un programa. Fue incluido por su extendido uso en estudios meta-analíticos como, por ejemplo, en Sánchez-Meca, Rosa y Olivares (1999).
- El resto de ítems (del 21 en adelante), no puntúa para la obtención del índice de calidad mencionado, al ser únicamente descriptivos. Algunos de ellos se han incluido a posteriori, sin haber pasado previamente por el estudio de validez, por ser considerados de interés a nivel metodológico y/o sustantivo y por tratarse de posibles variables moduladoras a la hora de realizar un meta-análisis; así:
- Los **ítems 21, 22 y 23** versan sobre los datos aportados y la significación estadística de los resultados.
- Los **ítems 24 y 25** hacen referencia al número de participantes en cada grupo y al número de grupos en los estudios.
- El **ítem 26** responde a si existen exclusiones después de haber recogido alguna medida posterior.
- El **ítem 32** se incluyó por ser considerado de interés en el área concreta de evaluación de programas de formación continua.

4.2. Estudio exploratorio.

4.2.1. Método.

4.2.1.1. Muestra.

Para la **realización del estudio bibliográfico** (Sanduvete, Chacón, Holgado, Gómez y Sánchez, 2006), se hizo una búsqueda de lo publicado acerca de programas de formación continua. De los **2379** estudios encontrados en las bases de datos utilizadas en la búsqueda, tras la lectura de los resúmenes, se incluyeron únicamente aquéllos que versaban sobre formación continua en las organizaciones, desechando por tanto los programas de educación familiar, la formación ocupacional (dirigida a desempleados), las acciones formativas en las que se trabajaba con cuidadores no formales (como pueden ser los familiares de las personas mayores con dependencia) o los cursos realizados por la universidad para sus estudiantes, por ejemplo. Cumplieron este criterio de inclusión **287** resúmenes de trabajos, de los que finalmente se estudiaron **95** por ser los únicos de los que se consiguió obtener el texto completo.

4.2.1.2. Instrumentos.

Los instrumentos utilizados fueron la escala resultante tras ser depurada; el software “Procite-5” para el tratamiento de la información obtenida; el programa estadístico “SPSS 12.0” para la codificación de los datos y su posterior análisis; y Microsoft Excel para el cálculo del índice de calidad (ítem 20).

4.2.1.3. Procedimiento.

Se realizó una búsqueda en todas las bases de datos a las que se tenía acceso en la Universidad de Sevilla, que guardaran relación con el tema a estudiar. Concretamente, fueron EBSCO Online, Medline, Serfile, CABHealth, CINAHL, PsycINFO, Econlit, ERIC, MathSci, Current Contents y Humanities Index. Se recogieron artículos publicados hasta septiembre de 2006.

Las **palabras clave** introducidas fueron tres de manera conjunta; es decir, en búsquedas avanzadas, unidas por el nexa “AND” para que los resultados obtenidos contuvieran los tres términos. La búsqueda se solicitó en todos los posibles campos: título, resumen, palabras clave, texto completo, etc. Concretamente, las palabras clave fueron:

- **Training programs** (programas de formación): se eligieron estas palabras porque la intención era encontrar los programas de formación publicados para posteriormente estudiar su calidad a través del instrumento construido para tales fines.
- **Evaluation** (evaluación): al incluir esta palabra se intentó restringir la búsqueda para encontrar aquellos programas que mostraran los resultados encontrados y, posiblemente, algunas características metodológicas de diseño.
- **Work** (trabajo): incluyendo esta palabra en la búsqueda se intentó restringir los estudios encontrados al área concreta de las organizaciones, disminuyendo así la probabilidad de encontrar acciones formativas dirigidas a personas fuera del ámbito laboral.

Estas palabras clave no sirvieron para obtener unos resultados muy específicos, encontrándose **2379** estudios. Se procedió a la lectura de todos los resúmenes para tratar de acotar y estudiar únicamente lo que interesara, aplicando para ello un **criterio de exclusión**: se tomó como requisito imprescindible que la acción fuera dirigida a empleados y empleadas de alguna organización; se excluyeron, por tanto, aquellos trabajos que versaban sobre acciones formativas no dirigidas a empleados o empleadas como podían ser la formación ocupacional y los programas dirigidos a estudiantes o a cuidadores informales. Tras seguir este criterio, el número de estudios a incluir se redujo considerablemente, quedando únicamente **287**.

Tras esta criba, se procedió a la búsqueda de los textos completos a través de internet. Se encontraron definitivamente **95** estudios, que fueron codificados con la escala para medir la calidad de los estudios primarios previamente elaborada por tres codificadores independientes, que obtuvieron un grado de acuerdo bajo, con un índice de correlación intraclase de 0.5.

4.2.2. Resultados.

Los resultados se presentan en el anexo VI (pág. xxxi) (Sanduvete, Chacón, Holgado, Gómez y Sánchez, 2006). Se constató en el ámbito de la formación continua las mismas conclusiones obtenidas para los estudios psicológicos con el primer estudio exploratorio realizado: suele darse **poca especificación** de las características metodológicas (sin especificación de los criterios de selección de la muestra, definición no operativa de los constructos, sin inferencia de valores perdidos, sin especificación del tamaño de efecto y rango de edad no especificado); y un **grado medio de control y estandarización** (uso de diseños pre-experimentales y cuasiexperimentales, mortalidad global y diferencial, exclusiones posteriores a la asignación aleatoria y tras tomar una medida, sin seguimiento, sólo medida posterior, variables dependientes semiestandarizadas, sin uso del enmascaramiento y corto periodo de intervención).

Pero quizá más interesante que estos resultados en sí, fue el hecho de que este estudio exploratorio permitió detectar algunos aspectos relacionados con el funcionamiento de la escala que habrían de ser cambiados. Concretamente, se constató lo siguiente:

- No todos los ítems eran aplicables a todos los tipos de diseño de intervención por lo que, en la práctica, sólo los diseños experimentales podían alcanzar la puntuación máxima (19) en el índice de calidad metodológica. Además, los diseños cuasiexperimentales podían alcanzar una puntuación máxima mayor que aquellos estudios que no presentaban intervención. Se consideró que este trato desigual a favor de los diseños experimentales y, en segundo lugar, de los cuasiexperimentales, impedía la posibilidad de comparar resultados entre los distintos diseños.
- Algunas categorías e ítems resultaban algo ambiguos por falta de definición operativa de los constructos de los que se trataban o por poca especificación a la hora de marcar los intervalos, con lo que podrían surgir dudas acerca de la categoría concreta en que encuadrar un estudio determinado.

Por ello, se hizo necesario afinar más y hacer modificaciones. A continuación, se presenta la última modificación realizada a la escala de medición de la calidad de los estudios primarios, con la que se trató de lograr dos objetivos (Chacón, Sanduvete, Sánchez y Sánchez-Meca, 2007b):

- Homogeneizar su uso para todos los diseños, de tal manera que cualquier estudio pudiera puntuar de 0 a 19, independientemente del tipo de diseño que presentara en lugar de que, por el hecho de ser un diseño experimental, partiera de ventajas en la máxima puntuación posible a alcanzar en comparación con los demás diseños.

Se trató por tanto de adaptar los ítems a cualquier tipo de diseño, intentando así que el instrumento fuera flexible y que permitiera valorar de manera igualitaria la calidad de cualquier estudio, independientemente del diseño que presentara.

- Concretar, especificar y operativizar aquellos ítems o categorías de ítem que provocaron ambigüedades a los codificadores en la versión anterior.

5. DOBLE PROCESO DEDUCTIVO-INDUCTIVO: ELABORACIÓN DE UNA VERSIÓN INTEGRADORA DE LA ESCALA.

A continuación, se explica el método seguido para obtener la hasta ahora última versión de la escala.

5.1. Método.

5.1.1. Muestra.

La muestra utilizada fueron los 34 ítems que conformaron la escala de medición de la calidad en estudios primarios en su versión previa (tal y como se presenta en el apartado anterior).

5.1.2. Instrumentos.

Aparte de la escala previamente elaborada, no se utilizó ningún otro instrumento.

5.1.3. Procedimiento.

Se partió de los 34 ítems de los que constaba la escala en la versión previa. Tras ser aplicada en los estudios que versaban sobre formación continua, se prestó atención a cómo se comportaba cada ítem en los distintos estudios. Se detectaron:

- Aquellos ítems que mostraron diferencias en función del diseño que los estudios presentaban.
- Aquellos ítems que mostraron alguna dificultad de comprensión a la hora de decidir qué categoría escoger, una de las posibles causas por la que se obtuvo un bajo índice de acuerdo intercodificadores.

Una vez detectados estos ítems, se pasó a la realización de los cambios oportunos para lograr:

- La comparabilidad de los datos obtenidos en cada estudio, sin que el tipo de diseño presentado supusiera un aspecto diferenciador. Se logró que todos los ítems fueran aplicables a todos los tipos de diseño y que todos pudieran sumar en el índice de calidad de 0 a 19 (compuesto por 19 ítems, concretamente del 1 al 19, según la tabla 2.5 que a continuación se presenta).
- Claridad conceptual de los ítems y operacionalización para facilitar la elección de la categoría adecuada a la hora de la codificación.

5.2. Resultados.

A continuación, en la tabla 2.5, se presenta la escala resultante (Chacón, Sanduvete, Sánchez y Sánchez-Meca, 2007b). Además, para mayor claridad, en el anexo VII (pág. xxxviii) se muestran tres variantes de la misma escala: la primera adaptada a estudios observacionales y de encuesta (en aquellos casos donde no hay intervención, sino sólo recogida de la respuesta que da el sujeto); la segunda, a estudios cuasiexperimentales (donde sí hay intervención, pero la selección y/o la asignación a los distintos grupos no se hace al azar); y la última, a estudios experimentales (donde hay intervención y la selección y asignación se realizan aleatoriamente).

Para facilitar la interpretación, a continuación se explican las partes que componen la escala:

- La primera columna muestra el **ítem** que se está valorando.
- La segunda columna (**V**) presenta el **valor** que obtiene cada una de las opciones en cada ítem. Sólo los valores obtenidos en los ítems del **1 al 19** (en casillas sombreadas) son sumados para el cálculo del **índice global de calidad** (ítem 20); los demás, son únicamente descriptivos. Las valoraciones de estos 19 ítems oscilan entre 0 como menor puntuación y 1 como mayor por lo que, consecuentemente, este índice oscilará entre el valor 0 como menor y el 19 como más alto (los valores **8 y 9** no suman en ningún caso y son tratados en los análisis de manera diferente por hacer referencia a datos no disponibles y no aplicables respectivamente). Las diferencias de puntuación entre las distintas opciones son iguales (por ejemplo, 0, 0.5 y 1 para ítems con tres opciones de respuesta; 0, 0.33, 0.66 y 1 para ítems con cuatro opciones de respuesta), a excepción del ítem 4 referido al diseño donde, entre el 0.5 y el 1, se incluyeron con un 0.75 las “series temporales”, consideradas con más calidad que los “diseños cuasiexperimentales” pero no con tanta calidad como los “diseños de discontinuidad en la regresión” y los “diseños experimentales” (Shadish y Myers, 2004). No se pusieron las cuatro opciones con distancias iguales por considerarse que los “diseños cuasiexperimentales” estaban más cercanos a los “diseños de series temporales” que a los “pre-experimentales”. La explicación de este proceder, que en principio puede parecer arbitrario por lo novedoso, es que los diseños pre-experimentales presentan muchos problemas de validez que los “cuasiexperimentales” solventan (lo cual supone una gran diferencia entre ambos), mientras que las ventajas diferenciales entre las otras tres opciones son significativas pero más graduales: los diseños cuasiexperimentales no llegan a presentar la ventaja de los “diseños de series temporales” que suponen estudiar la evolución de un dato durante un largo periodo de tiempo (detección de posibles tendencias cíclicas, mayor seguridad sobre la consistencia de los datos, etc.); pero éstos últimos no llegan a tener la ventaja del conocimiento del criterio de asignación que los “diseños de discontinuidad en la regresión” presentan (Chacón, Shadish y Cook, en prensa).
- La tercera columna muestra la etiqueta de cada una de las **categorías**.
- Los ítems son diferenciados en función de las características a las que atiende; así, del ítem **0 al 23** se tratan las **características metodológicas**; y del **24 al 34** se hace referencia a **características sustantivas**.

ÍTEM	V	CATEGORÍAS
CARACTERÍSTICAS METODOLÓGICAS		

0. Tipo de estudio	1	Teórico (se describen modelos o no hay datos)
	2	Observacional
	3	Encuesta
	4	Cuasiexperimental
	5	Experimental
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
1. Grupo de comparación	0	No hay
	0.5	Inactivo
	1	Activo; en diseños observacionales y de encuesta , cuando sí hay
	8	No se especifica el dato
	9	No aplicable (teórico)
2. Criterios de selección de la muestra (inclusión/exclusión)	0	No especificados
	1	Especificados
	9	No aplicable (por el objeto de estudio o por ser teórico)
3. Azar	0	Observaciones o encuestas sin selección aleatoria de la muestra ni delimitación de los criterios de inclusión Pre-experimentales y cuasiexperimentales sin control de variables extrañas ni delimitación de los criterios de inclusión y asignación Experimentales con asignación y/o selección aleatorias inadecuados
	1	Observaciones o encuestas con selección aleatoria de la muestra o delimitación de los criterios de inclusión Pre-experimentales y cuasiexperimentales con control de variables extrañas o delimitación de los criterios de inclusión y asignación Experimentales con asignación y selección aleatorias adecuados
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
	0	Observaciones o encuestas con una o dos medidas Pre-experimentales Experimentales con un solo momento de medida
4. Diseño	0.5	Observaciones o encuestas con medidas entre 3 y 29 Cuasiexperimentales pre-post con grupo control no equivalente con medidas entre 2 y 29
	0.75	Series temporales: cuasiexperimentales con 30 o más medidas
	1	Observaciones o encuestas con 30 ó más medidas Discontinuidad en la regresión Experimentales con al menos dos momentos de medida
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
	0	$N < 12$
5. Muestra	0.5	$12 \leq N \leq 40$
	1	$N > 40$
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
	0	$\geq 20\%$
6. Mortalidad global	0.5	$0 < N < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
	0	$\geq 20\%$
7. Mortalidad diferencial	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (teórico; un solo grupo; objeto de estudio incompatible)
	0	$\geq 20\%$
8. Exclusiones posteriores	0	$\geq 20\%$

a la agrupación de la muestra a las distintas condiciones	0.5	0<n<20%
	1	0%
	8	No se especifica el dato
	9	No aplicable (teórico; sólo un grupo; objeto de estudio incompatible)
9. Seguimiento	0	Nada
	0.3	<6 meses
	0.6	[6-11] meses
	1	≥12 meses
	8	No se especifica el dato
	9	No aplicable (teórico; no es posible recoger el dato en más de una ocasión; objeto de estudio incompatible)
10. Momentos de medida	0	Sólo posterior; una medida cuando no hay intervención
	1	Previo y posterior; más de una medida cuando no hay intervención
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
11. Medidas que aparecen en todos los momentos de registro	0	Más de una medida no aparece en todos los momentos de registro; también cuando ocurre esto en la única variable que se mide
	0.5	Una medida no aparece en todos los momentos de registro (siempre que se mida más de una variable)
	1	Todas las medidas son tomadas en todos los momentos de registro
	8	No se especifica el dato
	9	No aplicable (teórico; sólo un momento de registro; objeto de estudio incompatible)
12. Variables dependientes normalizadas (uso de instrumentos normalizados)	0	Sólo autoinformes sin estandarizar
	0.5	Ninguno normalizado pero al menos uno es cuestionario o autoinforme estandarizado (explícitamente expresado)
	1	Al menos uno es objetivo o normalizado
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
13. Enmascaramiento del evaluador	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
14. Enmascaramiento del usuario	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
15. Enmascaramiento del formador (como evaluador interno cuando no hay intervención)	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
16. Homogeneidad de la intervención o del proceso de registro cuando no hay intervención	0	No mismos proceso, intensidad (n° sesiones), duración y/o profesionales
	1	Mismos proceso, intensidad, duración y profesionales
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
17. Definición de los constructos	0	Ninguno definido conceptual y/o empíricamente
	0.5	Al menos 1 definido conceptual y/o empíricamente
	1	Todos definidos conceptual y empíricamente
	8	No se especifica el dato
18. Métodos estadísticos para inferir los valores perdidos	0	Ninguno. Sólo se analizaron los datos completos
	1	Sí o análisis "por intención de tratar"
	8	No se especifica el dato
	9	No aplicable (teórico; no se dieron valores perdidos; objeto de estudio incompatible)
19. Tamaño de efecto y	0	No se especifica

valor	1	Se especifica (o índices derivables)
	9	No aplicable (teórico; sólo una medida en un grupo; objeto de estudio incompatible)
20. Índice de calidad	0-19	Suma
21. Índice estadístico calculado		Anotar valor concreto
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
22. Diferencias estadísticamente significativas entre medidas (explicitar qué se compara)	0	No
	1	Sí
	8	No se especifica el dato
	9	No aplicable (teórico; una sola medida; objeto de estudio incompatible)
23. Índice de variabilidad		Anotar valor concreto
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo		Anotar valor concreto (la suma ha de dar el valor de la muestra)
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
25. Número de grupos en el estudio		Anotar valor concreto
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
26. Exclusiones tras medidas posteriores	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (teórico; una sola medida; objeto de estudio incompatible)
27. Rango de edad especificado	1	No se especifica
	2	Sí se especifica
	9	No aplicable (por el objeto de estudio o por ser teórico)
28. Media de edad		Anotar valor concreto
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
29. Periodo de estudio	1	≤ 6 meses
	2	> 6 meses
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
30. Intensidad del tratamiento/registro		nº de horas/ periodo de tiempo
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
31. Unidades de intervención o de registro	1	Individual (o grupo considerado como unidad)
	2	Grupal
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio o por ser teórico)
32. Área formativa		Anotar valor concreto
33. Campo de intervención: destinatarios		Anotar valor concreto
34. Tipo de publicación	1	Revista
	2	Libro
	3	Tesis
	4	Congreso
	5	Otras publicaciones
	6	Trabajos no publicados

Tabla 2.5. Versión integradora de la escala de medición de la calidad en estudios primarios.

Una diferencia de la última versión respecto a la previa es que, anteriormente, los estudios sin intervención no eran tenidos en cuenta con lo que, aplicando la escala, dichos estudios habrían sido no calificados en muchos de los ítems presentados; y los diseños mejor valorados eran los experimentales, seguidos de los cuasi y, finalmente, los que no mostraban intervención.

Partiendo de la idea de que el tipo de diseño no necesariamente correlaciona con la calidad metodológica, se trató de ampliar las categorías de cada ítem para adaptarlas a cualquier tipo de diseño de tal manera que el índice de calidad pudiera oscilar entre 0 y 19 independientemente del tipo de diseño que presentara el estudio (Chacón, Sánchez-Meca y Sanduverte, 2007).

En el apartado de discusión, se presenta una descripción detallada de los ítems que fueron modificados, los cambios concretos que se realizaron y las razones de dichos cambios.

A continuación, al entenderse que la definición clara y concisa de cada uno de los ítems y sus categorías es fundamental pues, para optimizar su utilidad, es muy importante que cualquier persona que quiera utilizar esta escala sepa a qué se refiere cada apartado y que todos tengan el mismo concepto respecto a lo que se está midiendo en cada momento, se presenta el **manual de codificación** que pretende aportar claridad al significado de cada uno de los conceptos para aumentar así la fiabilidad entre codificadores cuando se aplique esta escala.

CARACTERÍSTICAS METODOLÓGICAS

Ítem 0. Tipo de estudio. Es una categoría ómnibus en la que se trató de etiquetar de manera genérica el trabajo a categorizar tras hacer una primera lectura superficial, con la intención de facilitar la codificación del resto de variables.

a) Teórico: se consideraron teóricos todos aquellos trabajos en los que únicamente se describían modelos y/o donde no existían datos empíricos directos del contexto de intervención.

b) Observacional: fueron aquellos estudios en los que no se llevó a cabo ninguna intervención, sólo se registró, el contexto de los sujetos apenas fue modificado y la respuesta registrada no fue provocada.

c) Encuesta: fueron aquellos estudios en los que no se llevó a cabo ninguna intervención pero, para recoger información, se requirió a las personas que respondieran a determinados estímulos (en general, ítems por escrito o entrevistas).

d) Cuasi-experimental: fueron aquellos trabajos en los que se manipuló al menos una variable independiente para estudiar su efecto en una dependiente (por tanto, hubo intervención) y no hubo selección y/o asignación aleatoria de las personas que conformaron la muestra a las distintas condiciones.

e) **Experimental:** fueron aquellos estudios en los que se manipuló al menos una variable independiente para estudiar su efecto en una dependiente (hubo intervención) y las personas que conformaron la muestra fueron seleccionadas y posteriormente asignadas a los distintos grupos de manera aleatoria.

f) Cuando no se dio la información, se asignó un 8.

g) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 1: Grupo de comparación. Con este ítem se determina si existía un grupo aparte del considerado foco de atención del estudio y, en caso de que lo hubiera, qué nivel de participación tiene.

a) **Sin grupo de comparación:** cuando no hubo grupo de comparación, se asignó un 0.

b) **Grupo de comparación inactivo:** Los sujetos que conformaron el grupo de comparación no recibieron ninguna formación alternativa al grupo de intervención, bien porque estaban esperando a que comenzara la acción formativa o porque sencillamente no estaba prevista su realización. También se consideró grupo control inactivo al que no se le cambió la forma habitual de formarse (Shadish y Ragsdale, 1996). En este caso, se asignó un 0.5.

c) **Grupo de comparación activo:** Los sujetos del grupo control recibieron algún tipo de formación alternativa al grupo de intervención. Se valoró con un 1. En caso de que no se llevara a cabo intervención en el estudio, se valoró con esta máxima puntuación la existencia de un grupo de comparación observado o encuestado, aunque no fuera posible catalogarlo de activo.

d) Cuando no se dio la información, se asignó un 8.

e) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 2. Criterios de selección de la muestra (inclusión/exclusión): hace referencia a si se explicitaron los criterios de decisión por los que algunas personas participaron en el estudio, mientras que otras quedaron fuera. Es lo que Shadish, Cook y Campbell (2002b) denominan “assignment criteria”.

a) **No se especificaron:** en los casos en los que no se explicitaron dichos criterios de selección, se valoró con un 0.

b) **Se especificaron:** en los casos en los que se explicitaron dichos criterios, se valoró con un 1.

c) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 3. Azar: hace referencia a si la selección de la muestra fue aleatoria y si los participantes fueron o no asignados al azar a las distintas condiciones en los casos en los que había más de una.

- a) Se valoraron con un 0 los siguientes casos: las **observaciones y encuestas** cuya selección de la muestra **no** fue **aleatoria**; los **diseños pre-experimentales y cuasiexperimentales** (del tipo “pretest-posttest con grupo control no equivalente”) **sin control de variables extrañas**; los **diseños experimentales** con un **proceso de aleatorización inadecuado** en la selección de la muestra o la posterior asignación a los diferentes grupos.
- b) Se valoraron con un 1 los casos siguientes: los diseños **observacionales y de encuestas** en los que la selección de la muestra se hizo **aleatoriamente**; los **diseños pre-experimentales y cuasiexperimentales** (del tipo “pretest-posttest con grupo de control no equivalente”) **con control de variables extrañas** como, por ejemplo, balanceo, bloqueo y estratificación; los **diseños experimentales** en los que la selección de la muestra y la posterior asignación de las personas a los distintos grupos se hizo con un sistema de aleatorización adecuado.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 4. Diseño: Se valoró en este ítem la validez interna, la calidad y la estabilidad del dato (Cronbach, 1982).

- a) Se asignó un 0 a las **observaciones o encuestas** con 1 ó 2 medidas; a los **diseños pre-experimentales**, caracterizados por la ausencia de selección y/o asignación aleatoria y porque sólo existe un grupo de personas que es medido transversalmente; por último, también a los **diseños experimentales** (en los que existe más de un grupo formado aleatoriamente) cuando sólo se da un momento de medida.
- b) Se asignó un 0.5 cuando se trató de **metodología observacional** o de **encuestas**, con un número de medidas entre 3 y 29 (ambos valores inclusive); y los **diseños cuasiexperimentales** cuando se toman varias medidas a lo largo del tiempo, pero menos de 30.
- c) Se asignó 0.75 a las **series temporales**, un tipo de diseño cuasiexperimental en el que un grupo es medido en 30 o más ocasiones a lo largo del tiempo en el pretest y el posttest (Shadish y Myers, 2004).
- d) Fueron valorados con un 1 los estudios **observacionales y de encuestas** con 30 o más medidas; los **diseños de discontinuidad en la regresión**, que son diseños cuasiexperimentales en los que se establece un punto de corte a partir del cual las personas van a recibir la intervención, por lo que se puede decir que los criterios de asignación a los grupos son totalmente conocidos; y los **diseños experimentales** con al menos dos momentos de medida.

- e) Cuando no se especificó el dato, se asignó un 8.
- f) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 5. Muestra: Se valora en este ítem el tamaño total de la muestra del estudio (N). Se encuadró cada caso en el siguiente sistema de puntuación:

- a) **N<12:** cuando participaron en el estudio menos de doce personas, se asignó un 0.
- b) **12≤N≤40:** cuando participaron entre 12 y 39 personas (ambos valores inclusive), se asignó un 0.5.
- c) **N>40:** cuando participaron más de cuarenta personas, se asignó un 1.
- d) En las ocasiones en las que no se especificó este dato, se asignó un 8.
- e) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 6. Mortalidad global (Shadish, Cook y Campbell, 2002b). Se valora el número de personas que comienzan el estudio pero que, por diversas razones, no llegan a concluirlo. Las distintas categorías son:

- a) **≥20%:** cuando el 20% o más de la muestra inicial no llegó a concluir el estudio, se asignó un 0.
- b) **0<N<20%:** cuando entre el 0 y el 20% de la muestra inicial no llegó a concluir el estudio, se asignó un 0.5.
- c) **0%:** cuando no se dio mortalidad (es decir, cuando todas las personas que comenzaron el estudio lo finalizaron), se asignó un 1.
- d) Cuando no se especificó el dato, se asignó un 8.
- e) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 7. Mortalidad diferencial: Este ítem valora la mortalidad diferencial sufrida entre dos grupos de comparación en un segundo momento de medida o posterior, con el siguiente sistema de puntuación:

- a) **≥20%:** cuando la diferencia de mortalidad experimental entre grupos fue del 20% o más, se valoró este ítem con un 0.
- b) **0<n<20%:** cuando la diferencia de mortalidad experimental entre grupos fue de entre el 0 y el 20%, se valoró con un 0.5.

- c) **0%:** Cuando la mortalidad experimental fue igual en todos los grupos existentes, se valoró con un 1.
- d) Cuando no se especificó el dato, se asignó un 8.
- e) Cuando se consideró este ítem como no aplicable por tratarse de un estudio teórico o por tener un solo grupo, se asignó un 9.

Ítem 8. Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones: este ítem hace referencia al porcentaje de sujetos que, habiendo sido asignados (o perteneciendo por sus propias características) a una de las condiciones de estudio, posteriormente son excluidos. Las puntuaciones son las siguientes:

- a) **$\geq 20\%$:** cuando el 20% o más de la muestra fue excluida tras la agrupación en las distintas condiciones, se valoró este ítem con un 0.
- b) **$0 < n < 20\%$:** cuando entre el 0 y el 20% de la muestra fue excluida tras la agrupación en las distintas condiciones, se asignó un 0.5.
- c) **0%:** cuando no se excluyó posteriormente a ninguna de las personas agrupadas en las distintas condiciones, se valoró este ítem con un 1.
- d) Cuando no se especificó el dato, se asignó un 8.
- e) En las ocasiones en las que el estudio era teórico o sólo existía un grupo de personas o de medidas, se consideró este ítem como no aplicable, asignándosele un 9.

Ítem 9. Seguimiento: Se determina con este ítem durante cuánto tiempo se toman medidas, ya sea de la variable dependiente *una vez finalizada* la intervención, o de las variables registradas en los estudios en los que no hay intervención. Si se presentara más de un periodo de seguimiento, se tomaría el más amplio para responder a este ítem. Se distingue entre las siguientes posibilidades:

- a) **0:** No se tomó ninguna medida de seguimiento. En estos casos, se asignó un 0.
- b) **<6 meses:** Se tomaron medidas durante menos de 6 meses. En este caso, se asignó un 0.3.
- c) **[6-11] meses:** Se tomaron medidas entre los 6 y 11 meses, ambos incluidos. En este caso, se asignó un 0.6.
- d) **≥ 12 meses:** Se tomaron medidas durante un año o más, en cuyo caso se valoró este ítem con un 1.
- e) Cuando no se especificó el dato, se asignó un 8.

- f) Cuando se consideró este ítem como no aplicable por tratarse de un estudio teórico, se asignó un 9.

Ítem 10. Momentos de medida: Se pretende concretar cuándo se tomaron las medidas. Las categorías son:

- a) **Sólo posterior o sólo una medida cuando no hay intervención:** se consideraron parte de esta categoría aquellos estudios en los que todas las medidas se tomaron tras la intervención y aquellos en los que, no habiendo intervención, se daba una sola medida. En estas ocasiones, se valoró con un 0.
- b) **Previo y posterior o más de una medida cuando no hay intervención:** se eligió esta categoría cuando las medidas fueron tomadas antes y después de la intervención o cuando, sin haber intervención, se daba más de una medida de tal manera que se posibilitaba su comparación. En estas ocasiones, se valoró con un 1.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 11. Medidas que aparecen en todos los momentos de registro: se hace en este ítem el conteo de medidas que, siendo tomadas en el primer momento, también aparecen en todos los demás momentos de registro. Se distingue entre:

- a) **Más de una no se mide en todos los momentos de registro:** más de una medida inicial no se repite en todos los registros posteriores. Estos casos se valorarán con un 0, al igual que aquellos casos en los que **sólo se mide una variable y ésta no se registra en todos los momentos de medida.**
- b) **Todas menos una:** una medida inicial no se repite en todos los registros posteriores, siempre que se mida más de una variable; de lo contrario, formaría parte del apartado a) previamente comentado. Estos casos se valoraron con un 0.5.
- c) **Todas:** todas las medidas iniciales se repiten en todos los sucesivos registros. Estos casos fueron valorados con un 1.
- d) Cuando no se especificó el dato, se asignó un 8.
- e) Cuando se consideró este ítem como no aplicable por tratarse de un estudio teórico o por tener sólo un momento de registro, se asignó un 9.

Ítem 12. Variables dependientes normalizadas: se pretende determinar qué grado de estandarización y objetividad poseen las variables medidas en función del instrumento que se utiliza para su registro (Anguera, Chacón, Holgado y Pérez, en prensa). Se diferencian las siguientes categorías:

- a) Únicamente **no estándares**: hace referencia a aquellas medidas tomadas con instrumentos elaborados para una situación de evaluación concreta y no validados ni baremados, por lo que no pueden asegurar la validez de la medida y no existe una población de referencia con la que comparar el dato obtenido en una situación concreta. Algunos instrumentos con estas características son los registros de conducta, las fuentes documentales y los autoinformes sin estandarizar (para que se consideren estandarizados, se ha de comentar este hecho explícitamente en el estudio). Estos casos se valoraron con un 0.
- b) Ninguno normalizado, pero al menos uno es **estándar**: aunque no hubo ninguna variable dependiente medida con instrumento normalizado, al menos una de ellas fue medida con instrumentos que, aunque no estaban baremados ni validados, sí estaban estructurados, por lo que se pudo afirmar con cierto grado de seguridad que se estaba midiendo, mediante un procedimiento reglado y homogéneo para todos los usuarios, aquello que se pretendía medir. Algunos ejemplos de estos instrumentos son las entrevistas estructuradas, los cuestionarios, las encuestas y los autoinformes estandarizados (dicho explícitamente, pues de lo contrario se consideró un autoinforme sin estandarizar y formó parte de la opción a). Estos casos se valoraron con un 0.5.
- c) Al menos uno es **normalizado**: se incluyen en esta categoría aquellos casos en los que, al menos, una variable dependiente fue medida con pruebas objetivas, baremadas (con una población de referencia con la que comparar el dato obtenido) y/o validadas. Unos ejemplos de estos instrumentos son los registros fisiológicos, los tests y los registros de observación. Estos casos se valoraron con un 1.
- d) Cuando no se especificó el dato, se asignó un 8.
- e) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 13. Enmascaramiento del evaluador (codificador): se valora si los evaluadores desconocían las hipótesis del estudio, con el siguiente sistema de puntuación:

- a) **No**: los evaluadores conocían las hipótesis del estudio. En estos casos, se asignó un 0.
- b) **Sí**: los evaluadores desconocían las hipótesis del estudio. En estos casos, se asignó un 1.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 14. Enmascaramiento del usuario: se valora si las personas que participaron en el estudio desconocían las hipótesis de partida.

- a) **No:** el alumnado conocía las hipótesis del estudio. En estos casos, se asignó un 0.
- b) **Sí:** el alumnado desconocía las hipótesis del estudio. En estos casos, se asignó un 1.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 15. Enmascaramiento del profesional que realiza la intervención o del evaluador interno cuando no hay intervención: se valora si las personas encargadas de implementar la intervención desconocían las hipótesis del estudio; en los casos en los que no hubo intervención, se valoró con este ítem el conocimiento o desconocimiento de los objetivos por parte de un evaluador interno (en caso de que lo hubiera) que sería una persona que, formando parte de los sujetos estudiados, recoge datos para dicho estudio.

- a) **No:** los formadores o evaluadores internos conocían las hipótesis del estudio. En estos casos, se asignó un 0.
- b) **Sí:** los formadores o evaluadores internos desconocían las hipótesis del estudio. En estos casos, se asignó un 1.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 16. Homogeneidad de la intervención o proceso de registro: se valora la integridad de la intervención o proceso de registro; es decir, el grado en que todas las personas del estudio participaron en intervenciones con las mismas condiciones (intensidad -nº sesiones-, horas de cada acción formativa y formador) y, en caso de que no hubiera intervención, se valora si el proceso de registro fue para todos igual.

- a) **No:** las personas participaron en intervenciones con distinta intensidad, duración y/o profesionales. Cuando no hubo intervención, el proceso de registro varió. En estos casos, el ítem fue valorado con un 0.
- b) **Sí:** Todas las personas participaron en intervenciones con las mismas características en intensidad, duración y profesionales. Cuando no hubo intervención, todos los datos se registraron con el mismo proceso. En estos casos, se valoró con un 1.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 17. Definición del constructo: se valora si los constructos de interés en la investigación se han definido con claridad y de forma operativa. En concreto, debe hacerse referencia, al menos, al tipo de intervención implementada (variable/s independiente/s) si la hubiese, y a la medida de las variables dependientes o de respuesta, con el siguiente sistema de puntuación:

- a) **Ninguno definido conceptual y/o empíricamente:** Cuando ningún constructo implicado en el estudio se definió de forma conceptual ni empírica (operativa), se valoró con un 0.
- b) **Al menos uno definido conceptual y/o empíricamente:** cuando al menos uno de los constructos implicados en el estudio se definió de forma conceptual y/o empírica (operativa), se valoró con un 0.5.
- c) **Todos definidos conceptual y empíricamente:** los casos en los que todos los constructos implicados en el estudio se definieron tanto conceptual como empíricamente (operativamente) se valoraron con un 1.
- d) Cuando no se especificó el dato, se asignó un 8.

Ítem 18. Métodos estadísticos para inferir los valores perdidos: Se valora si se ha realizado algún tratamiento estadístico de los valores perdidos.

- a) **No:** los casos en los que no se aplicó ningún procedimiento estadístico de imputación de valores perdidos, o no se realizaron los análisis estadísticos incluyendo los datos “por intención de tratar”, sino sólo los de aquellas personas cuyo registro incluía todas las medidas posibles (“completer analysis”), se valoró con un 0.
- b) **Sí:** cuando se aplicó algún procedimiento estadístico de imputación de valores perdidos o se realizaron los análisis estadísticos sobre los datos “por intención de tratar” (“intention-to-treat analysis”), este ítem fue valorado con un 1. Se anotó el nombre del procedimiento concreto que se siguió.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Cuando se consideró el ítem como no aplicable por tratarse de un estudio teórico o por no haber valores perdidos, se asignó un 9.

Ítem 19. Tamaño de efecto y valor: consiste en discernir si los resultados del estudio fueron dados mediante un estadístico que muestra el tamaño de efecto.

- a) **No:** cuando los resultados del estudio no fueron dados mediante un estadístico que mostraba el tamaño de efecto, se valoró con un 0.
- b) **Sí:** cuando los resultados del estudio fueron dados mediante un estadístico que mostraba el tamaño de efecto como, por ejemplo, la diferencia de medias estandarizada “d”; o mediante índices derivables como “O Ratio” o la media, desviación tipo y tamaño de la muestra (son necesarios los tres valores para calcular el tamaño de efecto), se asignó un 1.

- c) Si se consideró este ítem como no aplicable por tratarse de un estudio teórico o en el que sólo existiera una medida en un solo grupo (hecho que impediría cualquier posible comparación), se asignó un 9.

Ítem 20. Índice de calidad: en esta variable se calculó la suma de las puntuaciones que el estudio alcanzó en los ítems anteriores (excluyendo el “ítem 0” considerado meramente descriptivo), con los que se evaluó la calidad metodológica. Las puntuaciones mínima y máxima en este ítem, por tanto, fueron 0 y 19 respectivamente.

Los ítems que a partir de ahora se presentan (3 referidos a características metodológicas y 11 a sustantivas) tienen como finalidad la realización de un estudio descriptivo de las características que suelen mostrar los trabajos que versan sobre formación continua para empleados y empleadas. Por esta razón, no son contabilizados de manera cuantitativa y no aportan valor al índice de calidad (ítem 20). Esta es también la causa por la que no se aporta un valor numérico a cada categoría, como aparecía en los 20 ítems anteriores (en la escala aparece por facilitar la codificación, pero son números que no aportan valor respecto al constructo “calidad”).

Ítem 21. Índice estadístico calculado y valor por grupo: se presenta como ítem de formato abierto, y hace referencia a los índices estadísticos que se presentaron en el estudio y su valor. En los casos en los que existía más de un grupo, o más de una medida en el tiempo, se especificaron los valores en cada uno de éstos. No se trata de una categoría mutuamente excluyente, por lo que para un solo estudio se pueden haber marcado varios estadísticos. Cuando no se especificó el dato, se asignó un 8. Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 22. Diferencias estadísticamente significativas entre medidas (de grupos y/o registros): este ítem responde a si las diferencias entre medidas (de varios grupos y/o varios momentos) son estadísticamente significativas o no. Se explicitó en cada caso qué se estaba comparando concretamente.

- a) **No:** se recogieron en este primer apartado los estudios que no mostraron diferencias estadísticamente significativas.
- b) **Sí:** los estudios en los que se encontraron diferencias estadísticamente significativas marcaron en esta categoría.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Cuando se consideró que este ítem no era aplicable por tratarse de un estudio teórico o con una sola medida, se asignó un 9.

Ítem 23. Índice de variabilidad facilitado y valor por grupo: con este ítem de formato abierto se recogió el índice de variabilidad que se mostraba en el estudio y su valor concreto. Cuando se hallaron varios grupos o medidas con este dato especificado, se recogieron todos ellos por separado. Cuando no se especificó el dato, se asignó un 8.

Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

CARACTERÍSTICAS SUSTANTIVAS

Ítem 24. N° de participantes por grupo: hace referencia al número de personas que conformaban un grupo de tal manera que, al sumar todos los componentes de todos los grupos, el valor obtenido coincidirá con el tamaño de la muestra. Se recogió el valor concreto y, posteriormente, se crearon intervalos para agrupar dichos valores. Además, se asignó un 8 en los casos en que no se especificaba el dato y si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 25. N° de grupos en el estudio: en este ítem se determinó el número de grupos que aparecían en el estudio. Al igual que en el ítem anterior, se recogió el valor concreto y, posteriormente, se crearon intervalos para agrupar dichos valores. Se asignó un 8 en los casos en que no se especificaba el dato. Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 26. Exclusiones tras medidas posteriores: en este ítem se constató si alguna persona era excluida del estudio una vez iniciado éste con, al menos, una medida ya tomada.

- a) **No:** Se incluyeron aquellos casos en los que no se excluyó ninguna persona del estudio en ningún momento una vez iniciada la recogida de información.
- b) **Sí:** En esta categoría se incluyeron aquellos casos en los que se daban exclusiones a partir de la segunda medida.
- c) Se asignó un 8 en los casos en que no se especificaba el dato.
- d) Se otorgó un 9 cuando se consideró un ítem no aplicable porque fuera un estudio teórico o porque se recogiera una única medida.

Ítem 27. Rango de edad especificado: este apartado se centra en una característica de la muestra; concretamente, el rango de edad de los participantes.

- a) **No:** En la primera categoría, se incluyeron aquellos estudios en los que no se especifica el rango de edad de los participantes.
- b) **Sí:** En esta categoría, se incluyeron aquellos estudios en los que sí se especificó rango de edad. Se anotó además el valor concreto.
- c) Se otorgó un 9 cuando se consideró un ítem no aplicable porque fuera un estudio teórico o porque se recogiera una única medida.

Ítem 28. Media de edad (valor concreto): al igual que el anterior, este ítem también hace referencia a una característica de la muestra, concretamente la media de edad de

los participantes en el estudio. Se anotó el dato concreto y posteriormente se crearon rangos con los que se aunaron los resultados en grupos. Cuando no se dio el dato, se asignó un 8; si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 29. Periodo de estudio: este ítem hace referencia a la duración total del estudio, desde la primera medida o inicio de la intervención hasta la última medida incluyendo, por tanto, periodo de seguimiento si lo hubiera.

- a) **≤ 6 meses:** el primer apartado recogió a aquellos estudios en los que la duración fue igual o inferior a 6 meses.
- b) **> 6 meses:** en este apartado se incluyeron aquellos estudios cuya duración fue mayor a 6 meses.
- c) Cuando no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 30. Intensidad del tratamiento o del registro cuando no hay intervención: este ítem, de formato abierto, pone en relación el número de horas de duración de la acción formativa con el periodo de tiempo durante el que ésta se llevó a cabo, preferiblemente en semanas si así estuviera especificado, aunque quedó abierto para recoger aquellos valores expresados en otras unidades. Además, cuando no hubo intervención, se recogió información acerca de cuántas medidas, de qué duración, se hacían en qué periodo de tiempo concreto. Cuando no se especificó este dato, se asignó un 8. Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 31. Unidades de intervención o de registro: consiste en especificar si la intervención se realiza a una única persona (de manera individual) o a varias (a un grupo); en caso de no darse intervención, se respondería a si se toma medidas de una sola persona o de un grupo de ellas.

- a) **Individual:** la primera categoría recogió a aquellos casos en los que la intervención o la medida se realizaba a un individuo o a un grupo considerado como una unidad.
- b) **Grupal:** esta categoría recogió a aquellos casos en los que en la intervención participaba un grupo de personas o se tomaban medidas de más de una persona.
- c) En los casos en los que no se especificó el dato, se asignó un 8.
- d) Si no fue posible valorar esta categoría por no ser aplicable, ya sea por tratarse de un estudio teórico o por la naturaleza del objeto de estudio, se asignó un 9.

Ítem 32. Área formativa: este ítem de formato abierto hace referencia a los contenidos sustantivos sobre los que versaba el estudio.

Ítem 33. Campo de intervención: en este apartado de respuesta abierta se hace referencia a los destinatarios de los que hablaba el estudio.

Ítem 34. Tipo de publicación: se recoge, de manera genérica, la fuente de información de la que se tomaron los estudios incluidos. Las diferentes opciones son las siguientes:

- a) **Revista**
- b) **Libro**
- c) **Tesis**
- d) **Congreso**
- e) **Otras publicaciones**
- f) **Trabajos no publicados**

6. DISCUSIÓN.

A continuación, en la tabla 2.6 se resume el procedimiento utilizado en cada una de las fases de elaboración de la escala de medición de la calidad de estudios primarios y sus resultados, hasta llegar a la hasta ahora última versión del instrumento (Chacón, Sanduvete, Sánchez y Sánchez-Meca, 2007b). Las columnas hacen referencia a cada una de las fases realizadas, en orden cronológico, empezando por la de más antigüedad; y las filas diferencian entre los distintos aspectos de la metodología y los resultados.

Y, posteriormente, en la tabla 2.7 se presenta un resumen de la evolución por la que han pasado los distintos ítems hasta llegar a la última versión de la escala de calidad de estudios primarios. Concretamente:

- La primera columna muestra una **etiqueta** general para reconocer el contenido del ítem rápidamente, con pocas palabras.
- En la segunda columna se muestran los ítems que componían el **cuestionario exploratorio** con el que se llevó a cabo el estudio de validez de contenido (Chacón, Sanduvete y Alarcón, 2005).
- En la tercera columna, se presentan los ítems de la versión depurada de la escala, creada a partir de los **resultados del estudio de validez de contenido** y tras la fusión con algunos ítems que generalmente son usados en los **estudios meta-analíticos** (Chacón, Sánchez-Meca, Sanduvete y Alarcón, 2006).
- La cuarta columna muestra los ítems definitivos tras **concretar algunas definiciones, operativizar** más los conceptos y realizar una **adaptación** para generalizar el uso de la escala a cualquier estudio, independientemente del **diseño** que presente (Chacón, Sánchez-Meca y Sanduvete, 2007; Chacón, Sánchez-Meca, Sanduvete y Alarcón, en elaboración).

- La quinta columna muestra una breve explicación de las **razones** por las que se hicieron los cambios, detallados con mayor detenimiento más adelante.

En negrita se marcan las diferencias en un mismo ítem entre los diferentes momentos.

Y tras la presentación de la tabla, se describen los cambios acaecidos, las razones que llevaron a ellos y las consecuencias que sucedieron a dichos cambios.

	1ª FASE. ELABORACIÓN DEL CUESTIONARIO PARA EL ESTUDIO DE VALIDEZ DE CONTENIDO	2ª FASE. ESTUDIO DE VALIDEZ DE CONTENIDO	3ª FASE. ELABORACIÓN DE LA VERSIÓN DEPURADA DE LA ESCALA	4ª FASE. ELABORACIÓN DE LA VERSIÓN INTEGRADORA DE LA ESCALA
MUESTRA	<p>1. Elaboración de la escala: 27 documentos disponibles acerca de la medición de la calidad en estudios primarios.</p> <p>2. Estudio bibliográfico: 1899 resúmenes sobre estudios psicológicos, sociales o de educación.</p>	<p>30 expertos en meta-análisis y revisiones sistemáticas, calidad, evaluación y diseño.</p>	<p>1. Elaboración de la escala: - Los 23 ítems que cumplieron el criterio de inclusión en el estudio de validez de contenido. - Otras escalas disponibles (ej., Sánchez-Meca, 1998)</p> <p>2. Análisis bibliográfico: 95 trabajos completos acerca de formación continua.</p>	<p>Escala depurada (34 ítems)</p>
INSTRUMENTOS	<p>1. Elaboración de la escala: - Bases de datos electrónicas. - Procite.</p> <p>2. Estudio bibliográfico: - Bases de datos electrónicas. - Procite. - SPSS 12.0. - Sistema de categorías (19 ítems).</p>	<p>- Cuestionario exploratorio con 43 ítems, tres opciones de respuesta (de -1 a +1) y 3 conceptos a evaluar (representatividad, utilidad y viabilidad del dato).</p> <p>- Internet para envío y recogida de los datos.</p> <p>- Microsoft Excel para el análisis de datos.</p>	<p>1. Elaboración de la escala: - Ítems que cumplieron el criterio de inclusión. - Ítems de otras escalas.</p> <p>2. Análisis bibliográfico: - Escala, versión depurada. - Bases de datos informatizadas. - Procite. - SPSS 12.0.</p>	<p>Escala depurada (34 ítems)</p>
PROCEDIMIENTO	<p>1. Elaboración de la escala: - Búsqueda de artículos. - Recogida de ítems. - Estructuración en dominios y subdominios.</p> <p>2. Análisis bibliográfico: - Búsqueda de resúmenes. - Elección de los ítems que se codificarían. - Codificación (3 codificadores).</p>	<p>- Selección de la muestra. - Distribución del instrumento. - Análisis de datos.</p>	<p>1. Elaboración de la escala: - Recopilación de ítems con la aplicación del criterio de inclusión. - Recopilación de ítems de otras escalas disponibles. - Comparación entre ambos resultados. - Modificación para crear un instrumento más completo.</p> <p>2. Análisis bibliográfico: - Búsqueda de artículos completos referidos a la formación continua. - Codificación.</p>	<p>Cambios en los términos utilizados y algunas categorías para lograr dos objetivos:</p> <p>- Comparaciones homogéneas entre diseños (la calidad de cualquier tipo de diseño podría oscilar entre 0 y 19).</p> <p>- Mayor concreción y operacionalización.</p>

Capítulo 2. Elaboración de una escala para medir la calidad de estudios primarios

RESULTADOS	<p>1. Cuestionario exploratorio (43 ítems) con tres dimensiones: características extrínsecas, sustantivas y metodológicas.</p> <p>2. Estudio exploratorio: características más frecuentes respecto a la calidad del diseño en los estudios publicados.</p>	<p>Estudio de validez de contenido: cada ítem presentó tres índices que podían oscilar entre -1 y +1 en los tres conceptos estudiados: representatividad, utilidad y viabilidad del dato.</p> <p>Criterio de inclusión: un valor de 0.5 o mayor en al menos dos de los tres conceptos.</p>	<p>1. Escala con 34 ítems.</p> <p>2. Estudio exploratorio: características más frecuentes respecto a la calidad del diseño en los estudios publicados sobre formación continua.</p>	<p>- La nueva escala resultante (35 ítems), instrumento que se usará para conocer las características de calidad que suelen presentar los estudios en formación continua y proponer mejoras (capítulo 3).</p>
-------------------	--	--	---	---

Tabla 2.6. Método y resultados en cada fase de elaboración de la escala de medición de la calidad en estudios primarios.

IDENTIFICADOR	EXPLORATORIO (2005)	V. DEPURADA (2006)	V. INTEGRADORA (2007)	RAZONES DE CAMBIOS
Tipo de estudio	---	---	0. Tipo de estudio: 0-teórico (se describen modelos o no hay datos); 1-observacional; 2-encuesta; 3-cuasi-experimental; 4-experimental	Incorporación. Descriptivo, a cumplimentar a posteriori, ayuda a tener una visión general
Grupo de comparación	---	1. Grupo control: 0-inactivo; 1-activo	1. Grupo de comparación: 0-no hay; 0.5-inactivo; 1-activo o sí hay (en diseños sin intervención)	Incorporación, 1º total y después parcial. 1º, necesidad de determinar la existencia o no de grupo control y el tipo; 2º, adaptación a diseños sin intervención
Criterios de selección de la muestra	21. Criterios de inclusión y exclusión de las unidades de la muestra explicitados: 1-no;2-sí	2. Criterios de selección de la muestra (inclusión/exclusión) (Assignment criteria; en Shadish, Cook y Campbell, 2002b, p.323): 0- no especificados; 1-especificados	2. Criterios de selección de la muestra (inclusión/exclusión) (Assignment criteria; en Shadish, Cook y Campbell, 2002b, p.323): 0- no especificados; 1-especificados	---
Azar	22. Asignación aleatoria de las unidades a los grupos: 1-no y sin control de variables extrañas; 2-no pero con control de variables extrañas; 3-sí	3. Azar: 0- pre-experimentales y cuasiexperimentales sin control de vvee; 0'5- cuasiexperimentales con control de vvee (balanceo, bloqueo, estratificación); 1- experimentales (asignación aleatoria de las unidades a grupos)	3. Azar: 0- experimentales con asignación y/o selección aleatoria inadecuados; pre-experimentales y cuasiexperimentales sin control de vvee; observaciones o encuestas sin selección aleatoria de la muestra: 1-experimentales (selección de la muestra y asignación aleatoria a grupos adecuados); cuasiexperimentales y pre-experimentales con control de vvee (balanceo, bloqueo, estratificación); observaciones o encuestas con selección aleatoria de la muestra	Incorporación parcial: 1º, definición más concreta; 2º asignación + selección y diferentes grados de calidad en cada diseño Consecuencias: - Cualquier tipo de diseño puede puntuar del 0 al 1 en función de su calidad - Anteriormente, sólo se tenía en cuenta el método de asignación; posteriormente, también la selección

Diseño	23. Tipo de metodología/ diseño: 1-pre-experimental (sólo un grupo; una medida); 2-cuasi-experimental (dos grupos sin asignación aleatoria o con grupo control no equivalente con pre-test y post-test; 3-experimental (aleatorio); 4-otros (cuestionarios, observación)	4. Tipo de metodología/ diseño: 0-pre-experimental; 0.5-cuasi-experimental; 0.75-serie temporal (obs pre ≥ 30 y post ≥ 30) y discontinuidad en regresión; 1-experimental-aleatorio	4. Diseño: 0-observaciones o encuestas con una o dos medidas; pre-experimentales; experimentales con un solo momento de medida ; 0.5-Pre-post con grupo control no equivalente con medidas entre 2 y 29: observaciones o encuestas con medidas entre 3 y 29 ; 0.75-series temporales; 1- observaciones o encuestas con 30 ó más medidas; discontinuidad en la regresión ; experimentales con al menos dos momentos de medida	Incorporación parcial en cada una de las opciones. Consecuencias: - Cualquier tipo de diseño puede puntuar del 0 al 1 en función de tres criterios: 1. Si se conoce el método de asignación 2. Número de medidas 3. Número de grupos
Muestra	24. Tamaño de la muestra: 1-n ≤ 5 ; 2-5 <n <10; 3-n ≥ 10	5. Muestra: 0 - n<12; 0.5 - n=[12-40]; 1- n>40	5. Muestra: 0 - n<12; 0.5 - n=[12-40]; 1- n>40	Modificación del punto de corte por no resultar informativo: antes, poca sensibilidad; no se encontraba una muestra menor de 5 Consecuencias: Ahora es más informativo y es más fácil obtener un valor medio en este ítem, pero más complicado obtener el valor máximo.
Mortalidad global	26. Mortalidad experimental: 1- $\leq 30\%$; 2->30%	6. Mortalidad global: 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	6. Mortalidad global: 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	Modificación del punto de corte: antes, poca sensibilidad; no se encontraba una mortalidad mayor de 30 Consecuencias: Mayor concreción y más informativo
Mortalidad diferencial	28. Mortalidad experimental entre grupos: 1- homogéneo; 2- no homogéneo	7. Mortalidad diferencial: 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	7. Mortalidad diferencial: 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	Modificación , por creación del punto de corte. Consecuencia: mayor concreción y operacionalización
Exclusiones posteriores	29. Exclusiones tras la asignación aleatoria (por ejemplo, no codificados): 1- no; 2-sí (número)	8. Exclusiones posteriores a la asignación aleatoria (post-assignment attrition ; en Shadish, Cook y Campbell, 2002b, p.323): 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	8. Exclusiones posteriores a la agrupación de la muestra (post-assignment attrition; en Shadish, Cook y Campbell, 2002b, p.323): 0- $\geq 20\%$; 0.5 - %=]0-20[; 1- 0%	1º, modificación por creación de puntos de corte; 2º, modificación terminológica Consecuencia: 1º, mayor concreción y operacionalización; 2º, adaptación del ítem a los diseños sin intervención

Seguimiento	31. Periodo de seguimiento: 1 < 5 meses; 2-[6-11] meses; 3- > 12 meses	9. Seguimiento: 0 – no se da seguimiento; 0.3< 6 meses; 0.6 – meses = [6-11]; 1- ≥12 meses	9. Seguimiento: 0 – no se da seguimiento; 0.3< 6 meses; 0.6 – meses = [6-11]; 1- ≥12 meses	Modificación de los puntos de corte. Consecuencias: - En primer lugar, un estudio de 5 meses de seguimiento puntuaría como si no lo tuviera; ahora puntúa 0.6 más - Mayor concreción: se pasó de dos a cuatro posibilidades
Momentos de medida	32. Momentos de medida: 1- post intervención; 2-pre y post intervención	10. Momentos de medida: 0- posterior; 1- previo y posterior	10. Momentos de medida: 0- sólo posterior o una medida cuando no hay intervención; 1- previo y posterior o más de una medida si no hay intervención	Incorporación parcial. Consecuencia: adaptación del ítem a los diseños sin intervención
Medidas que aparecen en los distintos momentos de registro	33. Las medidas del pre-test aparecen en el pos-test: 1- ninguna; 2-algunas; 3-todas	11. Las medidas del pre-test aparecen en el pos-test: 0- falta más de uno; 0.5- falta 1; 1- todas aparecen en todos los momentos	11. Medidas que aparecen en todos los momentos de registro : 0->1 no aparece en todos los momentos; 0.5-1 no aparece en todos los momentos de registro (siempre que se mida más de una variable); 1-todas aparecen en todos los momentos	1º, modificación parcial de las categorías; 2º, modificación terminológica; 3º, ampliación de la explicación de la 2ª categoría. Consecuencia: 1º, más concreción y operatividad; 2º, adaptación del ítem a los diseños sin intervención; 3º, atención especial cuando sólo se mide una variable
Variables dependientes normalizadas	34. Variables dependientes normalizadas: 1-no hay (autoinformes y medidas a posteriori); 2-cuestionarios o autoinformes estandarizados; 3-al menos una es objetiva (medidas psicofisiológicas)	12. Variables dependientes normalizadas: 0- sólo autoinformes sin estandarizar; 0.5-ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado; 1-al menos un instrumento es objetivo o normalizado (p.ej. registro fisiológico, pruebas baremadas)	12. Variables dependientes normalizadas: 0- sólo autoinformes sin estandarizar; 0.5-ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado; 1- al menos un instrumento es objetivo o normalizado (p.ej. registro fisiológico, pruebas baremadas)	---
Enmascaramiento del evaluador	---	13. Enmascaramiento del evaluador: 0-no; 1-sí	13. Enmascaramiento del evaluador: 0-no; 1-sí	Incorporación total. Consecuencias: mayor concreción: diferenciación entre quienes realizan la intervención y quienes evalúan

Capítulo 2. Elaboración de una escala para medir la calidad de estudios primarios

Enmascaramiento del usuario	36. Técnicas de control: 1- ciego (de beneficiarios); 2- ciego (de implementadores); 3-doble ciego (ambos); 4-otros (especificar)	14. Enmascaramiento del usuario: 0-no; 1-sí	14. Enmascaramiento del usuario: 0-no; 1-sí	Modificación por desagregación. Consecuencias: un solo ítem pasó a dos
Enmascaramiento del profesional	---	15. Enmascaramiento del profesional que realiza la intervención: 0-no; 1-sí	15. Enmascaramiento del profesional que realiza la intervención (del evaluador interno cuando no hay intervención): 0-no; 1-sí	1º, modificación por desagregación; 2º, incorporación parcial para generalizar su uso a todo tipo de diseño Consecuencias: 1º, un solo ítem pasó a dos; 2º, adaptación para su uso en todo tipo de diseños
Homogeneidad	35. Homogeneidad de la intervención/estudio: 1-las personas no reciben el tratamiento en las mismas condiciones contextuales; 2- las personas reciben el tratamiento en las mismas condiciones contextuales	16. Homogeneidad de la intervención: 0- no todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales ; 1- todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales	16. Homogeneidad de la intervención o proceso de registro: 0- no todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales o proceso de medición cuando no hay intervención ; 1- todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales o proceso de medición cuando no hay intervención	Incorporación, 1º parcial para definición más concreta y 2º incorporación parcial para generalizar el uso también a los diseños sin intervención Consecuencias: - Mayor especificación del constructo y mayor operacionalización. - Adaptación del ítem a los diseños sin intervención
Definición del constructo	37. Definición del constructo del resultado: 1-replicable por el lector en su propio contexto; 2-definición vaga; 3-no definición	17. Definición del constructo: 0-no; 0.5-no operativa; 1-operativa	17. Definición del/los constructo/s (variable/s dependiente/s e independiente/s cuando hay intervención): 0-ninguno definido conceptual y/o empíricamente; 0.5-sí, al menos uno definido conceptual y/o empíricamente; 1-todos definidos conceptual y empíricamente	2 modificaciones para mayor especificación. Consecuencias: mayor especificación y operacionalización
Inferencia de valores perdidos	38. Métodos estadísticos para tratar los valores perdidos: 1-no; 2-sí	18. Métodos estadísticos para inferir los valores perdidos: 0-no; 1-sí	18. Métodos estadísticos para inferir los valores perdidos: 0-no; 1-sí	---

Innovaciones metodológicas en la evaluación de la formación continua

Tamaño de efecto y valor	40. Tamaño de efecto y valor	19. Tamaño de efecto y valor: 0-no; 1-sí	19. Tamaño de efecto y valor: 0-no; 1-sí	Incorporación parcial: especificación de dos opciones para que pueda puntuar entre 0 y 1 Consecuencias: inclusión para el cálculo del índice de calidad
Índice de calidad	---	20. Índice de calidad: suma de puntuaciones: 0-19	20. Índice de calidad: suma de puntuaciones: 0-19	Incorporación total: índice cuantitativo que determina la calidad de cada estudio Consecuencia: información cuantitativa integradora
Índice estadístico	---	21. Índice estadístico calculado	21. Índice estadístico calculado	Incorporación total: para estudiar la posibilidad de integración en meta-análisis Consecuencia: recogida de dato relevante para meta-análisis
Diferencias significativas	---	22. Los resultados mostraron diferencias estadísticamente significativas: 0-no; 1-sí	22. Las medidas mostraron diferencias estadísticamente significativas: 0-no; 1-sí	1º, incorporación total: descriptivo para estudiar la posibilidad del “sesgo de publicación” de cara a realizar un meta-análisis; y 2º, adaptación del ítem a los diseños sin intervención Consecuencia: recogida de dato relevante para meta-análisis
Índice de variabilidad	---	23. Índice de variabilidad facilitado	23. Índice de variabilidad facilitado	Incorporación total: para estudiar la posibilidad de integración en meta-análisis Consecuencia: recogida de dato relevante para meta-análisis
Nº participantes en cada grupo	---	24. Número de participantes en cada grupo	24. Número de participantes en cada grupo	Incorporación total: especificación mayor de la muestra Consecuencia: mayor especificación
Nº grupos	---	25. Número de grupos en el estudio	25. Número de grupos en el estudio	Incorporación total: especificación mayor del diseño Consecuencia: mayor especificación
Exclusiones tras medidas posteriores	---	26. Exclusiones tras medidas posteriores: 0-no; 1-sí	26. Exclusiones tras medidas posteriores: 1-no; 2-sí	Incorporación total: especificación mayor del proceso de recogida de datos Consecuencia: mayor especificación

Capítulo 2. Elaboración de una escala para medir la calidad de estudios primarios

Rango de edad	7. Edad (rango) referido: 1-no; 2-sí	27. Rango de edad especificado: 0-no; 1-sí	27. Rango de edad especificado: 1-no; 2-sí	---
Media de edad	8. Edad (media)	28. Media de edad (valor concreto)	28. Media de edad (valor concreto)	---
Periodo	17. Periodo de tratamiento	29. Periodo de tratamiento: 0.5-≤6 meses; 1->6 meses	29. Periodo de estudio: 1- ≤6 meses; 2->6 meses	1º, incorporación parcial: puntos de corte; 2º, Modificación terminológica Consecuencia: 1º, facilitación de recogida del dato; 2º, adaptación del ítem para su uso con diseños sin intervención
Intensidad	18. Intensidad del tratamiento (p.ej. número de dosis)	30. Intensidad del tratamiento: número de sesiones a la semana	30. Intensidad del tratamiento o del registro cuando no hay intervención: número de sesiones a la semana	Incorporación, 1º parcial para definición más concreta y 2º incorporación parcial para generalizar el uso también a los diseños sin intervención Consecuencias: 1º, mayor operacionalización; 2º, adaptación del ítem a los diseños sin intervención
Unidad	19. Unidades: 1-en grupo; 2-individual	31. Unidades de intervención: 0-individual; 1-grupal	31. Unidades de intervención o de registro: 1-individual; 2-grupal	Incorporación parcial por generalización a diseños sin intervención Consecuencias: Adaptación del ítem a los diseños sin intervención
Área formativa	12. Contexto de intervención: 1-urbano; 2-rural; 3-mixto	32. Área formativa	32. Área formativa	Modificación terminológica y eliminación parcial: se pasó de hablar de “contexto” a “área formativa” y de opción múltiple a formato abierto. Consecuencia: concreción en el ámbito sustantivo y mayor especificación.
Campo de intervención	13. Campo de intervención: 1-sanitario; 2-educacional; 3-social; 4-clínico; 5-organizacional; 6-otros	33. Campo de intervención (destinatarios)	33. Campo de intervención (destinatarios)	Eliminación parcial: se pasó de opción múltiple a formato abierto. Consecuencia: mayor especificación; posibilidad de adaptación a distintos contextos
Tipo de publicación	1. Tipo de publicación: 1-revista, 2-libro, 3-tesis, 4-congreso, 5-otros	34. Tipo de publicación: 1-Revista; 2-Libro; 3-Tesis; 4-Congreso; 5-Otros	34. Tipo de publicación: 1-Revista; 2-Libro; 3-Tesis; 4-Congreso; 5-Otras publicaciones; 6-Trabajos no publicados	Modificación parcial por desagregación: la última opción “otros” fue dividida en dos: “otras publicaciones” y “trabajos no publicados” Consecuencia: mayor especificación

Año de publicación	2. Año de publicación	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Índice de impacto	3. Índice de impacto (en revistas)	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Base de datos	4. Base de datos (especificar)	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Entrenamiento de los investigadores	5. Entrenamiento de los investigadores: 1-especificado; 2-no hay datos suficientes	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Estructura APA	6. Estructura del artículo recomendada por la APA	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables

Capítulo 2. Elaboración de una escala para medir la calidad de estudios primarios

Edad, desviación típica	9. Edad (desviación típica)	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Origen cultural	10. Origen cultural: 1-sólo uno; 2-más de uno; 3-no hay datos suficientes	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Nivel socioeconómico	11. Nivel socioeconómico: 1-bajo; 2-medio; 3-alto	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
País	14. País	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Orientación teórica	15. Orientación teórica: 1-especificada; 2-inferida; 3-no hay datos suficientes	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables

Evidencia empírica previa	16. Evidencia empírica previa: 1-especificada; 2-no hay datos suficientes	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Discusión	20. Los puntos fuertes y débiles son discutidos: 1-no; 2-sí	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Cálculo del tamaño de la muestra	25. Cálculo estadístico del tamaño de la muestra (magnitud del error): 1-sí; 2-no	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Sin mortalidad	27. Sin mortalidad: 1-no; 2-sí	---	---	Fusión con el ítem referido a la “mortalidad experimental” (número 26 del cuestionario exploratorio para el estudio de validez de contenido) Conclusiones: simplificación de la escala por fusión de los ítems que se refieren al mismo constructo
Periodo de línea base	30. Periodo de línea base: 1-< 5 meses; 2- 6-11 meses; 3-> 12 meses	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables

Intervalos de confianza	39. Especificación de los intervalos de confianza en los análisis estadísticos: 1-no; 2-sí	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Efectividad	41. Otros datos además de los objetivos marcados: 1-efectos positivos; 2-efectos negativos; 3-ambos; 4-ninguno	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Interpretación de los resultados	42. Interpretación de los resultados: 1-todos; 2-algunos de ellos; 3-ninguno	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables
Interpretación de los sesgos en los resultados	43. Interpretación de los sesgos en los resultados: 1-todos; 2-algunos de ellos; 3-ninguno	---	---	Eliminación total por incumplimiento del criterio de inclusión (estudio de validez de contenido) Conclusiones: simplificación de la escala por eliminación de los ítems menos representativos, útiles y/o viables

Tabla 2.7. Evolución acaecida hasta llegar a la versión integradora de la escala de calidad.

Como puede apreciarse echando un vistazo a la última columna de la tabla 2.7, las principales razones por las que introdujeron cambios fueron las siguientes:

- **Eliminación:** consistió en la exclusión de ítems o categorías anteriormente presentados. Generalmente, con estas eliminaciones se consiguió hacer de la escala un instrumento más concreto y preciso, en tanto se descartó aquello que no fue considerado representativo, útil y/o viable; en definitiva, aquello que no aportaría información relevante. La eliminación pudo ser:
 - **Total** (de un ítem completo): fueron aquellos casos en los que el ítem fue descartado en su totalidad.
 - **Parcial** (de alguna categoría): se trataron de aquellos casos en que no se eliminó un ítem al completo, sino sólo alguna categoría (el ítem se mantendría, con diferente número de opciones).
- **Incorporación:** supuso incluir ítems o categorías que no aparecían al principio. Generalmente, con esto se consiguió mayor concreción de los constructos en estudio, mayor utilidad al incorporarse aspectos relevantes anteriormente omitidos y, por tanto, lograr un instrumento más completo. La incorporación pudo ser:
 - **Total** (de un ítem completo): se incluyó un ítem completo que no aparecía al principio.
 - **Parcial** (de alguna categoría): se incluyó alguna categoría que anteriormente no se daba.
- **Modificación:** supuso el cambio de algo que ya estaba incluido, ya fuera la redacción de un ítem o de alguna categoría dentro de éste. Generalmente, con esto se logró la adaptación de la escala para su aplicación en todo tipo de diseños (anteriormente, estaba principalmente pensada para su aplicación a diseños con intervención, especialmente experimentales); mayor concreción al definir más claramente y de manera más generalizada los conceptos; en consecuencia, mayor operacionalización; y más utilidad por haberse logrado un uso más generalizado. Fundamentalmente, se dieron tres tipos de modificaciones:
 - **Cambio terminológico:** consistió en el cambio de un concepto por otro, generalmente para lograr la generalización del uso de los ítems en todo tipo de diseño.
 - **Fusión/desagregación:** consistió en unir varios ítems en uno o, al contrario, formar de un solo ítem varios.
 - **Cambio en puntos de corte:** supuso la modificación de los intervalos de algunos ítems cuantitativos, generalmente para lograr una información más útil, a la vez de una codificación sencilla.

A continuación se concreta, para cada ítem que fue modificado, las razones de dicho cambio y las consecuencias que acarrearón.

- En primer lugar, cabe destacar el hecho de que lo que en la escala depurada se codificó como “categoría 9, valor perdido”, en la versión integradora se cambió por dos posibles opciones: “**8**, no se especifica el dato” y “**9**, no aplicable”, con lo que se especificó en mayor grado la causa por la que no pudo recogerse el dato concreto.
- El **ítem 0**, de carácter descriptivo referido al **tipo de estudio**, se incluyó por facilitar el trabajo de codificación tras una primera lectura; para mayor claridad, se crearon diferentes modalidades de la escala de calidad, pensando en utilizar una u otra en función del tipo de estudio (ver anexo VII, pág. xxxvii). En definitiva, la idea de este primer ítem 0 fue la de clasificar cada trabajo según el tipo de diseño tras una primera lectura superficial y, a partir de ahí, decidir qué modalidad de escala utilizar para su codificación.
- El **ítem 1**, referido al **grupo de comparación**, se incluyó después del estudio de validez de contenido, al completar el resultado con otras escalas que previamente no estaban disponibles. Tras su inclusión, se modificó para generalizar su uso en todos los tipos de diseño: en primer lugar, se incluyó una tercera categoría para aquellos casos en que no había grupo de comparación; y, además, se amplió la categoría que puntuaba uno para que se incluyeran también los casos en que no había intervención, ya que en estos casos no es posible distinguir entre un control inactivo o activo, puesto que en ningún grupo se introducen cambios en la vida cotidiana de las personas que son estudiadas.
- Las modificaciones realizadas en el **ítem 3**, referido al **azar**, responden a la necesidad de no considerar los diseños experimentales como los de mayor calidad sin estudiar mínimamente si el proceso de aleatorización fue o no adecuado (Letón y Pedromingo, 2001). Por otro lado, se considera que si no se utiliza azar pero el método de selección y asignación es conocido, el diseño resultante será de mayor calidad que cuando se desconoce. Finalmente, para que los estudios en los que sólo había un grupo también pudieran ser puntuados, se tuvo en cuenta no sólo la asignación, sino también el método de selección de la muestra. De este modo, el resultado fue pasar de tres opciones en que se graduaba la asignación aleatoria a dos opciones, donde se valoraron de igual modo los diseños experimentales con selección y asignación aleatorias adecuados, los pre-experimentales y cuasiexperimentales con control de variables extrañas y los diseños observacionales o de encuestas con selección aleatoria de la muestra.
- El **ítem 4** referido al **diseño**, en la primera versión tenía 4 opciones: 1. Pre-experimental (solo un grupo o una medida); 2. Cuasiexperimental (dos grupos sin asignación aleatoria) o grupos no equivalentes con pretest y posttest; 3. Experimental; aleatorio; y 4. Otros (estudios observacionales, naturalistas y de encuestas). Sin embargo, las 4 opciones elaboradas posteriormente para el mismo ítem fueron distintas: 1. Pre-experimental, 1 ó 2 medidas cuando no hay intervención y experimental con un solo momento de medida; 2. Cuasiexperimental (pre-post con grupo control no equivalente) entre 2 y 29; observaciones y encuestas con medidas entre 3 y 29; 3. Series temporales; 4. Experimental con al menos dos momentos de medida, diseño de discontinuidad en la regresión y observaciones o encuestas con más de 30 medidas. Estas modificaciones se realizaron con la intención de que la calidad del diseño fuera puntuada gradualmente de 0 a 1, independientemente del tipo de diseño que tuviera el estudio. Las **observaciones y**

encuestas fueron consideradas de mayor calidad a medida que aumentaba el número de medidas; este mismo criterio se siguió para graduar la calidad de los **diseños experimentales**; respecto a los **diseños cuasiexperimentales**, los de discontinuidad a la regresión fueron los más valorados, seguidos de los de series temporales, los cuasi (al menos dos momentos con grupo control no equivalente) y, finalmente, los pre-experimentales fueron los valorados en nulo (siguiendo varios criterios, concretamente, el número de grupos, el número de medidas y el conocimiento o no del método de asignación).

- En el **ítem 5**, referido al tamaño de la **muestra**, se mantuvo con tres opciones, pero los puntos de corte variaron tratando así de lograr mayor sensibilidad, más información; anteriormente, los puntos de corte fueron 5 y 10; posteriormente, pasaron a 12 y 40, debido a que en muy pocas ocasiones el número de personas era menor que 5 y en la mayoría de las ocasiones era mayor a 10, por lo que prácticamente todos los casos se encuadraban en la tercera opción (mayor o igual que 10); poniendo el primer punto de corte en 12 y el segundo en 40, se consiguió que los casos se distribuyeran entre los tres intervalos. Realmente, la creación de intervalos óptimos para este ítem variará en función del contexto de intervención que se esté tratando; una posible solución sería anotar los valores concretos de la muestra y crear los intervalos a posteriori.
- En el **ítem 6**, referido a la **mortalidad global**, también se modificaron los puntos de corte con la intención de hacer más informativa la recogida del dato: en muy pocas ocasiones la mortalidad experimental era mayor del 30%, por lo que se bajó el primer punto de corte al 20%; además, se dio la máxima puntuación a aquellas ocasiones en que no había mortalidad experimental. Al igual que se comentó en el ítem anterior, el punto de corte adecuado variará según el contexto de intervención, por lo que un procedimiento a seguir sería anotar el porcentaje concreto en cada caso y crear los intervalos a posteriori.
- En el **ítem 7**, referido a la **mortalidad diferencial**, se pasó de valores dicotómicos (sí o no homogeneidad) a tres intervalos para lograr así una mayor concreción y operacionalización.
- El **ítem 8**, referido a las **exclusiones posteriores**, pasó por dos transformaciones: en primer lugar, hubo un cambio terminológico con la intención de generalizar la aplicación de este ítem a casos en que la agrupación de la muestra no se realizaba aleatoriamente, de tal manera que se pasó de decir “exclusiones posteriores a la asignación aleatoria” a decir “posteriores a la agrupación de la muestra”; por otro lado, se pasó de opciones dicotómicas (sí/no) a tres intervalos, con lo que se aumentó la concreción de la información obtenida.
- El **ítem 9**, referido al **seguimiento** tuvo, al igual que otros ítems anteriores, una modificación en sus puntos de corte para ganar en concreción y especificación; así, por ejemplo, antes de la modificación entraban en la misma categoría un estudio con cuatro meses de seguimiento y otro sin seguimiento; tras la modificación, sin embargo, puntuaría más bajo la segunda situación.
- En el **ítem 10**, referido a los **momentos de medida**, se hizo una inclusión en cada una de las categorías para adaptarlo así a las situaciones en que no se daba

intervención. Así, anteriormente se diferenció entre “posterior” y “previo y posterior”; más tarde, se incluyó a la categoría posterior “o una medida cuando no hay intervención”, y a la categoría de previo y posterior, “o más de una medida si no hay intervención”.

- Al **ítem 11**, referido a las **medidas que aparecen en los distintos momentos de registro**, por un lado se le modificaron algunas opciones para lograr mayor concreción; así, por ejemplo, se eliminaron términos poco específicos como “algunos” por valores concretos; por otro lado, se le modificó la terminología usada para así adaptarlo a los diseños sin intervención: como hablar de pre-test y pos-test implicaba la existencia de una intervención como punto de separación entre el antes y el después, se optó por cambiar esta nomenclatura por “medidas que aparecen en todos los momentos de registro”, con lo que ya no necesariamente tenía que haber intervención; por último, se especificó que un estudio puntuaría en la categoría intermedia siempre y cuando la variable que no aparece en todos los momentos no fuera la única, en cuyo caso puntuaría “0”.
- El **ítem 13**, que hace referencia al **enmascaramiento del evaluador**, se incorporó tras el estudio de validez de contenido pues, proveniente de escalas que previamente no se encontraron disponibles, fue considerado de interés ya que recogía información acerca del enmascaramiento del evaluador, diferenciando así a esta figura de quien realiza la intervención; en ocasiones, ambas funciones pueden recaer en la misma persona o grupo de personas, pero no siempre tiene que ser así.
- Los **ítems 14 y 15**, referidos al **enmascaramiento del usuario y del profesional** respectivamente, previamente estaban fusionados en uno solo. De este modo, se hizo un estudio diferenciado del enmascaramiento de estas dos figuras y del evaluador, previamente comentado. Además, al ítem 15 se le añadió la posibilidad de codificar el enmascaramiento del evaluador interno cuando no había intervención, haciendo así que el ítem fuera codificable ante cualquier tipo de diseño.
- El **ítem 16**, referido a la **homogeneidad**, sufrió dos modificaciones en dos momentos distintos: en primer lugar, se incluyó aquellos aspectos en los que era preciso fijarse para concluir si el estudio era homogéneo para todas las personas participantes, concretamente la intensidad, duración y profesionales, con lo que se ganó en operacionalización; en segundo lugar, para lograr la aplicabilidad en todos los diseños, se concretó que la homogeneidad podía hacer referencia también al proceso de registro, con lo que se logró adaptar el ítem a las situaciones en las que no se daba intervención.
- El **ítem 17**, referido a la **definición del constructo**, pasó por dos modificaciones para tratar de conseguir mayor especificación y operacionalización. Así, se pasó de explicitar si se definía el constructo y si esta definición era o no vaga a determinar si se definía y si era o no operativa, para finalmente distinguir entre la definición conceptual y la empírica y, a partir de ahí, hacer algunas posibles combinaciones.
- En el **ítem 19**, referido al **tamaño de efecto**, en principio sólo se recogía el valor concreto; posteriormente, además, se codificó de manera dicotómica (si se daba o

no), para tener la posibilidad de contabilizar el resultado de 0 a 1 y así poder incluirlo en el cálculo del índice de calidad, ítem que a continuación se comenta.

- El **ítem 20**, referido al **índice de calidad**, se incluyó después de obtener los resultados del estudio de validez de contenido con la intención de calcular un índice cuantitativo que diera una visión general del nivel de calidad que aportaba cada estudio.

Es preciso recordar que los ítems que a partir de ahora se presentan son únicamente descriptivos; no forman parte del cálculo del índice de calidad, pero aportan información interesante a nivel sustantivo. Hay que tener en cuenta, por tanto, que en función del problema que se esté tratando, quizá podrían eliminarse ítems que aquí aparecen e incluirse otros que no se encuentran. Sea como sea, a continuación se siguen comentando los cambios que presentaron cada uno de los ítems y las razones de tales modificaciones.

- Los **ítems 21 y 23**, referidos al **índice estadístico calculado** y al **índice de variabilidad** respectivamente, se aportaron después de haber realizado el estudio de validez de contenido con la intención de ir tanteando qué estudios podrían ser incluidos en un meta-análisis y cuantos de ellos no eran útiles para ello por carecer de datos.
- El **ítem 22**, referido a la **significación de las diferencias**, se incluyó de igual manera después del estudio de validez de contenido para, de cara a la posibilidad de realizar un meta-análisis, determinar si el “sesgo de publicación” puede estar influyendo en los resultados. Este sesgo puede darse porque actualmente se tiende a publicar aquellos estudios que obtuvieron resultados significativos de tal manera que, al recopilar los estudios disponibles acerca de una temática para la realización de un meta-análisis, es posible que se esté tomando una muestra no significativa, con más resultados positivos de los que en la realidad se dan. Por otra parte, se realizó un cambio terminológico: concretamente, se sustituyó la palabra “resultados” por “medidas”, por si la primera pudiera parecer demasiado dirigida a los diseños con intervención exclusivamente.
- El **ítem 24**, referente al **número de participantes que hay en cada grupo**, se incluyó también después del estudio de validez de contenido para obtener una mayor especificación de la muestra, aspecto imprescindible para estudiar la posibilidad de generalización de los resultados.
- El **ítem 25**, referido al **número de grupos**, también se incluyó a posteriori para recoger datos más concretos acerca del diseño.
- El **ítem 26**, referente a las **exclusiones que se realizaron tras medidas posteriores**, también se incluyeron después para lograr mayor especificación acerca de los datos recogidos.
- El **ítem 29**, referido al **periodo de estudio**, pasó por dos cambios. En primer lugar, se pasó del formato abierto a la inclusión de dos intervalos, con lo que se trató de agilizar su cumplimentación. Como en otras ocasiones, es interesante recordar que la elección del intervalo óptimo dependerá en gran medida del ámbito sustantivo,

por lo que es lógico tomar dicha decisión tras haber recogido la información concreta. En segundo lugar, se modificó la palabra “tratamiento” por “estudio”, para así hacer aplicable este ítem a aquellos casos en los que no se daba intervención.

- El **ítem 30**, referido a la **intensidad**, se cambió en dos ocasiones: en primer lugar, se dio una definición de “intensidad del tratamiento” más acorde a todos los posibles contextos, ya que en el cuestionario con el que se realizó el estudio de validez de contenido se hablaba de “número de dosis”, con una orientación claramente clínica; en segundo lugar, se incluyó la posibilidad de estudiar la “intensidad del registro” para aquellos casos en los que no había “tratamiento”, con lo que se logró generalizar el uso de este ítem a los diseños sin intervención.
- Al **ítem 31**, referido a las **unidades de intervención**, se le añadió en último término la posibilidad de que las unidades fueran de registro para así adaptarlo a cualquier tipo de diseño.
- El **ítem 32**, referido al **contexto**, se cambió en dos sentidos. Por un lado, se especificó que el aspecto que se recogía iba a ser el área formativa. Esta modificación se realizó en busca de una mayor especificación, teniendo en cuenta el ámbito de estudio, por lo que no sería generalizable a otros contextos. Por otro lado, se pasó de un formato de opción múltiple a un formato abierto, con lo que se trató de recoger la respuesta de la manera más concreta posible. Esto mismo ocurrió en el **ítem 33**, referido al **campo de intervención**.
- Al **ítem 34**, referido al **tipo de publicación**, se le incluyó en último término una categoría más. Concretamente, la opción “otros” se desagregó en “otras publicaciones” y “trabajos no publicados”, para distinguir así entre aquellos estudios no publicados y los que sí lo estaban.
- Los siguientes ítems del cuestionario utilizado para la validez de contenido fueron eliminados por no cumplir el criterio de inclusión a partir de los índices obtenidos en el estudio de validez de contenido, ya que no alcanzaron un índice del 0.5 en al menos dos de los tres conceptos estudiados (representatividad, utilidad y viabilidad del dato): **ítem 2** (referido al **año de publicación**), **3** (**índice de impacto**), **4** (**base de datos**), **5** (**entrenamiento de los investigadores**), **6** (**estructura del artículo recomendado por la APA**), **9** (**desviación típica de la edad**), **10** (**origen cultural**), **11** (**nivel socioeconómico**), **14** (**país**), **15** (**orientación teórica**), **16** (**evidencia empírica previa**), **20** (**discusión de puntos fuertes y débiles**), **25** (**cálculo estadístico del tamaño de la muestra**), **27** (**sin mortalidad**), **30** (**periodo de línea base**), **39** (**intervalos de confianza de los análisis estadísticos**), **41** (**otros datos además de los objetivos marcados**), **42** (**interpretación de los resultados**) y **43** (**interpretación de los sesgos en los resultados**).

7. CONCLUSIONES.

Según estudios bibliográficos realizados acerca de la temática que aquí nos ocupa, los programas de formación continua suelen presentar algunas características que podrían estar afectando negativamente a la calidad de su diseño, como puede ser la baja

especificación de cómo se realizan dichos programas o la poca operacionalización y concreción de aquello en lo que se trata de intervenir y del procedimiento que se lleva a cabo.

Para determinar empíricamente en qué medida esto es así, en este capítulo se ha tratado de aportar una escala para medir la calidad de los estudios primarios. A pesar de que son comunes este tipo de instrumentos en la medición de la calidad, esta propuesta presenta algunos **beneficios** con respecto a otras versiones encontradas en la literatura:

- Se pretendió elaborar un instrumento que no sólo aportara un **índice cuantitativo** acerca del nivel de calidad sino que, teniendo en consideración sus distintos **ítems por separado**, aportara ideas sobre qué puntos débiles se pueden estar presentando y cómo podrían ser éstos solventados. En definitiva, se intentó que el instrumento sirviera en un primer momento para evaluar, pero también que fuera útil para mejorar los programas, aportando para ello información concreta acerca de los aspectos en los que incidir para provocar dicha mejora, e incluso proponiendo ideas sobre el sentido en el que intervenir concretamente.
- Se intentó crear un instrumento que fuera **aplicable a cualquier tipo de diseño** por diversas razones, entre las que destacan las siguientes:
 - Generalmente las escalas están optimizadas para ser aplicadas con diseños experimentales, cuando éstos no suelen ser usados en los ámbitos de la psicología, la educación y la intervención social. Creando esta escala aplicable a todos los diseños, se aumentó la **utilidad** del instrumento.
 - Hay que acabar con el mito de que el **diseño**, simplemente por ser experimental, ya presentará mejor índice de calidad que un diseño de otro tipo. La aplicación de esta nueva escala quiso poner en tela de juicio esta idea preconcebida.
 - Realmente, el **límite** entre los distintos tipos de diseño puede ser muy **difuso**; así, por ejemplo, cuando una intervención se prolonga durante un largo periodo de tiempo, ¿no se está pasando a realizar un diseño de baja intervención? Puede que la intervención comenzara siendo un elemento extraño en la vida de las personas participantes pero, pasados unos meses, ¿esta intervención no pasa a formar parte de la vida cotidiana de estas personas? En este sentido, no parece justificable el hecho de puntuar como de mayor calidad unos estudios u otros tomando como criterio el “supuesto diseño” que presentan.
- Se intentó crear un instrumento que fuera **aplicable a todo tipo de estudio, independientemente de la temática sobre la que versara**. Al tratarse de un índice de calidad metodológica, se hizo posible su aplicación a distintos ámbitos, con lo que se logró aportar propuestas de mejora de las intervenciones en distintos ámbitos. Así, por ejemplo, en este capítulo se describió un estudio exploratorio aplicado a los ámbitos psicológico, social y educacional; pero, además, se aplicó el instrumento al ámbito de intervención con personas mayores (Sanduvete, 2004) y, en el próximo capítulo, se presentará su aplicación en formación continua.

El **procedimiento** concreto que se siguió para elaborar esta escala se caracterizó por ser inductivo-deductivo ya que se fueron alternando principalmente dos fuentes de

información: por un lado, la bibliografía referida a la evaluación de la calidad y, por otro lado, los datos obtenidos con cada las distintas versiones de la escala en diferentes contextos de intervención; concretamente, constó de cuatro **fases** que a continuación se resumen:

- En la **primera fase**, se **elaboró un cuestionario** tras recopilar todos los ítems encontrados acerca de la medición de la calidad en estudios primarios en todos los documentos encontrados que versaban sobre esta temática.

En este momento, a modo de estudio piloto, se llevó a cabo la **aplicación de la escala**. Concretamente, se recogieron resúmenes referidos a estudios aplicados en ciencias sociales, psicología y educación. Este estudio, además de dar una visión general del estado de poca especificación y grado medio de control que presentan generalmente las cuestiones psicológicas en los trabajos que se realizan, sirvió para ir definiendo más concretamente los conceptos utilizados y para hacer una exploración acerca del nivel de viabilidad de la información necesaria para responder a cada ítem.

- En la **segunda fase**, se realizó un **estudio de validez de contenido** para determinar cuáles de éstos ítems eran considerados por los expertos como representativos de la dimensión en la que se ubicaba, útiles y cuáles, según su opinión, solían estar disponibles en los informes y publicaciones. De este modo, se hizo una criba para reducir el amplio número de ítems del que constaba el cuestionario a sólo aquellos que resultaban de interés.

La aparente importancia otorgada a las cuestiones metodológicas mostrada en los resultados de este estudio de validez de contenido puede ser contraria a lo que asiduamente se encuentra en la práctica profesional: en la mayoría de las ocasiones, los distintos profesionales desarrollan su actividad en una dinámica tan rápida y cambiante que no tienen los recursos suficientes para dedicar tiempo a los aspectos relacionados con la planificación y el diseño. Ocurre que los problemas a resolver surgen de manera imprevisible y son de tal envergadura que requieren una respuesta rápida. Se ven, de este modo, inmersos en una práctica constante donde hay poca cabida a la reflexión y a la profundización de aspectos tan fundamentales como es la elaboración de un buen diseño previo a la intervención. Esta discrepancia entre los resultados obtenidos en el estudio de validez de contenido de los ítems y la realidad profesional pone de manifiesto la falta de conexión entre el desarrollo de la “academia” y el ejercicio profesional.

De todos modos, la mayor importancia dada a los ítems asignados al dominio de las características metodológicas puede tener una doble explicación: en primer lugar, puede haber habido un sesgo en la muestra de expertos, inclinada a temas metodológicos; pero también es cierto que difícilmente se pueden generalizar elementos de calidad en el terreno de las dimensiones extrínsecas y sustantivas, ya que suelen venir determinadas por el ámbito de intervención concreto (Chacón, Sánchez-Meca, Sanduvete y Alarcón, 2006).

- La **tercera fase** sirvió básicamente para detallar más la definición de los ítems encontrados y para realizar algunas otras inclusiones propuestas por expertos. De

este modo, se trató de que ningún ítem útil para medir la calidad de los estudios primarios quedara sin contemplarse en el instrumento elaborado.

Llegados a este punto, se desarrolló una **nueva aplicación de la escala**, donde se tomaron los textos completos que versaban sobre la formación continua. Los resultados obtenidos siguieron en la línea de lo encontrado en el primer estudio exploratorio realizado con los estudios psicológicos en general (poca especificación y grado medio de control y estandarización). Además, se detectó el sesgo que presentaba el instrumento a favor de los diseños experimentales y se vio que algunos ítems provocaban dudas a la hora de realizar la codificación. Para solventar estas deficiencias, se pasó a la cuarta e integradora versión de la escala.

- La propuesta de versión integradora de la escala para la medición de la calidad en estudios primarios obtenida en esta **cuarta fase** presentó diferencias respecto a la versión anterior que se centraron en la búsqueda de mayor concreción y operacionalización y en la necesidad de que todos los ítems fueran aplicables a los distintos tipos de diseño para que los resultados obtenidos fueran comparables (para que el valor del índice de calidad pudiera oscilar de 0 a 19 indistintamente del tipo de diseño que presentara el estudio).

Esta última versión de la escala no se considera como algo cerrado y finalizado, sino que se seguirá perfilando en el futuro, pues actualmente presenta ciertas **debilidades** que habrían de ser solventadas:

- Aún no se ha probado en profundidad la capacidad de **generalización** del instrumento. Se intentará en el futuro realizar nuevas aplicaciones a programas en distintos ámbitos de intervención y hacer los ajustes necesarios para que su uso sea recomendable en cualquier contexto.
- No se ha comprobado las diferencias existentes entre los resultados encontrados con esta versión de la escala y otras escalas encontradas en la literatura. En el futuro se realizará un estudio de **validez convergente**.

Esta versión integradora de la escala se utiliza para determinar en el próximo capítulo, de manera exploratoria, el grado de calidad metodológica que presentan los estudios relacionados con la formación continua, concretar las características principalmente metodológicas que suelen presentar estos trabajos y, a partir de esta información, proponer mejoras.

CAPÍTULO 3. ESTADO DE LA CUESTIÓN EN EVALUACIÓN DE LA FORMACIÓN CONTINUA: APLICACIÓN DE LA ESCALA DE CALIDAD

1. Introducción.....	139
2. Método.....	139
2.1. Muestra.....	140
2.2. Instrumentos	140
2.3. Procedimiento	140
3. Resultados	141
4. Discusión	160
5. Conclusiones.....	175

CAPÍTULO 3. ESTADO DE LA CUESTIÓN EN EVALUACIÓN DE LA FORMACIÓN CONTINUA: APLICACIÓN DE LA ESCALA DE CALIDAD.

1. INTRODUCCIÓN.

Una vez elaborada una versión integradora de la escala de medición de la calidad de estudios primarios, se pasa a aplicar dicho instrumento a las acciones formativas de formación continua publicadas (Sanduvete, Chacón y Sánchez, 2007; Chacón, Sanduvete, Sánchez y Sánchez-Meca, 2007a).

Como ya se ha comentado, a pesar de que la **evaluación** se considera **fundamental** para determinar si un programa de formación ha sido **eficaz** o no, pocas son las organizaciones que llevan a cabo una valoración con suficiente **control de variables extrañas** y basando su toma de decisiones en datos recogidos y analizados de manera **sistemática** (Moresky, Eliades, Bhimani, Bunney y VanRooyen, 2001; Follete y Beitz, 2003; Trevisan, 2004; Chacón, 2004; Chacón, Sanduvete y Alarcón, 2005; Shadish, Chacón y Sánchez-Meca, 2005; Sanduvete, Chacón, Holgado, Gómez y Sánchez, 2006). Por ello, los **objetivos** que se trataron de alcanzar en este capítulo fueron los siguientes:

1. Aplicar la hasta ahora última versión de la escala de valoración de la **calidad “metodológica”** en programas de formación continua publicados hasta septiembre de 2006 como un paso más en el proceso deductivo-inductivo para la elaboración de dicho instrumento.
2. Vislumbrar de manera exploratoria el grado de calidad que suelen presentar los estudios publicados de formación continua y detectar las características metodológicas más comúnmente presentadas.
3. Proponer **mejoras** a nivel **metodológico** en el diseño y evaluación de programas de formación para fomentar y potenciar la **toma de decisiones** basada en **referentes empíricos**, obtenidos desde la perspectiva del **meta-análisis** y **revisiones sistemáticas**.

La escala resultante sirvió para codificar todos los textos completos encontrados que versaban sobre formación continua, con lo que se realizó el estudio de las características que suelen presentar dichos programas; así, se cubrieron los **objetivos 1 y 2**; en la discusión se proponen las mejoras pertinentes a partir del estudio de los puntos negativos encontrados, con lo que se cubre el **objetivo 3**. Finalmente, en el apartado 5 del presente capítulo se resumen las **conclusiones** más importantes.

2. MÉTODO.

Realmente, la búsqueda bibliográfica que a continuación se presenta es la misma que ya se describió en la tercera fase de la elaboración de la escala para medir la calidad de estudios primarios (apartado 4 del capítulo 2). Básicamente, sólo hay una diferencia: el número de artículos completos encontrados pasó de 95 a 121 pues con la primera búsqueda se limitó a recoger lo disponible a través de Internet en la Universidad de Sevilla y, en esta segunda ocasión, se pasó a buscar físicamente en las hemerotecas de algunas facultades de la Universidad de Sevilla (Económicas y Psicología, entre otros), además de que se accedió a los recursos electrónicos de la Universidad de Barcelona.

A continuación, a modo de recordatorio, se presentan resumidamente los pasos seguidos para la realización de esta búsqueda bibliográfica.

2.1. Muestra.

Para la búsqueda de los artículos se utilizaron las **bases informatizadas** disponibles y a la vez consideradas de interés por la temática de los artículos que contienen. Concretamente, en este caso fueron EBSCO Online, Medline, Serfile, CABHealth, CINAHL, PsycINFO, Econlit, ERIC, MathSci, Current Contents y Humanities Index.

De los **2379** estudios encontrados en las bases de datos utilizadas en la búsqueda acerca de formación continua, tras la lectura de los resúmenes, se incluyeron únicamente aquéllos que versaban sobre formación continua en las organizaciones, desechando la formación a cuidadores no formales, a estudiantes y a desempleados, por ejemplo. Cumplieron este criterio de inclusión **287** trabajos, de los que finalmente se estudiaron **121** por ser los únicos que se consiguieron obtener con texto completo, tal como se ha comentado.

2.2. Instrumentos.

Con el **software Procite-5** se realizó el tratamiento de la información obtenida.

El **software SPSS 14.0** se utilizó para la codificación de los artículos y el análisis de resultados descriptivos, mientras que con **Microsoft Excel** se calcularon los índices de calidad.

El instrumento utilizado para la codificación de los estudios encontrados fue la **versión integradora de la escala para medir la calidad de los estudios primarios** previamente elaborada (ver tabla 2.5 en el capítulo anterior).

2.3. Procedimiento.

Se llevó a cabo la búsqueda de los estudios que versaban sobre formación continua. Las palabras clave utilizadas fueron “training programs”, “evaluation” y “work”. Una vez elaborada la escala definitiva de codificación tal y como se presentó

en el capítulo anterior, se recodificaron los 95 estudios encontrados con este nuevo instrumento y se codificaron por primera vez los 26 textos completos que en el anterior estudio bibliográfico no fueron encontrados.

3. RESULTADOS.

En la tabla 3.1, se muestran los porcentajes que se obtuvieron en las diferentes categorías de cada ítem. En la primera columna, aparece cada uno de los ítems y sus diferentes categorías, señalándose entre paréntesis y en negrita el valor cuantitativo de cada una (sólo de los ítems 1 al 19, pues son los únicos que puntúan para el cálculo del índice de calidad); la segunda columna presenta los porcentajes obtenidos en cada categoría contando con los valores perdidos, tanto por ausencia del dato como por no aplicabilidad, por lo que la N con la que se trabaja es el total de estudios recogidos (121); por último, la tercera columna muestra también el porcentaje referido a cada categoría, pero sin tener en cuenta los valores perdidos; es decir, sólo con los valores válidos. Como la N en este caso irá variando en función del ítem, se presenta entre paréntesis el valor concreto en cada una de las casillas.

ÍTEM	CON V.P. (N=121)	SIN V.P.
CARACTERÍSTICAS METODOLÓGICAS		
0. Tipo de estudio:		(N=121)
- Teórico	35.6%	35.6%
- Observacional	6.6%	6.6%
- Encuesta	15.7%	15.7%
- Cuasiexperimental	34.7%	34.7%
- Experimental	7.4%	7.4%
- Perdidos		
- No se especifica el dato	0%	---
- No aplicable (por el objeto de estudio)	0%	---
1. Grupo de comparación:		(N=75)
▪ No hay grupo de comparación (0)	43.8%	70.6%
▪ Inactivo (0.5)	9.1%	14.7%
▪ Activo (en diseños observacionales y de encuesta, cuando hay grupo) (1)	9.1%	14.7%
▪ Perdidos		
- No especificado	2.4%	---
- No aplicable (por el objeto de estudio o por ser teórico)	35.6%	---
2. Criterios de selección de la muestra (inclusión/exclusión):		(N=78)
▪ No especificados (0)	34.6%	53.8%
▪ Especificados (1)	29.8%	46.2%
▪ Perdidos: no aplicable (por el objeto de estudio o por ser teórico)	35.6%	---
3. Azar:		(N=73)
▪ Sin intervención sin selección aleatoria; pre-experimentales y cuasiexperimentales sin control de vvee; y experimentales con asignación y/o selección aleatoria inadecuados (0)	44.6%	74%
▪ Sin intervención con selección aleatoria; pre-experimentales y cuasiexperimentales con control de variables extrañas; y experimentales con asignación y selección aleatoria adecuados (1)	15.7%	26%
▪ Perdidos		
- No especificado	4.1%	---
- No aplicable (por el objeto de estudio o por ser teórico)	35.6%	---

4. Diseño:		(N=77)
<ul style="list-style-type: none"> ▪ Una o dos medidas sin intervención; pre-experimental; experimental con un momento de medida (0) 	46.3%	72.7%
<ul style="list-style-type: none"> ▪ [3-29] medidas sin intervención; cuasi-experimental con [2-29] medidas (0.5) 	9.1%	14.3%
<ul style="list-style-type: none"> ▪ Serie temporal (0.75) 	0.8%	1.3%
<ul style="list-style-type: none"> ▪ ≥ 30 medidas sin intervención; discontinuidad en la regresión; experimental con ≥ 2 medidas (1) 	7.4%	11.7%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	0.8%	---
	35.6%	---
5. Muestra:		(N=76)
<ul style="list-style-type: none"> ▪ $n < 12$ (0) 	3.3%	5.3%
<ul style="list-style-type: none"> ▪ $n = [12-40]$ (0.5) 	10.7%	17.1%
<ul style="list-style-type: none"> ▪ $n > 40$ (1) 	48.8%	77.6%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	1.6%	---
	35.6%	---
6. Mortalidad global:		(N=44)
<ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) 	19.8%	54.6%
<ul style="list-style-type: none"> ▪ $0 < N < 20\%$ (0.5) 	11.6%	31.8%
<ul style="list-style-type: none"> ▪ 0% (1) 	5%	13.6%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	28%	---
	35.6%	---
7. Mortalidad diferencial:		(N=7)
<ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) 	0.8%	14.3%
<ul style="list-style-type: none"> ▪ $0 < N < 20\%$ (0.5) 	1.7%	28.6%
<ul style="list-style-type: none"> ▪ 0% (1) 	3.3%	57.1%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio; teórico; un solo grupo) 	16.5%	---
	77.7%	---
8. Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones:		(N=24)
<ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) 	6.6%	33.3%
<ul style="list-style-type: none"> ▪ $0 < N < 20\%$ (0.5) 	5.8%	29.2%
<ul style="list-style-type: none"> ▪ 0% (1) 	7.4%	37.5%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio; teórico; un solo grupo) 	2.5%	---
	77.7%	---
9. Seguimiento:		(N=69)
<ul style="list-style-type: none"> ▪ No se da seguimiento (0) 	30.6%	53.6%
<ul style="list-style-type: none"> ▪ < 6 meses (0.3) 	13.2%	23.2%
<ul style="list-style-type: none"> ▪ [6-11] meses (0.6) 	5.8%	10.2%
<ul style="list-style-type: none"> ▪ ≥ 12 meses (1) 	7.4%	13%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	7.4%	---
	35.6%	---
10. Momentos de medida:		(N=79)
<ul style="list-style-type: none"> ▪ Sólo posterior; una medida cuando no hay intervención (0) 	32.2%	50%
<ul style="list-style-type: none"> ▪ Previo y posterior; más de una medida cuando no hay intervención (1) 	32.2%	50%
<ul style="list-style-type: none"> ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	---	---
	35.6%	---
11. Medidas que aparecen en todos los momentos de registro:		(N=36)
<ul style="list-style-type: none"> ▪ Falta más de uno; también cuando falta la única variable que se mide (0) 	0%	0%

Capítulo 3. Estado de la cuestión en evaluación de la formación continua: aplicación de la escala de calidad

<ul style="list-style-type: none"> ▪ Falta 1 (siempre que se mida más de 1 variable) (0.5) ▪ Todas aparecen en todos los momentos (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio; teórico; un solo momento de registro) 	<p>3.3%</p> <p>26.4%</p> <p>2.5%</p> <p>67.8%</p>	<p>11.1%</p> <p>88.9%</p> <p>---</p> <p>---</p>
<p>12. Variables dependientes normalizadas:</p> <ul style="list-style-type: none"> ▪ Sólo autoinformes sin estandarizar (0) ▪ Ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado (0.5) ▪ Al menos uno es objetivo o normalizado (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	<p>0.8%</p> <p>43.8%</p> <p>22.3%</p> <p>1.7%</p> <p>31.4%</p>	<p>(N=81)</p> <p>1.2%</p> <p>65.4%</p> <p>33.4%</p> <p>---</p> <p>---</p>
<p>13. Enmascaramiento del evaluador:</p> <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	<p>2.5%</p> <p>1.7%</p> <p>60.2%</p> <p>35.6%</p>	<p>(N=5)</p> <p>60%</p> <p>40%</p> <p>---</p> <p>---</p>
<p>14. Enmascaramiento del usuario:</p> <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	<p>2.5%</p> <p>1.7%</p> <p>60.2%</p> <p>35.6%</p>	<p>(N=5)</p> <p>60%</p> <p>40%</p> <p>---</p> <p>---</p>
<p>15. Enmascaramiento del formador (como evaluador interno cuando no hay intervención):</p> <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	<p>1.7%</p> <p>1.7%</p> <p>61%</p> <p>35.6%</p>	<p>(N=4)</p> <p>50%</p> <p>50%</p> <p>---</p> <p>---</p>
<p>16. Homogeneidad de la intervención o del proceso de registro cuando no hay intervención:</p> <ul style="list-style-type: none"> ▪ Distinto (0) ▪ Igual (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	<p>7.4%</p> <p>54.5%</p> <p>2.5%</p> <p>35.6%</p>	<p>(N=75)</p> <p>12%</p> <p>88%</p> <p>---</p> <p>---</p>
<p>17. Definición de los constructos:</p> <ul style="list-style-type: none"> ▪ No se define ninguno (0) ▪ Se define al menos uno conceptual y/o empíricamente (0.5) ▪ Todos definidos conceptual y empíricamente (1) 	<p>0.8%</p> <p>93.4%</p> <p>5.8%</p>	<p>0.8%</p> <p>93.4%</p> <p>5.8%</p>
<p>18. Métodos estadísticos para inferir los valores perdidos:</p> <ul style="list-style-type: none"> ▪ Ninguno (0) ▪ Se inferen los valores perdidos (1) ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio, teórico o no se dieron valores perdidos) 	<p>28.9%</p> <p>0.8%</p> <p>31.4%</p> <p>38.9%</p>	<p>(N=36)</p> <p>97.2%</p> <p>2.8%</p> <p>---</p> <p>---</p>
<p>19. Tamaño de efecto:</p> <ul style="list-style-type: none"> ▪ No se especifica (0) ▪ Se especifica (o índices derivables) (1) ▪ Perdidos: no aplicable (por el objeto de estudio, teórico o sólo una medida en un grupo) 	<p>38.9%</p> <p>9.1%</p> <p>52%</p>	<p>(N=58)</p> <p>81%</p> <p>19%</p> <p>---</p>
<p>20. Índice de calidad:</p> <ul style="list-style-type: none"> ▪ [0-3] ▪ [3.1-6] ▪ [6.1-9] 	<p>44.6%</p> <p>29.7%</p> <p>22.3%</p>	<p>(N=121)</p> <p>44.6%</p> <p>29.7%</p> <p>22.3%</p>

<ul style="list-style-type: none"> ▪ [9.1-12] 	3.4%	3.4%
21. Índice estadístico calculado: <ul style="list-style-type: none"> ▪ Datos cualitativos ▪ Descriptivos ▪ Comparación con t/F ▪ Otros métodos de comparación ▪ Fiabilidad ▪ Análisis factorial ▪ Modelo de ecuaciones estructurales ▪ Generalizabilidad ▪ Otros ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	2.3% 22.3% 18.3% 13.3% 2.3% 1.7% 1.7% 1.7% 0.8% --- 35.6%	5.1% 34.2% 26.5% 20.2% 5.1% 2.5% 2.5% 2.5% 1.4% --- ---
22. Los resultados mostraron diferencias estadísticamente significativas: <ul style="list-style-type: none"> ▪ No ▪ Sí ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio; teórico; una sola medida) 	9.1% 26.4% 9.9% 54.6%	(N=43) 25.6% 74.4% --- ---
23. Índice de variabilidad facilitado: <ul style="list-style-type: none"> ▪ Desviación típica ▪ Varianza ▪ Error estándar ▪ Intervalo de confianza ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	24% 1.7% 4.1% 0.8% 33.8% 35.6%	(N=39) 78.4% 5.4% 13.5% 2.7% --- ---
CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo: <ul style="list-style-type: none"> ▪ [0-50] ▪ [51-100] ▪ [101-200] ▪ [201-300] ▪ [301-400] ▪ [401-500] ▪ [501-600] ▪ [601-700] ▪ [701-800] ▪ [801-900] ▪ [901-1000] ▪ [1001-2000] ▪ [2001-3000] ▪ ≥ 3000 ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	21.5% 11.6% 5.8% 5.8% 4.1% 3.3% 0% 1.6% 1.6% 1.6% 0% 1.6% 1.6% 1.6% 2.5% 1.8% 35.6%	(N=76) 34.2% 18.4% 9.2% 9.2% 6.6% 5.3% 0% 2.6% 2.6% 2.6% 0% 2.6% 2.6% 4.1% --- ---
25. Número de grupos en el estudio: <ul style="list-style-type: none"> ▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 6 ▪ 10 ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	42.1% 13.3% 3.3% 4.1% 0.8% 0.8% --- 35.6%	(N=78) 65.4% 20.5% 5.1% 6.4% 1.3% 1.3% --- ---
26. Exclusiones tras medidas posteriores: <ul style="list-style-type: none"> ▪ No 	0.8%	(N=11) 9.1%

Capítulo 3. Estado de la cuestión en evaluación de la formación continua: aplicación de la escala de calidad

<ul style="list-style-type: none"> ▪ Sí ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio; teórico; una sola medida) 	8.3%	90.9%
27. Rango de edad especificado: <ul style="list-style-type: none"> ▪ No ▪ Sí ▪ Perdidos: no aplicable (por el objeto de estudio o teórico) 	42.2%	(N=70) 72.9%
	15.7%	27.1%
	42.1%	---
28. Media de edad: <ul style="list-style-type: none"> ▪ [17-30] ▪ [31-40] ▪ [41-45] ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o teórico) 	2.5%	(N=13) 23.1%
	6.6%	61.5%
	1.6%	15.4%
	47.2%	---
	42.1%	---
29. Periodo de estudio: <ul style="list-style-type: none"> ▪ ≤ 6 meses ▪ > 6 meses ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	33%	(N=71) 56.3%
	25.6%	43.7%
	5.8%	---
	35.6%	---
30. Intensidad del tratamiento/registro <ul style="list-style-type: none"> ▪ 1 vez ▪ 2 horas/semana ▪ 1 sesión/semana ▪ 18 horas/semana ▪ 6 horas/semana ▪ 1 sesión cada 6 semanas ▪ 1 sesión cada 10 días ▪ 32 horas/semana ▪ 1 sesión cada 12 semanas ▪ 40 horas/semana ▪ 2 horas/día ▪ Perdidos <ul style="list-style-type: none"> - No especificado - No aplicable (por el objeto de estudio o por ser teórico) 	19.8%	(N=61) 66.4%
	1.6%	5.6%
	1.6%	5.6%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	0.8%	2.8%
	35%	---
	35.6%	---
31. Unidades de intervención o de registro <ul style="list-style-type: none"> ▪ Individual (o grupo considerado como unidad) ▪ Grupal ▪ Perdidos <ul style="list-style-type: none"> - No aplicable (por el objeto de estudio o por ser teórico) 	2.4%	(N=78) 3.8%
	62%	96.2%
	35.6%	---
32. Área formativa <ul style="list-style-type: none"> ▪ Mejora de aptitudes ▪ Habilidades sociales ▪ Manufacturación ▪ Prevención de riesgos laborales ▪ Cuidados a personas mayores ▪ Ética ▪ Formación ▪ Nuevas tecnologías ▪ Cuidados a niños ▪ Diagnóstico y tratamiento ▪ Cuidados en general ▪ Economía ▪ Orientación escolar ▪ Evaluación de la formación ▪ Relajación ▪ Mejora de la calidad 	2.5%	(N=121) 2.5%
	2.5%	2.5%
	1.7%	1.7%
	9.1%	9.1%
	6.6%	6.6%
	0.8%	0.8%
	25.6%	25.6%
	1.7%	1.7%
	3.3%	3.3%
	5.8%	5.8%
	2.5%	2.5%
	1.7%	1.7%
	3.3%	3.3%
	29.6%	29.6%
	2.5%	2.5%
	0.8%	0.8%

33. Campo de formación: destinatarios		(N=121)
▪ Salud	19%	19%
▪ Sector terciario	2.5%	2.5%
▪ Recursos humanos	10.8%	10.8%
▪ Trabajadores de fábrica	5%	5%
▪ Profesorado	5.8%	5.8%
▪ Todos los profesionales que trabajan con personas mayores	6.6%	6.6%
▪ Formadores	4.1%	4.1%
▪ Contextos variados	21.4%	21.4%
▪ Salud mental	9.9%	9.9%
▪ Policías	1.7%	1.7%
▪ Trabajadores de la construcción	3.3%	3.3%
▪ Servicios sociales	3.3%	3.3%
▪ Empresas	5%	5%
▪ Otros	1.6%	1.6%
34. Tipo de publicación:		(N=121)
▪ Revista	90.1%	90.1%
▪ Libro	5.8%	5.8%
▪ Tesis	0%	0%
▪ Congreso	2.5%	2.5%
▪ Otras publicaciones	1.6%	1.6%
▪ Trabajos no publicados	0%	0%

Tabla 3.1. Resultados descriptivos de los ítems componentes de la escala de calidad.

A continuación, se representan los porcentajes obtenidos en cada uno de los ítems tras eliminar los valores perdidos y se muestra un breve comentario referido a estos resultados.

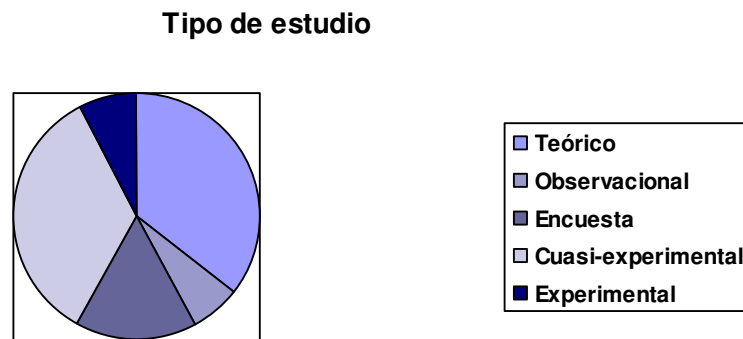


Figura 3.1. Tipo de estudio.

En la mayoría de las ocasiones, los estudios encontrados eran teóricos (35.6%) o cuasiexperimentales (34.7%). Los tipos de estudio menos utilizados fueron los observacionales (6.6%), seguidos a poca diferencia de los experimentales (7.4%).

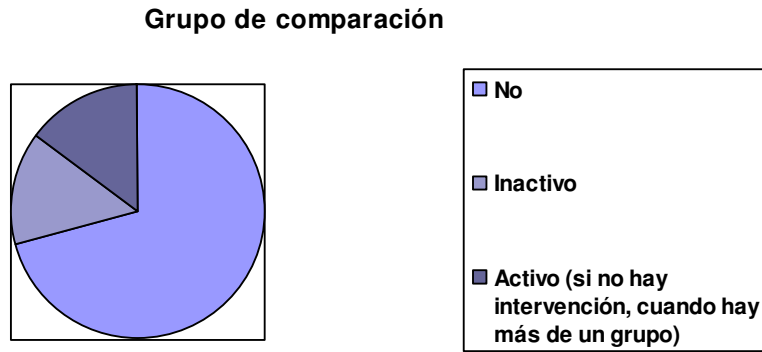


Figura 3.2. Grupo de comparación.

En la mayoría de las ocasiones, los estudios no mostraron grupo de comparación (70.6%). De las veces en que sí lo hubo (29.4%), en la mitad de las ocasiones dicho grupo fue inactivo y en la otra mitad, activo (aunque en este último porcentaje también se incluyeron aquellos casos en los que un estudio sin intervención mostraba como mínimo un segundo grupo).

Criterios de inclusión y exclusión de las unidades de la muestra

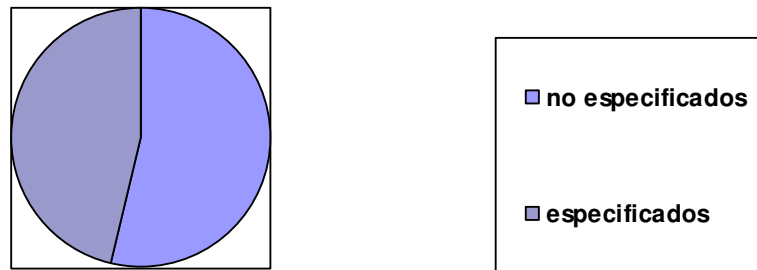


Figura 3.3. Criterios de inclusión y exclusión de las unidades de la muestra.

Los criterios de inclusión y exclusión de la muestra en el estudio se explicitaron únicamente en algo menos de la mitad de las ocasiones (46.2%).

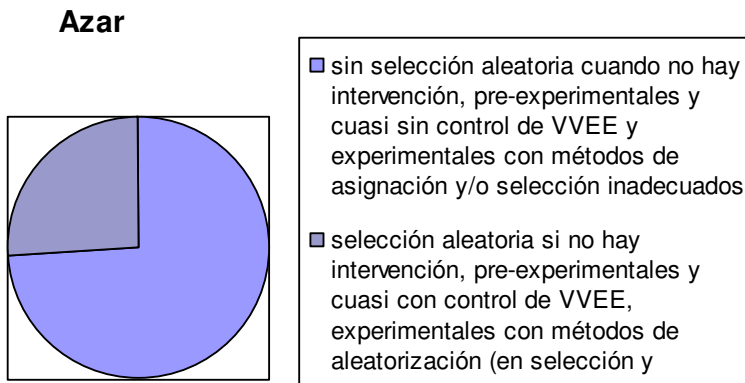


Figura 3.4. Azar.

En aproximadamente tres cuartas partes de las ocasiones (74%), un procedimiento adecuado basado en la aleatorización no fue el método utilizado para la selección de la muestra a estudiar y/o su asignación a los distintos grupos (cuando había más de uno).

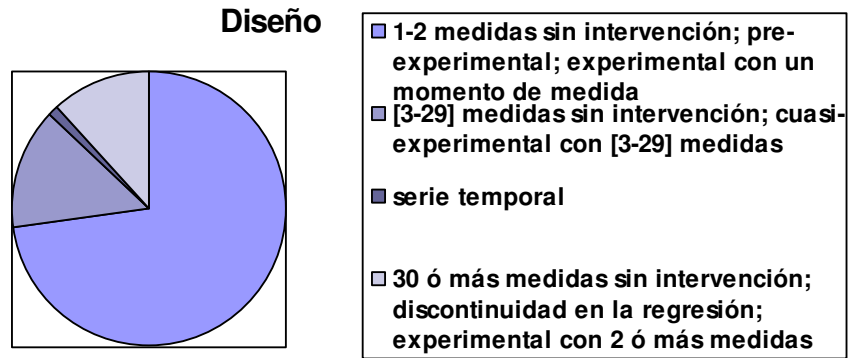


Figura 3.5. Diseño.

En aproximadamente tres cuartos de las ocasiones (72.7%), el diseño que presentaron los estudios tenían un momento de medida cuando se dio intervención y como máximo dos momentos cuando se trataba de estudios observacionales o de encuesta.

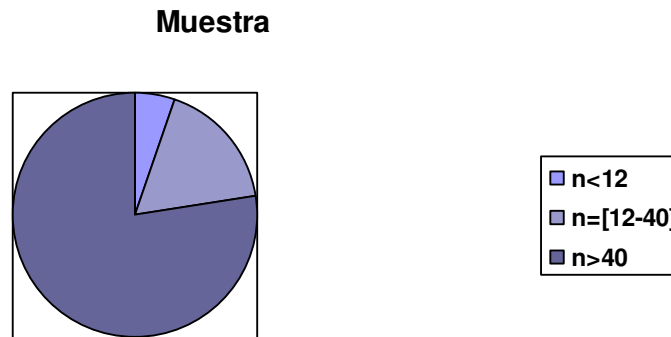


Figura 3.6. Muestra.

En algo más de las tres cuartas partes de las ocasiones (77.6%), el tamaño de la muestra de los estudios superó los 40 sujetos.

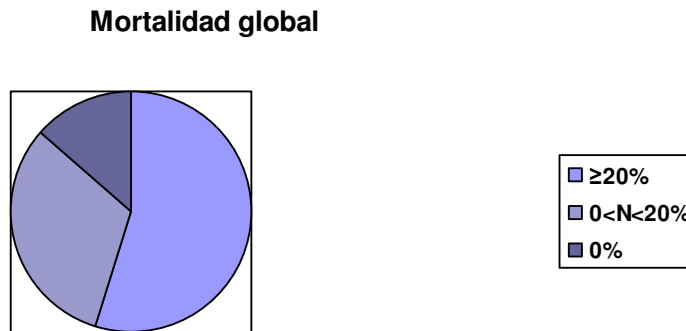


Figura 3.7. Mortalidad global.

En la mayoría de las ocasiones (54.6%), la mortalidad global de los estudios ascendió al 20% o más de la muestra inicial; en el 31.8% de las ocasiones presentó un grado intermedio de mortalidad global, inferior al 20%; y, en el 13.6% de los casos, se explicitó que la mortalidad global era nula. Un aspecto a destacar es el alto número de veces en el que este dato no fue especificado en los estudios (28%).

Mortalidad diferencial

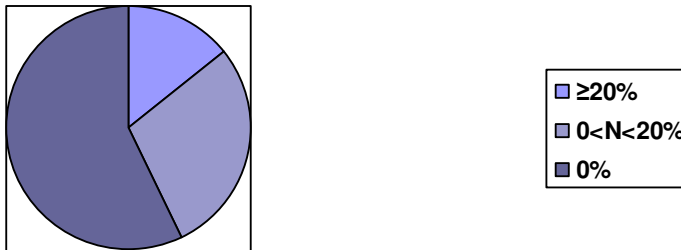


Figura 3.8. Mortalidad diferencial.

En la mayoría de las ocasiones (57.1%), no se dio mortalidad diferencial; en el 28.6% de los casos sí se dio, pero con una diferencia entre ambos grupos menor del 20%; en el 14.3% de los casos, dicha diferencia fue mayor al 20%.

Hay que tener en cuenta que estos porcentajes pueden no ser significativos, ya que sólo 7 estudios presentaron este dato: en el 77.7% de las ocasiones, este ítem resultó ser no aplicable por tratarse de estudios teóricos, con un solo grupo o por la casuística del objeto de estudio; y, además, en el 16.5% de las ocasiones, a pesar de ser un ítem aplicable, el dato no fue especificado.

Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones

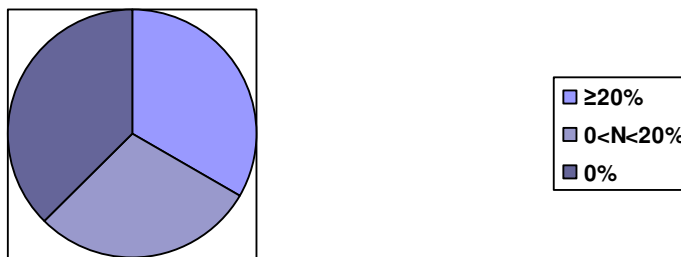


Figura 3.9. Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones.

Lo primero a destacar respecto a este ítem es que presentó un alto porcentaje de valores perdidos debido a que en el 77.7% de las ocasiones no era aplicable por tratarse de estudios teóricos, con un solo grupo o por la naturaleza del objeto de estudio (por lo que no había distintas condiciones en las que agrupar a la muestra).

De los 24 estudios que aportaron datos, los resultados en las distintas opciones quedó bastante igualado, ya que en el 37.5% de las ocasiones no hubo ninguna exclusión después de la agrupación; en el 33.3% la exclusión fue igual o superior al 20% de la muestra; y en el 29.2% de las ocasiones, este dato aportó un valor intermedio entre más del 0% y menos del 20%.

Seguimiento

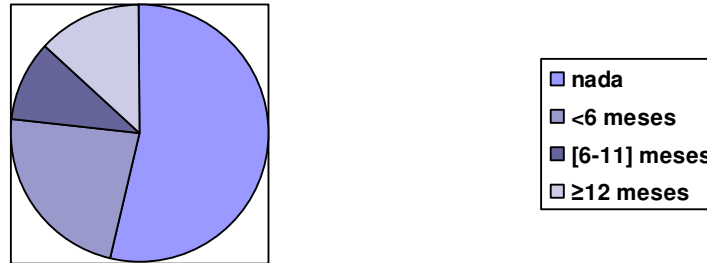


Figura 3.10. Seguimiento.

En algo más de la mitad de las ocasiones (53.6%), el estudio no presentaba periodo de seguimiento; de las veces que lo presentaba, en el 23.2% tenía una durabilidad menor a 6 meses, en el 10.2% de las ocasiones duraba entre 6 y 11 meses (ambos incluidos) y, en el restante 13%, la duración fue de un año o más.

Momentos de medida



Figura 3.11. Momentos de medida.

En la mitad de los estudios presentados, se tomaban medidas sólo tras la intervención y, en caso de los estudios de observación y encuesta (sin intervención), se tomaban más de una medida. En la otra mitad se tomaban medidas antes y después de la intervención o, en caso de tratarse de estudios observacionales o de encuesta, se recogía más de una medida.

Medidas que aparecen en todos los momentos de registro

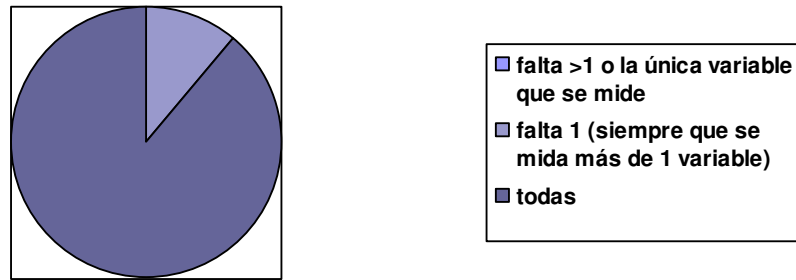


Figura 3.12. Medidas que aparecen en todos los momentos de registro.

En la gran mayoría de las ocasiones (88.9%), todas las medidas tomadas en un primer momento siguen recogiendo a lo largo de los restantes momentos de medida.

Variables dependientes normalizadas

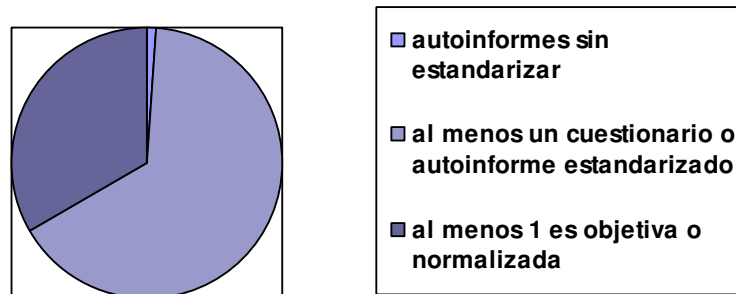


Figura 3.13. Variables dependientes normalizadas.

En la mayoría de las ocasiones, concretamente en el 65.4%, ninguna variable dependiente es medida a través de instrumentos objetivos o normalizados, aunque al menos una se mide con cuestionarios o autoinformes estandarizados. Casi todos los demás casos (33.4%) presentan al menos una variable dependiente medida de manera objetiva o normalizada.

Enmascaramiento del evaluador

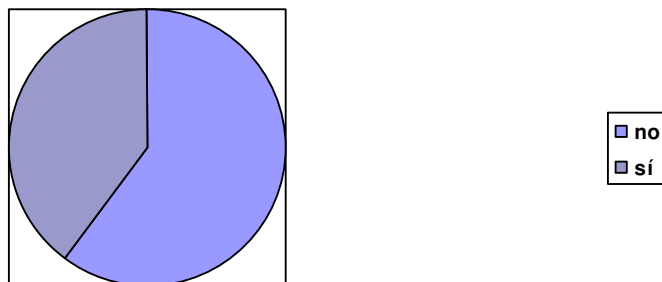


Figura 3.14. Enmascaramiento del evaluador.

Este ítem, referido al enmascaramiento por parte del evaluador, sólo fue recogido en cinco estudios, mostrando un porcentaje de pérdida del dato por falta de especificación del 60.2%, además de no aplicabilidad en el 35.6% de los casos. De los únicos cinco casos en los que se explicitó, en el 60% de las ocasiones no se daba enmascaramiento y en el restante 40% sí se daba.

Enmascaramiento del usuario

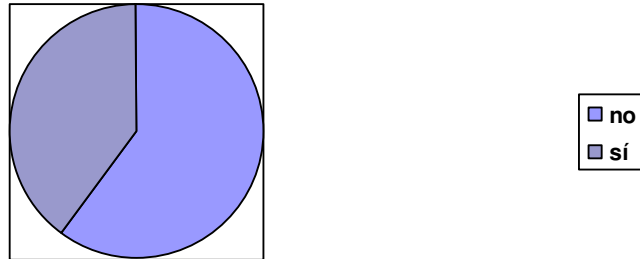


Figura 3.15. Enmascaramiento del usuario.

Este ítem, referido al enmascaramiento del usuario, presentó un perfil semejante al anterior, con altos porcentajes de valores perdidos (60.2% por no especificarse el dato y 35.6% por no ser aplicable). De las cinco ocasiones en las que se recogió el dato, en el 60% no se dio enmascaramiento y en el restante 40% sí.

Enmascaramiento del formador (como evaluador interno cuando no hay intervención)

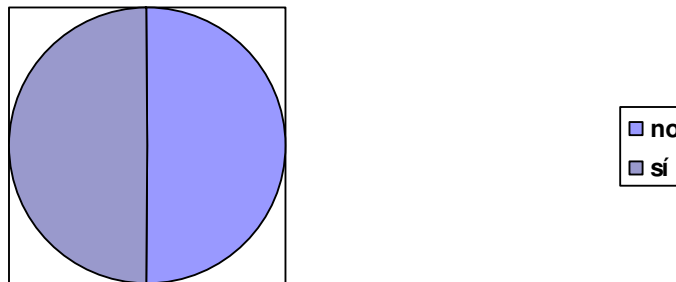


Figura 3.16. Enmascaramiento del formador (como evaluador interno cuando no hay intervención).

Este ítem, referido al enmascaramiento del formador o del evaluador interno cuando no existe intervención, se asemeja en sus resultados a los dos ítems anteriores, referidos al enmascaramiento del evaluador y del usuario: existe un alto porcentaje de datos perdidos (el 61% de no especificación y el 35.6% de no aplicabilidad); sólo cuatro casos presentaron la información solicitada, de los que dos de ellos mostraron enmascaramiento y los otros dos no.

Homogeneidad de la intervención o del proceso de registro cuando no hay intervención

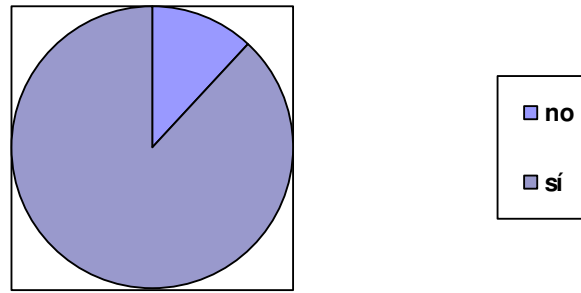


Figura 3.17. Homogeneidad de la intervención o del proceso de registro cuando no hay intervención.

En la mayoría de las ocasiones (88%), los estudios presentaron homogeneidad en la manera de llevar a cabo la intervención con los distintos usuarios respecto a intensidad, duración y profesionales que la desempeñaron, o en el proceso de recogida de datos cuando no hubo intervención.

Definición de los constructos

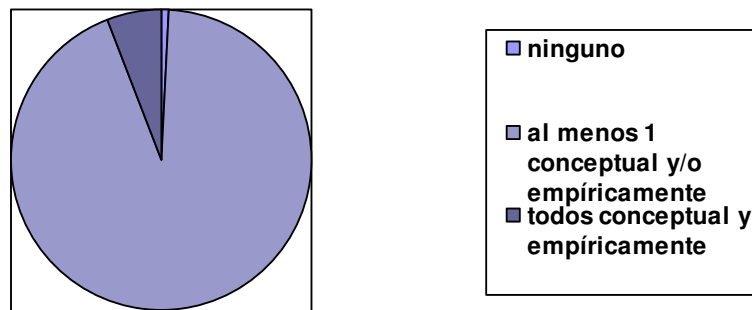


Figura 3. 18. Definición de los constructos.

Este ítem no presentó valores perdidos; pudo recogerse información de todos los estudios analizados. Concretamente, en la inmensa mayoría de las ocasiones (93.4%), al menos uno de los constructos era definido conceptual y/o empíricamente. En muy raras ocasiones, no se definía ninguno (0.8%) o se definían todos conceptual y empíricamente (5.8%).

Inferencia de valores perdidos

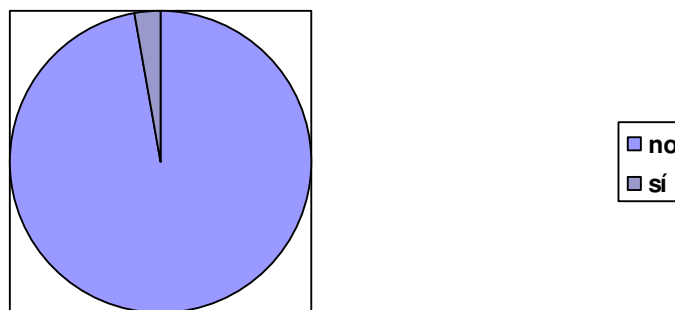


Figura 3.19. Métodos estadísticos para inferir los valores perdidos.

En la inmensa mayoría de las ocasiones en que se pudo recoger el dato (97.2%), no se utilizó ningún método estadístico para inferir los valores perdidos. En el 31.4% de los casos, esta información no se explicitaba; en el 38.9% de las ocasiones, no fue aplicable por tratarse de un estudio teórico, por no tener valores perdidos o por la casuística del objeto de estudio.

Tamaño de efecto especificado

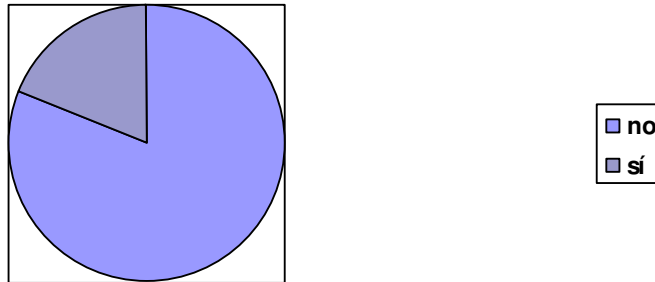


Figura 3.20. Tamaño de efecto especificado.

En el 52% de las ocasiones, el ítem referido a la especificación del tamaño de efecto resultó no ser aplicable por tratarse de estudios teóricos, por no haber posibilidad de comparación al tratarse de una sola medida en un solo grupo o por la naturaleza del objeto de estudio.

De las veces en que sí era posible recoger el dato, en la mayoría de las ocasiones no se especificaba (81%); en el restante 19% sí se hacía o, al menos, se aportaban datos con los que se podía calcular el tamaño de efecto.

Índice de calidad

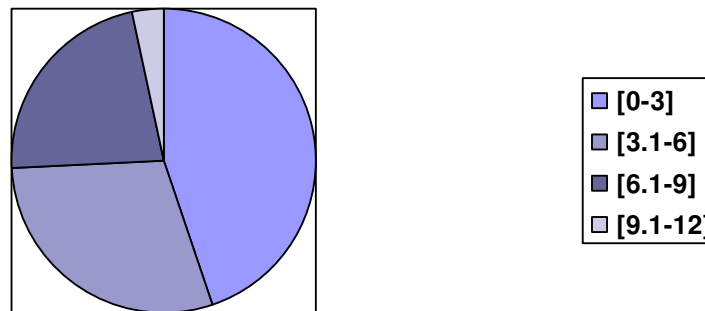


Figura 3.21. Índice de calidad.

Respecto al índice de calidad se aprecia que, a medida que el valor va siendo mayor, menos estudios se encuadran en el intervalo. Así, la mayoría de los estudios (44.6%) obtuvieron puntuaciones entre 0 y 3, ambos valores incluidos; el 29.7% llegó a obtener valores entre 3.1 y 6; el 22.3% obtuvo valores entre 6.1 y 9; y sólo el 3.4% de los estudios llegó a obtener entre 9.1 y 12 puntos; ningún estudio alcanzó una puntuación superior a 12.

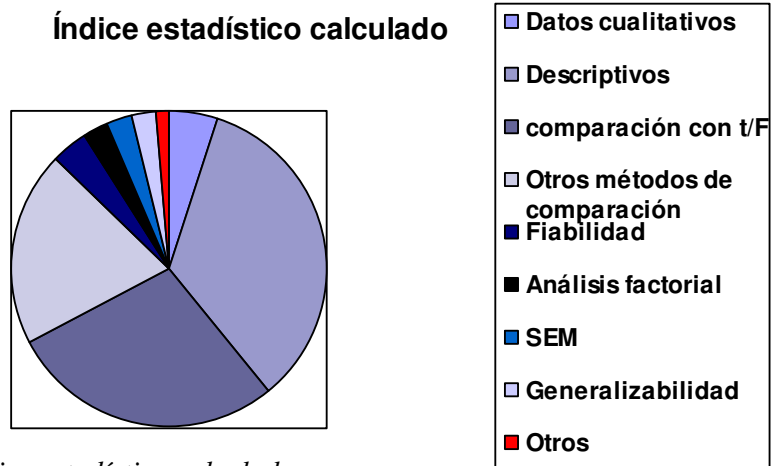


Figura 3.22. Índice estadístico calculado.

Hay gran disparidad de índices estadísticos utilizados para mostrar los resultados de los estudios. Los más frecuentes fueron los índices descriptivos (34.2%), seguidos de los métodos de comparación de medias (la t o la F en un 26.5% y otros métodos de comparación en un 20.2%).

Diferencias estadísticamente significativas

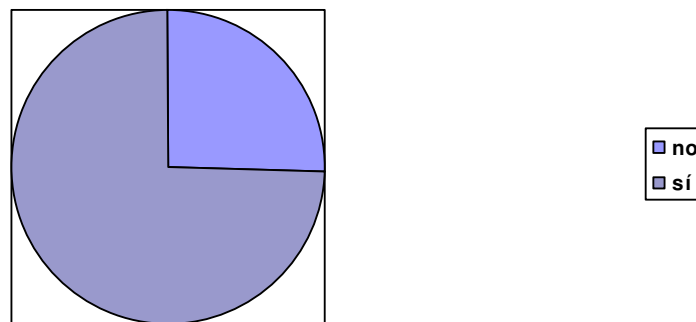


Figura 3.23. Diferencias estadísticamente significativas.

En el 54.6% de las ocasiones, el ítem no fue aplicable por ser teórico, presentar sólo una medida (hecho que imposibilitaba cualquier comparación) o por la naturaleza del objeto de estudio. De los trabajos en los que se pudo aplicar, en casi el 75% de las ocasiones los resultados mostraron diferencias estadísticamente significativas.

Índice de variabilidad

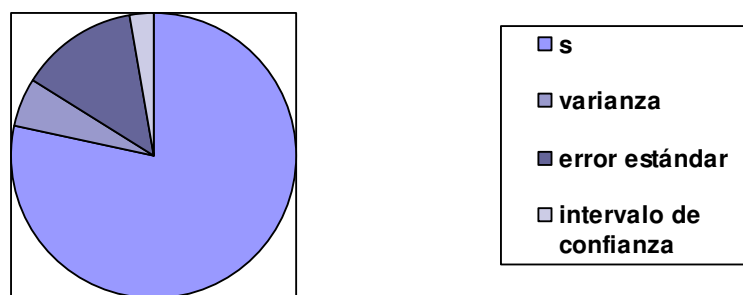


Figura 3.24. Índice de variabilidad facilitado.

Mayoritariamente, en las ocasiones en que se aplicó este ítem (concretamente el 78.4%), el índice de variabilidad que se dio fue la desviación típica. Es interesante destacar el alto porcentaje de ocasiones en las que no se especificó este dato (en el 33.8% del total).

Número de participantes en cada grupo

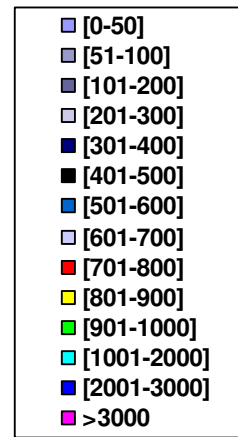
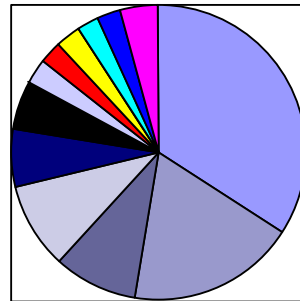


Figura 3.25. Número de participantes en cada grupo.

Algo más de la mitad de los estudios presenta menos de 100 participantes, estando la mayoría por debajo de 50, concretamente el 34.2%.

Número de grupos

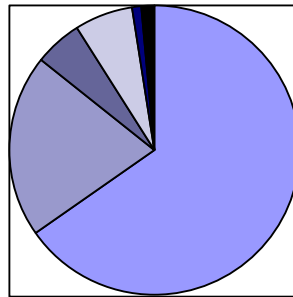


Figura 3.26. Número de grupos en el estudio.

En el 65.4% de las ocasiones, se trabajaba con un solo grupo de participantes. Cuando había más de un grupo, fundamentalmente era un par; en muy pocas ocasiones se daba un alto número de grupos.

Exclusiones tras medidas posteriores

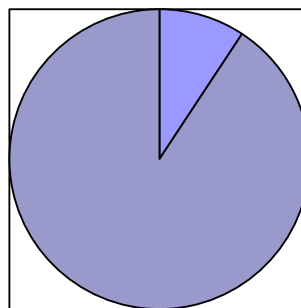


Figura 3.27. Exclusiones tras medidas posteriores.

En el 90.9% de ocasiones, el ítem referido a si se dieron exclusiones de algún sujeto después de haber tomado la primera medida no fue especificado o no fue aplicable por tratarse de un estudio teórico, con una única medida o por la casuística del objeto de estudio. De los pocos casos en los que sí se recogió el dato, prácticamente todos los estudios presentaron alguna exclusión tras la primera medida tomada.

Rango de edad especificado

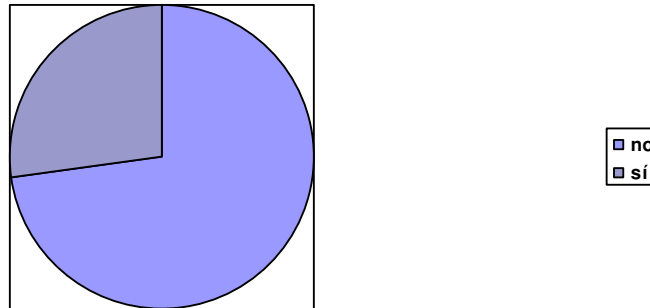


Figura 3.28. Rango de edad especificado.

En la mayoría de las ocasiones, concretamente en el 72.9% de los casos, el rango de edad no fue especificado.

Media de edad

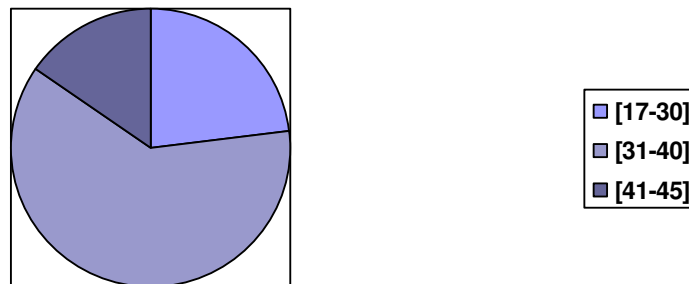


Figura 3.29. Media de edad.

Como primer aspecto a destacar respecto a la media de edad cabe señalar que, en la mayoría de las ocasiones (89.3% de los casos), se dieron valores perdidos ya sea porque no se especificó (47.2%) o por no ser aplicable (42.1%) por la naturaleza del objeto de estudio o en las ocasiones en que se trataron de estudios teóricos.

En todo caso, de los 13 estudios en que se concretó el dato, en la mayoría de las ocasiones (61.5%) las personas tenían entre 31 y 40 años (ambos valores incluidos), posteriormente le siguió el grupo de entre 17 y 30 años (23.1%) y, finalmente, el grupo de entre 41 y 45 años (15.4%).

Periodo de estudio

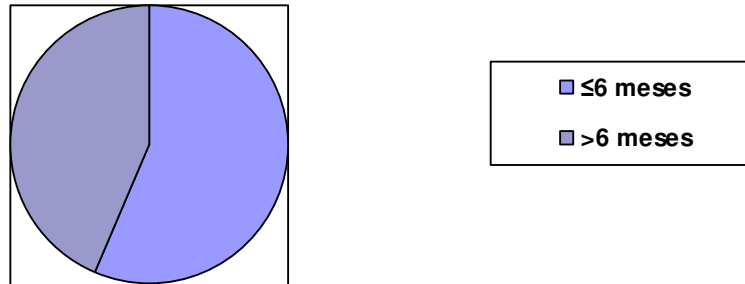


Figura 3.30. Periodo de estudio.

El periodo de estudio fue menor o igual de 6 meses en algo más de la mitad de las ocasiones (56.3%) y mayor en el resto (43.7%).

Intensidad del tratamiento/registro

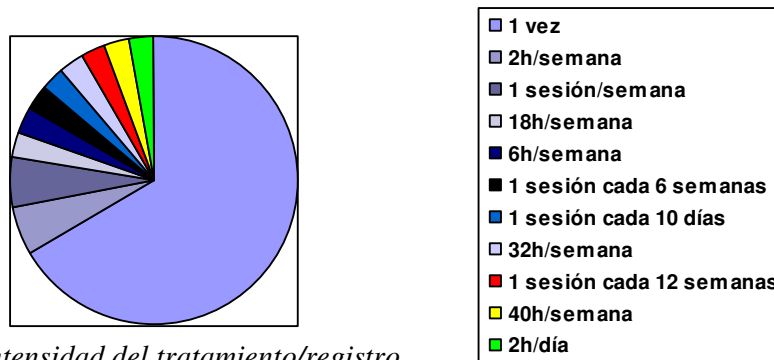


Figura 3.31. Intensidad del tratamiento/registro.

En la mayoría de las ocasiones (66.4%), el tratamiento o el registro se realizó en un único momento. Cuando se alargó en el tiempo, la intensidad que se presentó fue muy variada.

Unidades de intervención o de registro

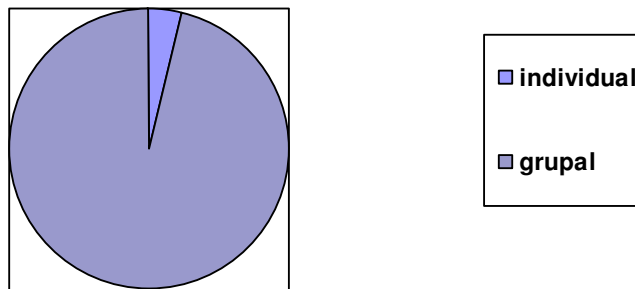


Figura 3.32. Unidades de intervención o de registro.

En casi la totalidad de las ocasiones (96.2%), el tratamiento y las medidas fueron grupales y no únicamente con individuos o un grupo considerado como una unidad.

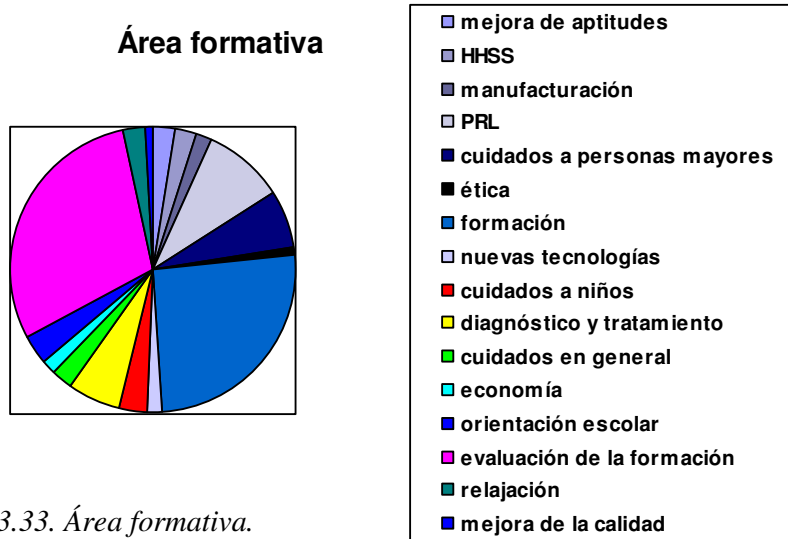


Figura 3.33. Área formativa.

Respecto al área formativa en la que se incluían los distintos estudios, en la mayoría de las ocasiones se trataron a nivel general de evaluación de la formación (29.6%) y formación (25.6%), siendo en muchas ocasiones estudios teóricos. Las áreas formativas más concretas fueron muy diversas destacando por su elevada frecuencia, aunque no con mucha diferencia respecto a las demás, la prevención de riesgos laborales (9.1%).

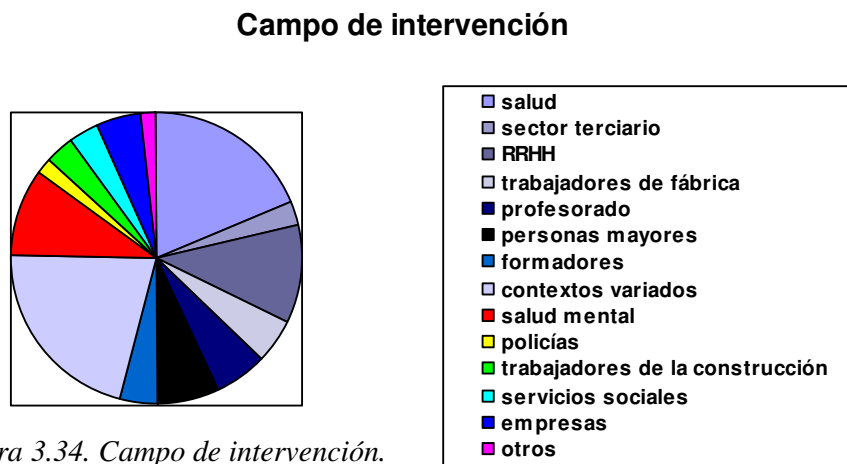


Figura 3.34. Campo de intervención.

En la mayoría de las ocasiones, el campo de intervención en los que se enmarcaron los estudios eran variados (21.4%). Como campos de intervención más concretos, destacan el ámbito de la salud con el 19% de los casos, seguido por el ámbito de los recursos humanos a bastante distancia (10.8%). Puede decirse que los campos de intervención encontrados fueron muy variados.

Tipo de publicación

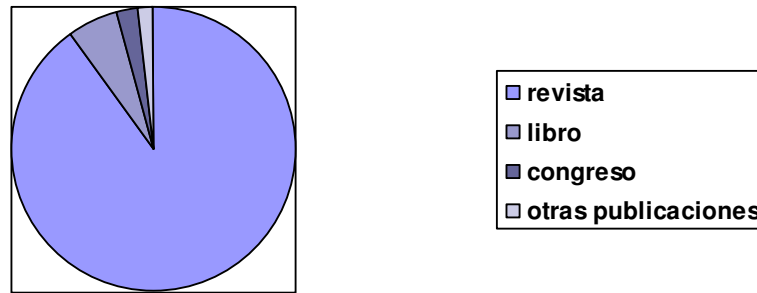


Figura 3.35. Tipo de publicación.

Respecto al tipo de publicación cabe destacar que, en la inmensa mayoría de las ocasiones (90.1%), los estudios se obtuvieron de revistas científicas.

4. DISCUSIÓN.

Idealistamente, el conocimiento científico estaría focalizado hacia la consecución del bienestar de la humanidad y estaría al alcance de todos. Sin embargo, esto no es exactamente así y de 287 trabajos que se detectaron referidos al tema, únicamente se tuvo acceso a los **textos completos** en 121, lo que supone algo menos de la mitad. Si bien es cierto que con las nuevas tecnologías (los ordenadores e internet) el acceso a la información se ha facilitado en gran medida, aún falta camino por recorrer.

En estos 121 estudios, en muchas ocasiones se encontraron **valores perdidos** tras la aplicación de la escala por dos razones diferenciadas en la tabla 3.1 de resultados: la **no aplicabilidad** del ítem, generalmente por tratarse de estudios teóricos (no empíricos) y/o por la casuística del objeto de estudio, y la **baja especificación** de los datos.

Concretamente, los ítems de los que no se explicitaba la información necesaria en mayor medida a pesar de ser aplicables en el estudio (a los que se les asignó el valor 9: no especificados) fueron los siguientes:

- Referidos a la **muestra**:
 - **Criterios de inclusión** (ítem 2): en el 34.6% de las ocasiones, no se explicitaron los criterios de inclusión y exclusión de las unidades de la muestra. Teniendo en cuenta que el 35.6% eran valores perdidos, sólo en el 29.8% de las ocasiones se explicitó el dato.
 - **Mortalidad**, tanto global (ítem 6) como diferencial (ítem 7): respecto al ítem referido al número de personas que, una vez empezado el estudio, no lo terminaron por alguna cuestión, no se especificó el dato en el 28% de las ocasiones y, además, no fue aplicable en el 35.6%, con lo que realmente se pudo recoger el dato de menos de la mitad de los estudios; más pronunciado se mostró este hecho respecto a la mortalidad diferencial entre grupos, puesto que

no podía estudiarse cuando sólo existía un grupo, por lo que las ocasiones en las que no era aplicable ascendieron al 77.7% y realmente sólo se obtuvo el dato en el 5.8% de los estudios.

- **Exclusiones posteriores**, tanto tras la agrupación de la muestra a las distintas condiciones (ítem 8), donde sólo se recogió el dato en el 19.8% de las ocasiones, como después de la primera medida tomada (ítem 26), donde se obtuvo el dato de aún menos estudios (el 9.1%).
- Especificación de la **edad**, tanto el rango (ítem 27) especificado en el 15.7% de las ocasiones, como la media (ítem 28), donde sólo se recogió el dato en el 10.7% de los casos.
- Referidos a las **técnicas de control** utilizadas, los **enmascaramientos** tanto del evaluador (ítem 13) como del usuario (ítem 14) y el formador (ítem 15), donde sólo se pudo codificar el dato claramente en cinco estudios cuando se trató del enmascaramiento del evaluador y del usuario, y en cuatro trabajos cuando se codificó la información referida al enmascaramiento del formador.
- Respecto a **cuestiones estadísticas**:
 - Los **métodos estadísticos para inferir los valores perdidos** (ítem 18) sólo fueron mencionados en el 29.7% de las ocasiones.
 - El **tamaño de efecto** (ítem 19) sólo fue aplicable y además se especificó en el 9.1% de los estudios.
 - El **índice de variabilidad** facilitado (ítem 23) sólo fue especificado y aplicable en el 30.6% de las ocasiones.
- Respecto a **cuestiones sustantivas**, el **área** (ítem 33) y el **campo** de formación (ítem 34), a pesar de que no presentan valores perdidos, concentran la mayor proporción de estudios en sus categorías más generales; concretamente, en “evaluación de la formación” y “formación” respecto al área, y en “contextos variados” respecto al campo. Por tanto, no puede considerarse que se aporte una información concreta acerca de un ámbito determinado.

Esta gran cantidad de valores perdidos puede estar afectando al índice de calidad por varias razones:

- Los **valores no especificados no puntuaron**. El bajo índice de calidad encontrado pudo ser debido a la baja calidad del reporte y no a la del estudio en sí mismo. Y como ya se ha visto, el alto porcentaje de datos no especificados fue especialmente llamativo en varios de los ítems que se contabilizaban para el cálculo del índice de calidad; concretamente, el ítem 2 referido a la exclusión/inclusión de los sujetos a la muestra; los ítems 6 y 7 referidos a la mortalidad global y diferencial respectivamente; el ítem 8 con el que se trató de recoger información acerca de las exclusiones posteriores tras la agrupación de la muestra a las distintas condiciones; los ítems 13, 14 y 15 referidos a los enmascaramientos; y los ítems 18 y 19, que hacían referencia respectivamente a

dos cuestiones relacionadas con el análisis de los datos: los métodos estadísticos para inferir los valores perdidos y el tamaño del efecto.

- Los **estudios teóricos** sólo podían puntuar en el ítem referido a la definición del constructo (ítem 17) porque el resto de los ítems eran **no aplicables** en estos casos. En este sentido, la puntuación máxima que podrían conseguir como índice de calidad se reduce a 1, por lo que este tipo de estudios no es comparable al resto, que podía alcanzar una puntuación máxima de 19 puntos. Sin tener en cuenta los estudios teóricos, la gráfica referida al índice de calidad quedaría de la siguiente manera:

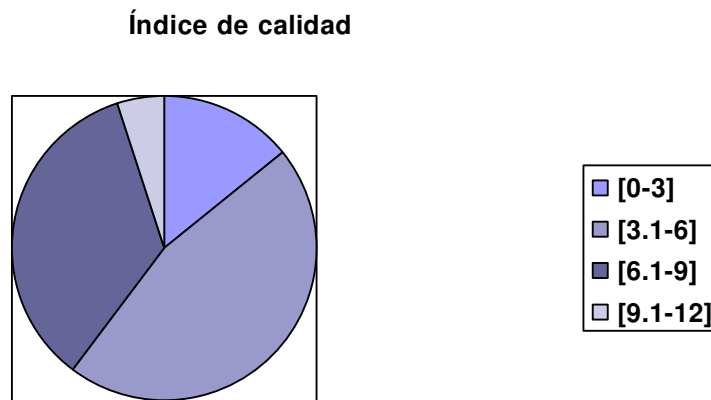


Figura 3.36. Índice de calidad tras excluir los estudios teóricos.

Esta figura representa a 78 estudios, tras excluir los 43 estudios teóricos. En comparación a la figura 3.21 en la que se representó el índice de calidad considerando todos los estudios recogidos, se puede ver que los porcentajes han mejorado. La razón es que los 43 estudios teóricos, al poder obtener únicamente un punto como valor máximo, se encontraban en el intervalo más bajo; al eliminarlos, el intervalo que incluye ahora más trabajos es el que comprende un índice de calidad de entre 3.1 y 6, ambos valores incluidos.

Sin tener en cuenta los estudios teóricos por las razones ya descritas, el resto de estudios (observacionales, de encuestas, cuasiexperimentales y experimentales) sí son comparables en lo que al índice de calidad respecta porque todos pudieron puntuar desde 0 hasta un máximo de 19 puntos. Si ahora comparásemos el índice de calidad en función del tipo de estudio, los resultados obtenidos serían los mostrados en la figura 3.37.

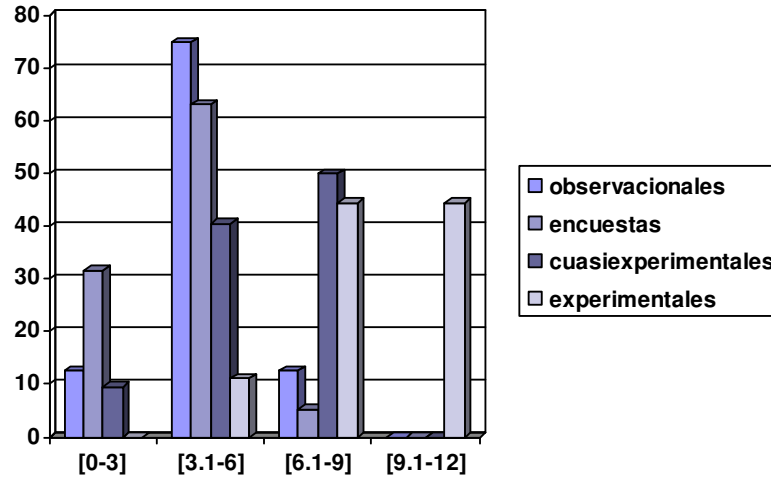


Figura 3.37. Índice de calidad distribuido según el tipo de diseño.

Concretamente, esta figura representa en porcentajes el modo en que se distribuye cada tipo de estudio entre los distintos intervalos del índice de calidad. Puede comprobarse que el índice de calidad mostrado no tiende a ser muy alto, ya que el valor máximo encontrado fue de 12 sobre un posible valor de 19. Teniendo en cuenta el intervalo de 0 a 12, parece ser que la mayoría de los estudios se encuadran entre el 3.1 y el 9, tendiendo a ser más numerosos en la mitad más baja (de 3 a 6). En relación a los diferentes tipos de estudio, los observacionales están en su mayoría en el intervalo de entre 3.1 y 6 puntos, al igual que los estudios de encuestas; los cuasi-experimentales presentan en su mayoría índices de calidad de entre 6.1 y 9; por último, la inmensa mayoría de los diseños experimentales se distribuyen en igual proporción entre el intervalo de entre 6.1 a 9 y el siguiente, de entre 9.1 a 12.

Las medias obtenidas en el índice de calidad por cada tipo de estudio se recogen en la tabla 3.2 que a continuación se presenta. El dato entre paréntesis hace referencia a la desviación típica. Además, junto al tipo de estudio se muestra la n; es decir, el número de estudios de cada tipo:

Observacional (N=8)	Encuesta (N=19)	Cuasi-experimental (N=42)	Experimental (N=9)
4.35 (2.18)	4.03 (1.29)	5.87 (1.89)	8.87 (2.4)

Tabla 3.2. Medias obtenidas en el índice de calidad respecto al tipo de estudio.

Existe una relación estadísticamente significativa al 99% de nivel de confianza entre el tipo de estudio y el valor obtenido en el índice de calidad ($F=15.29$, $p=0.000$). Se utilizó esta prueba paramétrica tras haber comprobado la homogeneidad de varianzas con el estadístico de Levene (Levene=1.73, $p=0.169$).

Tras la comparación por parejas de Scheffé se pudo comprobar, al mismo nivel de confianza, que los estudios experimentales mostraban diferencias con todos los demás y, además, existían diferencias entre estudios cuasiexperimentales y de encuesta. Las únicas parejas que no mostraron diferencias significativas fueron la relación entre estudios observacionales y de encuesta, y entre observacionales y cuasiexperimentales. Concretamente, las diferencias entre medias y la significación se

muestran en la tabla 3.3 que a continuación se presenta. Los asteriscos marcan las parejas en las que se encontraron diferencias significativas:

	Observacional	Encuesta	Cuasiexperimental	Experimental
Observacional	---	0.32 ($\rho=0.982$)	-1.52 ($\rho=0.222$)	-4.52 ($\rho=0.000$)*
Encuesta	---	---	-1.84 ($\rho=0.008$)*	0.75 ($\rho=0.000$)*
Cuasiexperimental	---	---	---	-3 ($\rho=0.001$)*

Tabla 3.3. Resultados de la comparación por parejas con la prueba de Scheffé.

¿A qué razones puede deberse que el índice de calidad no sea muy elevado? Algunas posibilidades son los **puntos negativos** encontrados:

- **Poca especificación** acerca de cómo se realizó el estudio que ya se ha comentado anteriormente: sin conocer cómo se llevó a cabo el programa de manera detallada, difícilmente será posible la generalización de dicha intervención en otros ámbitos u otros momentos, por lo que potenciales usuarios del programa aparte de los participantes originarios no podrán beneficiarse de los resultados positivos.
- No hay **grupo de comparación**: según el **ítem 1**, sólo se utilizó al menos un grupo de comparación en el 29.4% de los estudios recogidos.
- Infrecuencia del uso del **azar** en la selección de la muestra y/o su asignación a los distintos grupos (**ítem 3**). Concretamente, sólo en el 26% de las ocasiones se utilizó este procedimiento.
- **Mortalidad global elevada**. Según los resultados obtenidos en el **ítem 6** referido al número de personas que, habiendo comenzado en el estudio no finalizaron, en el 54.6% de las ocasiones (algo más de la mitad de los casos), ascendía al 20% de la muestra inicial o más.
- **No se confirma la estabilidad del dato**. En general, suele recogerse información en un solo momento en el tiempo, por lo que se corre el riesgo de estar incurriendo en una amenaza a la validez, concretamente en la **regresión estadística**, consistente en que la medida tomada resulta ser más baja o más alta de lo que en realidad debería haber sido debido a alguna circunstancia, como errores en la medida o porque algún hecho hubiera perturbado la naturalidad en la conducta de aquellas personas de las que se recoge el dato. Algunos datos que confirman lo expuesto son los siguientes:
 - Según el **ítem 4** referido al **diseño de los estudios**, sólo en el 11.7% de las ocasiones se tomaron 30 ó más medidas cuando no hay intervención o 2 ó más cuando se trata de un diseño experimental; sólo en el 14.3% de los casos se recogieron entre 2 y 29 medidas; y únicamente en el 1.3% de las ocasiones se trató de un diseño de series temporales (con más de 30 medidas recogidas a lo largo del tiempo).

- Según el **ítem 9** referido al **seguimiento**, en el 53.6% de las ocasiones (algo más de la mitad), no se daba; y sólo en el 13% de los casos era de una duración mayor al año.
- Según el **ítem 10** referido a los **momentos de medida**, la mitad de los estudios presentó únicamente medidas posteriores cuando se había dado intervención o una sola medida en caso de que fueran estudios observacionales o de encuesta.
- **Grado medio de estandarización:** la mayoría de los estudios se encuadraron en un punto intermedio entre la estandarización total y la ausencia de estandarización. Algunos datos que pueden dar base a esto son los siguientes:
 - **Carencia de variables dependientes normalizadas:** según el **ítem 12** referido a las **variables dependientes** utilizadas, en la mayoría de las ocasiones (65.4%) ninguna se midió con instrumentos normalizados, si bien es cierto que al menos una se medía con cuestionarios o autoinformes estandarizados; es decir, el uso de variables dependientes normalizadas no es lo más común.
 - **Carencia de definición sistemática de todos los constructos de manera conceptual y empírica:** según el **ítem 17** referido al modo en que se definieron los constructos, se encontró que prácticamente en todos los casos (93.4%) se definían al menos uno conceptual y/o empíricamente; sin embargo, sólo se definieron todos los constructos de ambas maneras en el 5.8% de las ocasiones.

En resumen, todos los ítems comentados (1, 3, 4, 6, 9, 10, 12 y 17) aportaban valor al índice de calidad, por lo que estarían directamente implicados en la baja puntuación conseguida.

Pero no todos son puntos negativos; también destacan algunos **aspectos positivos:**

- En la inmensa mayoría de las ocasiones (88.9%), todas las **variables** en estudio fueron **medidas** en **todos los momentos** de registro (**ítem 11**).
- En casi la totalidad de los estudios (el 88% de las ocasiones), según el **ítem 16**, se dio **homogeneidad** en la intervención o en el proceso de registro cuando no hubo intervención, lo que favoreció la comparabilidad de los datos obtenidos por los distintos participantes.
- Respecto a los análisis de datos que suelen realizarse, si bien es cierto que el ítem 19 mostró que el tamaño de efecto fue especificado únicamente en el 31% de las ocasiones, según el **ítem 21** referido al **índice estadístico calculado** se pudo observar que en la mayoría de las ocasiones (46.7%) se realizó al menos una comparación estadística, ya sea con la t, la F o con algún otro método.
- En la mayoría de los casos, atendiendo al **ítem 24**, se pudo comprobar que el número de **participantes** por grupo fue menor de 50. El bajo número de componentes en cada grupo es considerado un aspecto positivo, ya que esto da generalmente al formador la posibilidad de hacer un seguimiento más personalizado a cada uno de los participantes y permite a éstos la participación

activa. En este sentido, atendiendo esta vez al **ítem 31**, también puede considerarse positivo el hecho de que la **unidad de registro** fue en la mayoría de las ocasiones grupal, pues el aprendizaje en grupo tiende a ser más enriquecedor que el personalizado en aquellas ocasiones en las que compartir experiencias con los demás compañeros aporta tanto o más que el aprendizaje individual; también ocurre a veces que la resolución a una duda de una persona responde también a dudas que otros compañeros presentarían en algún momento posterior.

A continuación, y tomando como base los puntos negativos obtenidos, se enumeran algunas propuestas para **mejorar la calidad de los programas en formación continua** (Sanduvete, 2006).

1. Contra la poca especificación, explicación detallada:

A continuación se comentan las consecuencias que conlleva la poca especificación de las **características de la muestra**, de las **técnicas de control** utilizadas y de algunas **cuestiones estadísticas**, y cómo mejoraría la situación si estos datos se explicitaran.

La descripción difusa de las características de la **muestra** puede conllevar diversos problemas: uno de ellos es la dificultad para crear **grupos de control comparables**, ya que si se desconocen las características de las personas que componen el grupo objeto de estudio, difícilmente podrá estudiarse si un grupo alternativo parte del mismo punto que el anterior en características que pueden estar influyendo en los resultados obtenidos o si, por el contrario, pueden existir diferencias en posibles variables extrañas; otro problema derivado de la falta de especificación de las características de la muestra es la dificultad a la hora de **generalizar los resultados**. Así, según Shadish, Cook y Campbell (2002a), dos de los principios básicos para demostrar la posibilidad de generalización entre dos muestras están directamente relacionados con la concreción de sus características:

- **Excluir las irrelevancias:** identificar aquellas características irrelevantes porque el hecho de que sean diferentes en las distintas muestras no provocará cambios en los resultados, por lo que la generalización se podrá realizar sin riesgos.
- **Delimitar las discriminaciones:** clarificar las variables clave que limitan la posibilidad de discriminación.

Concretamente haciendo referencia a la explicitación de los criterios por los que algunas personas son incluidas en la muestra y otras quedan fuera, los denominados **criterios de inclusión y exclusión**, lo ideal sería poder atender a todos los potenciales usuarios, con lo que no se estaría hablando de muestreo, ya que se atendería a toda la población y no sería necesario excluir a nadie del estudio. Pero por carencia de recursos, en la mayoría de las ocasiones se hace necesaria la elección de una muestra más reducida, que no corresponde con el número de potenciales usuarios. Se suele plantear el muestreo aleatorio como la técnica de muestreo ideal desde los planteamientos de la equiprobabilidad; pero en la realidad, ocurre que esto no siempre se puede realizar, básicamente por cuestiones éticas, de tipo profesional o por restricciones técnicas (Shadish, Cook y Campbell, 2002c); en esos casos, se recomienda explicitar un criterio a partir del cual decidir a quiénes incluir en el

estudio y, posteriormente, a quienes asignar a cada grupo en los casos en que hay más de uno. Sin conocer los criterios utilizados para la selección de la muestra y su asignación a los distintos grupos, difícilmente podrá determinarse si los grupos creados son comparables o en qué tipo de usuarios podrían ser generalizables los resultados.

En resumen, la concreción de las características demográficas de la muestra como son la **media y rango de edad, los criterios de inclusión y exclusión, la mortalidad global y/o diferencial y sus causas, las exclusiones posteriores tras la agrupación de la muestra en los distintos grupos y tras la primera medida tomada, y otras características sustantivas**, pueden considerarse necesarias para crear grupos alternativos comparables, y para realizar un estudio de generalización y poder estimar probabilísticamente la eficacia del programa con otros usuarios distintos (en los casos en que hay intervención).

Respecto a las **técnicas de control** llevadas a cabo en los distintos estudios, se ha encontrado que no suele especificarse el dato, lo cual conlleva dificultades a la hora de determinar si las conclusiones a las que se llegó en el estudio pueden ser creíbles o alguna variable extraña pudo influir en los resultados.

A partir de esto las recomendaciones que podrían darse serían, en primer lugar, que se hiciera un estudio de posibles variables extrañas que pudieran influir en el estudio; en segundo lugar, que se eligieran unas técnicas de control adecuadas para controlar dichas variables extrañas y se llevaran a cabo; por último, que se detallara en el informe o publicación todo este procedimiento, lo que daría credibilidad a los resultados encontrados y posibilitaría a otros profesionales replicar el estudio. Aquello que no se escribe, para el lector, es como si no se hubiera hecho.

¿Qué técnicas de control específicas podrían recomendarse? Existen diversas posibilidades. Por ejemplo, en muy pocas ocasiones en los trabajos publicados se especifica la realización de alguna técnica de **enmascaramiento**, ya sea del **usuario**, del **evaluador** o del **profesional** que lleva a cabo la intervención en caso de que la haya, modos de control de variables extrañas característicos de los diseños experimentales.

Ante estos resultados encontrados, una posible mejora a proponer sería fomentar el uso de algún tipo de enmascaramiento en los estudios, pues esto permite proteger los resultados de gran cantidad de variables extrañas ya que, a través de la desinformación de quienes realizan la implementación, la evaluación y/o de los usuarios acerca de los resultados que se esperan tras la intervención, se evita la posibilidad de manipulación consciente o inconsciente de los resultados reales.

Además de las técnicas mencionadas, se recomienda poner en marcha otras también útiles y más susceptibles de aplicar en intervenciones no experimentales, como podrían ser las siguientes (López, Trigo y Arias, 1999):

- El **balanceo**, consistente en hacer que cada valor concreto de la variable a controlar esté igualmente representado en cada grupo y/o en cada momento de medida.

- El **mantenimiento constante** consistente en hacer que permanezca, durante toda la intervención y en todos los grupos, el mismo valor en una variable que podría influir en los resultados, sin estar incluida en el diseño.
- **La eliminación** (técnica de control genérica en la que se incluye el enmascaramiento), que consiste en hacer que el valor de la variable que tiene posibilidad de influir sea nulo en todas las medidas, ya sea entre grupos y/o en los distintos momentos.

Por último, otro aspecto en que generalmente no se especifica el dato es el referido a las **cuestiones estadísticas**. A continuación, se hacen algunas concreciones a este respecto:

En la mayoría de los trabajos que se estudiaron, el **tamaño de efecto** no se especificó; generalmente, sólo se dieron los resultados a nivel descriptivo como, por ejemplo, a través de la media, porcentajes o, a lo sumo, se realizó alguna comparación entre medias (t o F, por ejemplo).

Como primera propuesta de mejora a este respecto, se recomienda al menos una mayor especificación de los datos descriptivos, ya que en muchas ocasiones ni tan siquiera se aporta la desviación típica de los grupos, con lo que no se conoce la **variabilidad** en cada uno de ellos y no se posibilitan los estudios necesarios a la hora de elegir el estadístico de contraste a utilizar, como la prueba de homocedasticidad.

Yendo un poco más allá, una mejora más ambiciosa pero también más útil por aportar más información, sería no quedarse en el mero análisis descriptivo, sino llegar a la **estimación del tamaño de efecto** con sus intervalos de confianza, con lo que se lograría no sólo determinar si hubo diferencias entre las diversas medidas que se recogieron, sino también concretar el tamaño de dicha diferencia.

En algunos casos ocurre que no todas las personas que comenzaron el programa lo finalizaron, con lo que el número de datos recogidos previamente a la intervención es mayor que el recogido posteriormente. Cuando el porcentaje de personas que comienzan y no finalizan es muy elevado, pueden surgir problemas a la hora de realizar el análisis de datos, sobre todo cuando esta **mortalidad** se da de manera **diferencial** en los distintos grupos. Para intentar solventar este tipo de cuestiones, se sugiere que, a posteriori, una vez recogidos los datos, se estimen los **valores perdidos** a través de algún método estadístico adecuado y, por supuesto, que se explicita el **método** utilizado para ello. Más información acerca de algunas posibilidades se presenta más adelante en el punto 4.

2. Contra la carencia de grupos de comparación, uso de éstos:

En la mayoría de las ocasiones, los estudios constaban únicamente de un grupo de participantes, lo que suponía la imposibilidad de realizar estudios comparativos entre grupos.

La propuesta de mejora respecto a este punto sería la potenciación del uso de **al menos un grupo de comparación**, siendo aún más positivo que los grupos de comparación fueran **múltiples** a ser posible.

Para lograr un grupo comparable a otro, el primer paso necesario es la explicitación detallada de las **características de la muestra**, aspecto ya comentado en el apartado anterior.

Respecto al **método** con el que realizar la asignación de las personas a los distintos grupos (siempre y cuando haya intervención puesto que en los estudios observacionales y de encuesta la conformación de los grupos viene determinada de manera natural por las propias características de los participantes), la **asignación aleatoria** es el más recomendable por el control de variables extrañas que supone su uso. Cuando no es posible, ya sea por cuestiones éticas o por otras cuestiones, los grupos de **cohortes** son los grupos de comparación no equivalentes más recomendados (Shadish, Cook y Campbell, 2002a). Estos métodos y algunos otros se explicitan con más detalle en el siguiente punto.

Puede que se presentaran ocasiones en las que no fuera posible la realización de al menos un grupo más para posibilitar la comparación. Para estos casos, los **datos secundarios** son una posible alternativa. Consistiría, por ejemplo, en buscar estudios semejantes al de interés y tratar de encontrar resultados comparables, como los obtenidos al medir la misma variable dependiente.

3. Contra la infrecuencia del uso del azar, modo de asignación explicitado:

En los estudios en que hay intervención y más de un grupo, en muy escasas ocasiones los usuarios son asignados aleatoriamente a éstos, aunque suelen controlarse posibles variables extrañas durante el proceso. Los diseños cuasi-experimentales suelen ser mucho más utilizados que los experimentales.

Desde un planteamiento exclusivamente metodológico, respecto a la **asignación a las condiciones del programa**, se insiste generalmente en las ventajas de que se utilicen **criterios completamente conocidos, preferiblemente aleatorios (diseños experimentales)**, con lo que se potencia una estimación insesgada del efecto. Consistiría, por ejemplo, en realizar dos grupos de personas: uno que participara en el programa y otro que sirviera de comparación (grupo control, que no participaría en intervención alguna). Las personas concretas serían asignadas a uno u otro grupo en función del azar.

Sin embargo, no podemos estar seguros de que este muestreo “equiprobable” garantice la “equidad”, entendida ésta como la dotación de los recursos disponibles en función de las necesidades concretas de cada usuario, en lugar del reparto igualitario. ¿Qué ocurriría si la limitación de recursos no permitiera que todos los potenciales usuarios pudieran tener acceso al programa? Utilizando el azar, quizá quedarán fuera algunas personas que necesitaran con mayor premura participar en el programa. ¿Se puede entonces considerar justa esta asignación? Quizá no; tal vez fuera preferible haber usado otros métodos.

Así pues, para los casos en los que no es posible o no se considera ético utilizar la aleatorización como forma de asignación de las personas a los distintos grupos, se proponen distintas alternativas para la **conformación de grupos lo más**

similares posibles entre sí. Concretamente, éstas son algunas propuestas de actuación (Shadish, Cook y Campbell, 2002b):

- El **emparejamiento previo:** consiste en hacer grupos en función de las características de las personas y, una vez conformados, asignar a usuarios de características similares a los distintos grupos, de tal manera que así se está seguro de que éstos no quedan descompensados, al menos, respecto a las características estudiadas. Así, si existieran diferencias entre ambos grupos en los niveles de la variable dependiente después de la intervención, la variable compensada quedaría descartada como posible causa del cambio. Un ejemplo de diseño cuasiexperimental que toma como punto de partida este método es el **diseño con grupo control no equivalente y prueba previa y posterior.** Consistiría en, una vez conformados dos grupos similares, tomar al menos una medida de la variable dependiente en los dos grupos en el mismo momento antes de la intervención y otra después de ésta. Si el programa fuera realmente eficaz, se encontraría cambio al medir la variable dependiente en el grupo que participó en el programa, mientras que quienes no participaron mantendrían niveles semejantes a lo largo del tiempo; en el caso de que no hubiera diferencias en ninguno de los grupos, se podría concluir que el programa no fue eficaz; sin embargo, si en ambos grupos se encontraran cambios tras el momento de intervención, tampoco se podría concluir con seguridad que el programa fuera eficaz porque podría ser que las diferencias se debieran a otras vivencias de las que participaron ambos. Este diseño muestra ciertos problemas como, por ejemplo, que los cambios se deban a la maduración de las personas por el paso del tiempo en lugar de a la intervención en sí. Este y otros problemas se solventan con el uso del diseño que a continuación se presenta.
- Los **grupos de cohortes:** consiste en que, a pesar de que no son las mismas personas a lo largo del tiempo, el efecto logrado es semejante al encontrado en los estudios longitudinales, pues los grupos están formados por personas que, aun teniendo características similares y viviendo circunstancias muy parecidas, difieren en la edad. Tomando como punto de partida este método estaría el **diseño de cohortes básico.** Esto supone recoger datos con casi la misma validez que los conseguidos con un método longitudinal (las mismas personas a lo largo del tiempo) pero en un periodo más breve.
- La **explicitación del criterio de inclusión** a uno y otro grupo: **punto de corte.** El diseño considerado de mayor calidad de entre los diseños cuasiexperimentales es el **diseño de discontinuidad en la regresión,** ya que comparte una característica con los aleatorios, y es que se consigue la estimación insesgada del efecto del programa. Consiste en explicitar un valor concreto de la variable dependiente como punto crítico, de tal manera que todas las personas que sobrepasen este valor comenzarán a participar en el programa. Para que se pueda llevar a cabo este diseño, es necesario que se cumplan ciertas características, por lo que su uso queda restringido: han de existir relaciones lineales entre variables y se ha de asumir que la línea de regresión se prolongaría más allá del punto de corte si no se produjese la intervención. Además, presenta algunos problemas: uno de ellos es la dificultad de generalizar de unas situaciones a otras por lo específico que es cada diseño; el otro es la dificultad de disponer de un punto de corte adecuado: ¿por qué tomar un valor y no otro? Probablemente, dependiendo de las características de la sociedad en la que llevemos a cabo la intervención, este punto variaría.

¿Cómo saber en cada caso cuál es el adecuado? Una opción interesante sería oír la opinión de diversos expertos y llegar a un consenso.

Independientemente del tipo de asignación de las personas a los distintos grupos que se lleve a cabo, es fundamental no variar la asignación posteriormente o excluir a personas que previamente fueron asignadas al grupo de intervención porque esta forma de actuar, además de suponer la pérdida de la equivalencia probabilística intergrupos, implicaría importantes amenazas a la validez del estudio, como la mortalidad diferencial o la igualación compensatoria, consistente esto último en la tendencia a llevar a cabo al menos parte de la intervención y, en ocasiones, otra intervención, por no dejar a su suerte a las personas del grupo control, que generalmente va a presentar también, al igual que el grupo de intervención, la necesidad de cambio.

4. Contra la mortalidad global elevada, estudiar las razones e inferir los valores perdidos:

El número muy elevado de valores perdidos al recoger los resultados del estudio puede conllevar problemas a nivel estadístico a la hora de tratar de llegar a las conclusiones. Las propuestas de mejora ante esto son, en primer lugar, **estudiar las razones** por las que se dio para tratar de evitarlo en futuras implementaciones y porque de este modo se puede obtener información interesante acerca de cómo mejorar la intervención para otras ocasiones; como segundo paso, en caso de que fuera apropiado, sería recomendable utilizar **métodos de inferencia de valores perdidos**, algunos de los cuales son los que a continuación se presentan:

- El caso más burdo consiste en **eliminar** todos los datos de aquella persona de la que se obtuvo algún valor perdido, con la pérdida de información que esto conlleva. Técnicamente no se le podría denominar método de inferencia ya que, en lugar de inferir, simplemente elimina grupos de valores no completos.
- Cuando se trata de un estudio con pretest y postest, se pueden reemplazar los datos perdidos en el postest con los **mismos** obtenidos en el pretest, suponiendo así que no haya habido cambio.
- **Media de la serie:** este método consiste en sustituir los valores perdidos por la media de la variable.
- **Media de puntos adyacentes:** sustituye cada valor perdido por la media de sus valores adyacentes válidos.
- **Mediana de puntos adyacentes:** con la misma lógica que en el caso anterior, en esta ocasión se utiliza la mediana en lugar de la media.
- **Interpolación lineal:** en este caso se sustituyen los valores perdidos utilizando una interpolación lineal basada en los dos valores adyacentes al valor perdido. Si el valor perdido se encontrara en un extremo de la serie, no sería reemplazado por falta de uno de los referentes.

- **Tendencia lineal en el punto:** consiste en sustituir cada valor perdido por los pronósticos obtenidos al llevar a cabo un análisis de regresión lineal utilizando la serie como variable dependiente y cada sujeto como variable independiente.

La decisión de qué método utilizar dependerá de la situación concreta; por ejemplo, si hay muchos valores perdidos en las posiciones extremas, no sería recomendable usar la interpolación lineal; o si existe mucha variabilidad en los datos, no sería conveniente utilizar la media o la mediana al ser estos índices de tendencia central, pobres en dar información en estos casos. En líneas generales podría afirmarse que el método presentado en primer lugar referido a la eliminación de la serie sería el menos recomendado por la pérdida de información que conlleva su utilización.

5. Contra la ausencia de estudio sobre la estabilidad del dato, observaciones múltiples:

Como ya se ha comentado previamente, en muchas ocasiones se presenta un número muy bajo de medidas, dándose a menudo únicamente una.

En caso de que se trate de un estudio sin intervención (observacional o de encuesta), la mejora general que se propone es que aumente el número de momentos de medida, con lo que podría realizarse un estudio de estabilidad del dato o de cambio a lo largo del tiempo.

En caso de que sí exista intervención, la propuesta de mejora es la misma, el aumento del número de medidas, pero distinguiendo entre medidas **previas**, **posteriores** y **de seguimiento**, con la idea de evaluar a lo largo de todo el proceso, llevando a cabo así la denominada evaluación formativa en contraposición a la sumativa o final, que permite no sólo la posibilidad de detectar posibles dificultades (que se lograría meramente con un proceso de monitorización), sino también la capacidad de actuar para mejorar la situación en el mismo momento, con las consecuencias positivas que esto conlleva (Anguera y Chacón, en prensa, b).

Respecto a las **medidas previas** a la intervención del programa cabe destacar que, en algo más de la mitad de las ocasiones no se dan, recogiendo únicamente medidas después de la implementación del programa.

La primera recomendación a este respecto es recoger **al menos una** medida previa aunque, siempre que sea posible, es preferible recoger **múltiples**, cuanto más numerosas mejor, dentro del marco de la **validez** y con la mejor **calidad de medida** posible (Shadish, Cook y Campbell, 2002b). Aunque en algunos ámbitos se considera que los registros fisiológicos y estandarizados son los de mayor calidad por su objetividad, la realidad es que los instrumentos semiestandarizados pueden ser más útiles e igual de fiables para medir determinadas variables subjetivas, cambiantes en función de la persona aunque el contexto físico sea el mismo, como podría ser la satisfacción de un alumno con el curso en el que participó.

Una medida previa es algo sumamente importante porque no se podrá determinar si el programa fue o no eficaz, o si provocó o no cambios en la variable sobre la que se pretendía intervenir, si no se conoce el punto desde el que parten los

usuarios del programa. Cuando sólo se recoge una medida previa, sin embargo, amenaza la denominada **regresión estadística**, consistente en que la medida tomada, por alguna circunstancia, resulta ser más baja o más alta de lo que en realidad debería haber sido. Al comparar esta medida previa tan extrema con la posterior a la intervención, los cambios pueden resultar irrealmente significativos. Es por ello que se recomienda tomar varias medidas anteriores a la intervención, hasta detectar una línea base estable que permita determinar con un cierto grado de certeza cuál es el nivel del que se parte.

Cada vez existe mayor acuerdo en afirmar que partir de datos previos a la intervención es fundamental. De ahí la importancia creciente que se está dando a la **detección de necesidades** previa al diseño de la intervención: por una parte, es una forma de justificar, a partir de evidencias sistemáticamente recogidas, que el programa es necesario y no se hace de manera fortuita, sin causa; por otra parte, sirve para disponer de datos que se pueden comparar, con posterioridad, con la información obtenida tras la intervención. En el capítulo 4, se hablará más pormenorizadamente de esta fase de evaluación.

En los momentos en que no sea posible recoger ni tan siquiera una medida previa, existen varias alternativas para solventar en cierta medida esta carencia. Algunas son las siguientes:

- Recoger **medidas previas de muestras independientes** con lo que, aunque no se tenga un dato previo de las personas que recibieron la intervención, al menos sí se tienen datos de referencia.
- Usar **medidas retrospectivas**, consistentes en preguntar a las personas que recibieron la intervención en qué punto se encontraban en el aspecto que se pretendió modificar con el programa antes de participar en éste. Hay que tener en cuenta que la comparabilidad de los datos obtenidos por este método puede estar sesgada, ya que el paso del tiempo puede provocar distorsión en la respuesta por diversas cuestiones como pueden ser el olvido, la falta de claridad de la percepción que tenían antes o la idealización de la situación presente en comparación con la anterior.

Respecto a las **medidas posteriores** a la intervención, cabe destacar que no suelen faltar por el uso generalizado de la evaluación sumativa. Sin embargo, una posible mejora sería recoger información repetidamente a lo largo del tiempo, para descartar así la posibilidad de que los resultados fueran positivos pero percederos de tal manera que, poco después de la intervención, el nivel de la variable dependiente volviera al punto de partida. En este sentido, la mejora propuesta sería realizar **seguimiento** de amplia duración.

Tras la intervención, en la mayoría de las ocasiones no se hace seguimiento y, cuando se hace, dura menos de seis meses.

Sin embargo, la recogida de esta información es fundamental porque, ¿cuál es el objetivo de una intervención que no mantiene los resultados positivos obtenidos a lo largo del tiempo? ¿Qué sentido tiene afirmar que un programa es eficaz sin antes estudiar si las personas han salido beneficiadas con la intervención y no sólo de modo

puntual, volviendo más tarde al punto del que se partió? Es imprescindible que las intervenciones no busquen la mejora fácil, puntual en un momento concreto en el tiempo y superficial, sino que persigan el desarrollo autónomo y continuado de las personas, fomentándose de esta forma la autosuficiencia, la independencia y el empowerment, porque no siempre van a tener a un experto a su lado para ser ayudados.

Por tanto, la mejora propuesta no es sólo la realización de seguimiento sino que además, si las limitaciones de los recursos disponibles lo permiten, se **augmente el tiempo de seguimiento** como manera de demostrar con más fuerza que efectivamente los resultados se mantienen, especialmente en el caso de ciertos programas como son los de **prevención**. Actualmente, la investigación sobre sus efectos a largo plazo es escasa. Probablemente, con un seguimiento duradero se demostraría el gran ahorro que supondría al estado este tipo de programas. La mayor duda que provocan es que puede parecer innecesario intervenir en una necesidad que aún no ha surgido. Sin embargo, son sobre seguro muy eficientes porque, con una inversión inicial menor que en el caso de las intervenciones de otro tipo, además del potencial beneficio aportado al usuario se podría evitar que en muchos casos llegara a surgir el problema, por lo que la institución ahorraría muchos recursos a largo plazo, aunque los efectos se hicieran visibles de una forma menos inmediata.

La posibilidad del seguimiento se puede complementar con distintas variantes de diseño, de entre las que podemos entresacar como más usuales las siguientes (Shadish, Cook y Campbell, 2002b):

- La realización del **diseño de repeticiones de tratamiento múltiple intercambiado**, consistente en implementar un programa distinto en cada grupo y, tras cierto periodo de tiempo, intercambiar los tratamientos. Se tomarían medidas y se trataría de comprobar si alguna de las técnicas fue más eficaz que la otra.
- El **diseño de repeticiones cambiadas** consiste en ir intercambiando la aplicación de la intervención en uno o más grupos de usuarios de tal manera que, por ejemplo, un grupo de usuarios que recibía la intervención se convierte en el grupo control y viceversa (quien era el grupo control ahora recibe la intervención); esto se puede hacer cuantas veces se vea conveniente, con objeto de analizar si es realmente el programa el que provoca el cambio. Es un diseño semejante al anteriormente comentado; la mayor diferencia radica en que, en este caso, sólo existe un tipo de intervención. A pesar de que este diseño es muy útil para obtener inferencias válidas, presenta muchos problemas de aplicabilidad: por un lado, porque el efecto del programa debe desaparecer rápidamente una vez eliminado éste, lo que requiere un alto grado de control; por otro lado, por el problema ético que supone no intervenir con un programa que potencialmente provoca resultados positivos. La idea es que, una vez validada la eficacia del programa mediante este diseño, el profesional que lleva a cabo la intervención busque algún tipo de sustituto natural al programa que provoque el mantenimiento de los resultados deseados una vez que se deje de implementar éste.

6. Estandarización del proceso:

En general, se encuentra un nivel medio de estandarización. Algunos aspectos podrían ser mejorados:

En referencia a la **carencia de variables dependientes normalizadas**, se recomienda el uso de variables dependientes objetivas aunque teniendo en cuenta, como ya se ha comentado que, en ocasiones, los instrumentos semiestandarizados pueden ser más útiles e igual de fiables para medir determinadas variables subjetivas.

Otra alternativa es el uso de **variables dependientes no equivalentes** en los casos en que se realiza una intervención (Chacón, Shadish y Cook, en prensa). Consiste en tomar medidas de variables que, aunque no son concretamente las que se pretendían estudiar, guardan estrecha relación con éstas. La idea es que si las mejoras se han debido a la intervención, la variable sobre la que se pretendía actuar habrá cambiado, mientras que la variable dependiente no equivalente mantendrá sus valores sin cambio.

En relación a la **carencia de definición empírica y teórica de los constructos**, se recomienda que se definan detalladamente a nivel teórico y, sobre todo, empíricamente, lo que facilitará la validez de la medición; es decir, que se mida lo que realmente se quiere medir.

Ante la gran variedad posible de acciones formativas a realizar, la mejor manera de actuar antes de intervenir es delimitar con precisión qué aspectos se van a considerar relevantes (en qué constructos concretos se va a querer encontrar mejoras) y una vez elegidos, definirlos de manera operativa y medible. Acotando el objeto de intervención desde el principio y determinando cómo van a medirse los cambios de cada constructo, se potenciará la coherencia a lo largo de todo el programa y se podrá tener una idea más precisa sobre los aspectos en los que se ha de intervenir para aumentar la probabilidad de eficacia del programa.

Hay que tener en cuenta que no hay definiciones de los constructos verdaderas o falsas porque, en realidad, ¿quién tiene la verdad a su alcance? Como mucho, se pueden obtener subjetividades compartidas (opiniones que, llevadas a consenso, pueden parecer más ciertas que otras). Por ello, se considerará un constructo bien definido cuando se base en un modelo teórico o al menos tenga otra justificación, otra manera de argumentación aparte de una mera opinión personal (como podría ser la experiencia profesional en el tema a tratar) y cuando se haga de manera operativa, con lo que se posibilita estudiar los cambios.

5. CONCLUSIONES.

Con la aplicación de la escala elaborada, aun teniendo en cuenta que no se trata de un instrumento cerrado y definitivo se obtuvo, a modo de estudio exploratorio, un índice global de calidad para cada trabajo en estudio; pero aún de mayor interés fue la recopilación de los resultados obtenidos en cada uno de los ítems, lo que dio la posibilidad de proponer mejoras concretas para mejorar la calidad de los estudios que suelen realizarse y publicarse acerca de la formación continua.

Uno de los aspectos más llamativos encontrados fue el alto porcentaje de valores perdidos hallados en muchos ítems, en primer lugar porque muchos de los estudios recopilados acerca de la temática eran de carácter teórico; y en segundo lugar, porque no existía un alto grado de especificación de los datos requeridos.

Respecto a la cuestión de si existía relación entre el tipo de diseño presentado y el nivel de calidad del estudio, los resultados muestran que existen diferencias estadísticamente significativas, a favor de los estudios experimentales. Quizá esto sea debido a que la inespecificación de los datos es menor en los estudios que presentan este diseño.

Los resultados apoyaron también algunas ideas extendidas en la literatura, como el escaso uso de los diseños experimentales en el ámbito de la psicología o la carencia de estandarización de los procesos en los programas de formación.

Tomando como base los resultados encontrados, en la tabla 3.4 se presentan los que fueron considerados los puntos negativos encontrados y las posibles propuestas de mejora de cara a aumentar la calidad de los programas de formación continua.

PUNTOS NEGATIVOS	PROPUESTAS DE MEJORA
Poca especificación:	Mayor especificación para poder crear grupos de comparación y facilitar la generalización de los resultados:
De la muestra.	Criterios de inclusión y exclusión; mortalidad (global y diferencial); exclusiones posteriores tras la agrupación de la muestra en las distintas condiciones y tras la primera medida tomada, edad (rango y media) y otras características sustantivas.
De las técnicas de control utilizadas.	Enmascaramientos y otras técnicas.
De cuestiones estadísticas.	Inferencia de valores perdidos, cálculo del tamaño de efecto e índices de variabilidad.
No hay grupo de comparación.	Múltiples grupos de comparación. Necesidad de especificación de la muestra. Como 1ª opción si es posible y pertinente, asignación aleatoria. Como 2ª opción, asignación conocida (diseño de discontinuidad en la regresión). Como 3ª opción, grupo de cohortes. Como última opción, uso de datos secundarios.
Infrecuencia de uso de azar.	Uso de azar. Si no es posible, al menos explicitación clara del modo de selección y asignación de las personas a los distintos grupos.
Mortalidad global elevada.	1º. Estudiar las razones a las que se debe. 2º Métodos de inferencia de valores perdidos.
No estudio de la estabilidad del dato:	Estudiar la estabilidad del dato:
Pocas medidas, a menudo una.	Observaciones múltiples.
No hay seguimiento.	Seguimiento de larga duración.
Sólo se dan medidas posteriores.	Observaciones antes de la intervención (cuando la hay): Múltiples, al menos una. Otras opciones: medidas a muestras alternativas o medidas retrospectivas.
Poca estandarización:	Mayor grado de estandarización:
No hay variables dependientes normalizadas.	Variables dependientes objetivas/normalizadas, siempre que sean las más óptimas para medir el constructo.
Los constructos no se definen empírica y teóricamente.	Definición teórica y empírica de los constructos.

Tabla 3.4. Puntos negativos y propuestas para mejorar la calidad de los estudios primarios.

Aunque estas mejoras son aplicables a estudios con cualquier tipo de diseño, se recomienda su uso especialmente en los diseños observacionales y de encuestas, ya que son los que presentan un índice de calidad medio menor.

Hasta aquí, se han estudiado posibilidades para mejorar la calidad de los estudios primarios en formación continua, pero cabría preguntarse si los programas de formación que suelen hacerse son eficaces. Para responder a esta cuestión, en el **capítulo 4** se presenta un **estudio meta-analítico** en el que se unifican todos los resultados que se especificaron en los trabajos estudiados.

La carencia de la disponibilidad del dato puede convertirse en una verdadera barrera en algunas situaciones como, por ejemplo, a la hora de generalizar los resultados a otras muestras o en otros momentos, o también en el momento de realizar un estudio meta-analítico y tratar de determinar las variables moderadoras en una intervención concreta con resultados heterogéneos. Así, como se verá, de 121 trabajos encontrados referidos a la evaluación de la formación continua, sólo 17 mostraron los datos suficientes como para ser incluidos en dicho estudio.

CAPÍTULO 4. EFICACIA DE LOS PROGRAMAS DE FORMACIÓN CONTINUA: UN ESTUDIO META-ANALÍTICO

1. Introducción	181
2. Método	181
2.1. Muestra.....	181
2.2. Instrumentos.....	182
2.3. Procedimiento.....	182
3. Resultados.....	186
3.1. Descriptivos.....	186
3.2. Eficacia	189
3.3. Variables moderadoras	190
4. Discusión.....	193
5. Conclusiones	198

CAPÍTULO 4. EFICACIA DE LOS PROGRAMAS DE FORMACIÓN CONTINUA: UN ESTUDIO META-ANALÍTICO.

1. INTRODUCCIÓN.

El **meta-análisis** es una metodología que se basa en la integración cuantitativa de los resultados de las investigaciones empíricas sobre un determinado tema (Sánchez-Meca, Rosa y Olivares, 1999; Letón y Pedromingo, 2001).

Esta técnica estadística permite, por una parte, aunar resultados de diversos estudios para llegar a una conclusión general acerca de la eficacia de los programas en un contexto determinado y, además, tratar de determinar **variables influyentes o moderadoras** de los resultados cuando existe heterogeneidad entre los distintos estudios aunados; es decir, identifica las razones por las que los resultados individuales fueron diversos e incluso contradictorios en caso de que así haya sido (Hedges y Olkin, 1985).

En este capítulo, se presenta un estudio meta-analítico aplicado a los programas de formación continua, con el que se tratan de alcanzar dos **objetivos**:

1. Obtener un dato a partir del cual poder concluir acerca de la eficacia de los programas de formación continua.
2. Detectar posibles variables moderadoras (extrínsecas, sustantivas o metodológicas) que estén influyendo en los resultados, en caso de que exista heterogeneidad entre ellos.

2. MÉTODO.

2.1. Muestra.

Los estudios analizados se recogieron a partir de la muestra obtenida para la aplicación de la escala a los programas de formación continua (ver capítulo anterior). Las bases de datos utilizadas a las que se consiguió tener acceso y a la vez consideradas de interés por la temática de los artículos que contienen fueron EBSCO Online, Medline, Serfile, CABHealth, CINAHL, PsycINFO, Econlit, ERIC, MathSci, Current Contents y Humanities Index.

Los criterios de inclusión fueron los siguientes:

1. Los trabajos tuvieron que versar sobre formación continua a empleados, y no sobre otros tipos de formación, como la ocupacional (a desempleados) o a cuidadores no formales (padres o hijos, por ejemplo).

2. Los estudios tuvieron que estar disponibles en su versión completa; se desecharon aquellos casos de los que sólo pudieron obtenerse los resúmenes.
3. Al menos, tuvieron que tomarse dos medidas, ya fueran del mismo grupo a lo largo del tiempo o de grupos diferentes, para poder realizar un estudio comparativo.
4. Los trabajos tuvieron que presentar datos suficientes como para poder calcular los tamaños de efecto.

De los 2379 trabajos localizados inicialmente, tras excluir los que no cumplieron los cuatro criterios, únicamente quedaron para el estudio 11. Concretamente, fueron los siguientes: Arranz, Ulla, Ramos, del Rincón y López-Fando (2005); Coogle, Parham, Welleford y Netting (2002); Irvine, Ary y Bourgeois (2003); Khan (2002); Lintzeris, Ritter y Dunlop (2002); Newes-Adeyi, Helitzer, Roter y Caulfield (2004); Quinet, Nunn y Kincaid (2003); Wilkinson, Leliopoulou, Gambles y Roberts (2003); Bray y Howard (1980); LaLonde (1986); y Panter (1994).

Estos 11 trabajos dieron lugar, por un lado, a 14 estudios independientes con medida posterior tomada inmediatamente al finalizar la intervención, puesto que algunos de estos trabajos presentaron más de una acción formativa diferente implementada con diferentes participantes; y, por otro lado, a 2 estudios en los que la medida posterior se realizó durante el seguimiento, un tiempo después de la finalización de la formación.

2.2. Instrumentos.

Los instrumentos utilizados para la realización de este estudio meta-analítico fueron los siguientes:

El **software Procite-5**, con el que se realizó el tratamiento de la información obtenida. Sirvió, entre otras cuestiones, para detectar y posteriormente eliminar los artículos duplicados tras la búsqueda en las diversas fuentes.

La **escala para medir la calidad de los estudios primarios** previamente elaborada (ver tabla 2.5 en el capítulo 2) se utilizó para la codificación de los estudios encontrados.

El **software SPSS 14.0** se utilizó para la codificación de los artículos y el análisis de resultados descriptivos, mientras que con **Microsoft Excel** se calcularon los índices de calidad, las medias, las desviaciones típicas y las “d” individuales de cada estudio.

El **software Comprehensive Meta Analysis** fue utilizado para el cálculo del tamaño de efecto global y para el estudio del efecto de las posibles variables moderadoras sobre los resultados.

2.3. Procedimiento.

Como se comentó previamente, este estudio meta-analítico partió de la búsqueda bibliográfica realizada para determinar las características que suelen presentar los estudios sobre formación continua descrita en el capítulo 3. A continuación, se recuerda brevemente los pasos seguidos y se continúa con los demás realizados hasta alcanzar los resultados del meta-análisis.

1°. **Búsqueda de los artículos.** Se llevó a cabo la búsqueda de los estudios que versaban sobre formación continua en las bases informatizadas ya comentadas. Las palabras clave utilizadas fueron “training programs”, “evaluation” y “work”. Se encontraron 2379 resúmenes, tras eliminar las duplicaciones.

2°. **Aplicación de los criterios de inclusión.** Partiendo de los resúmenes encontrados, se procedió a la aplicación de los criterios de inclusión: tras excluir los trabajos que no trataban sobre formación continua, quedaron 287 resúmenes; sólo se consiguió obtener la versión completa de 121 de estos trabajos; y tras eliminar aquellos en los que no se aportaba al menos dos medidas comparables y datos suficientes como para poder calcular los tamaños de efecto, la muestra se redujo a 11 artículos publicados.

3°. **Codificación de las características de los estudios.** Se obtuvo información de los 11 artículos incluidos siguiendo la escala para la medición de la calidad de estudios primarios elaborada previamente (ver capítulo 2). A modo de recordatorio, cabe destacar que concretamente se valoraron, siguiendo la clasificación de Lipsey (cit. en Sánchez-Meca, Rosa y Olivares, 1999), las siguientes características:

- Las **metodológicas** fueron el tipo de estudio, si tenía grupo de comparación, los criterios de selección de la muestra, el uso de azar para la asignación, el tipo de diseño, el tamaño de la muestra, la mortalidad (global y diferencial), las exclusiones posteriores a la agrupación de la muestra a las distintas condiciones, el seguimiento, los momentos de medida, las medidas que aparecen en todos los momentos de registro, las variables dependientes normalizadas, el enmascaramiento del evaluador, del usuario y del formador, la homogeneidad, la definición de constructos, los métodos estadísticos utilizados para inferir los valores perdidos, la especificación del tamaño de efecto, el índice de calidad, el índice estadístico calculado, la significación del resultado y el índice de variabilidad.
- Las características **sustantivas** fueron, **propias del sujeto**, el número de participantes en cada grupo, el número de grupos en el estudio, si hubo exclusiones tras medidas posteriores y el rango y la media de edad de la muestra; **propias del tratamiento**, el periodo de estudio, la intensidad y las unidades; y **propias del contexto**, el área formativa, el campo de intervención y el tipo de publicación.

4°. **Cálculo del tamaño de efecto.** Con el fin de homogeneizar los resultados encontrados en los diferentes estudios (14 con medida tras la intervención y 2 con medida durante un seguimiento tras la finalización de ésta), se calculó para cada uno de ellos el índice de tamaño de efecto “**diferencia media tipificada**” o “**d**”.

En los diseños de medida previa y posterior de un solo grupo, este índice fue calculado con la siguiente fórmula (Wolf, 1986):

$$d = \frac{|\bar{x}_{pre} - \bar{x}_{post}|}{s}$$

donde:

- \bar{x}_{pre} = media de la medida tomada antes de la realización de la acción formativa.
- \bar{x}_{post} = media de la medida tomada inmediatamente después de la realización de la acción formativa o transcurrido un tiempo tras la finalización de ésta en las ocasiones en que la medida fue tomada durante el seguimiento.
- Las || señalan que el numerador se toma en valores absolutos.
- s = desviación típica intragrupo de ambos grupos de datos en conjunto.

Cuando se trató de un diseño de dos grupos (uno control y uno de intervención) con medida únicamente posterior, la fórmula anterior se modificó levemente para su adaptación a estos casos:

$$d = \frac{|\bar{x}_{tratamiento} - \bar{x}_{control}|}{s}$$

Como puede apreciarse, la única diferencia entre ambas fórmulas radica en que en la anterior se restaba las medias de los resultados obtenidos antes y después de la intervención por un mismo grupo mientras que, en ésta última, son las medias de dos grupos diferentes tomadas en un mismo momento las que se restan.

Cuando se trató de un diseño con un grupo control y otro de tratamiento y dos momentos de medida (antes y después de la acción formativa), la fórmula que se utilizó fue la siguiente (Sánchez-Meca, Rosa y Olivares, 1999):

$$d = d^t - d^c = \frac{|\bar{x}_{pre(tratamiento)} - \bar{x}_{post(tratamiento)}|}{s_{(tratamiento)}} - \frac{|\bar{x}_{pre(control)} - \bar{x}_{post(control)}|}{s_{(control)}}$$

En otras palabras, los índices que se restaron fueron:

- d^t = cambio logrado comparando los valores obtenidos en el momento previo y el posterior en el grupo que participó en una acción formativa.
- d^c = cambio logrado comparando los valores obtenidos en el momento previo y el posterior en el grupo control.

Incluir en un meta-análisis varias medidas de un solo estudio puede provocar problemas de dependencia. Para evitar este posible sesgo, se aunaron en sólo una todas las medidas dadas (por ejemplo, de diferentes grupos de personas, de distintas variables dependientes o de cada uno de los ítems de un cuestionario). Dos métodos principales se utilizaron para este fin:

1. Cuando se encontraron en un solo estudio diversas medias con escala común con sus desviaciones típicas y sus tamaños muestrales, se unieron todas las medias previas en sólo una y todas las posteriores también en sólo una, mientras que las desviaciones típicas se promediaron en un único valor. Concretamente, las fórmulas utilizadas para esto fueron las siguientes:

Para calcular una única media de valores en bruto o de otras medias (siempre y cuando la escala fuera la misma), se utilizó la fórmula de la media ponderada,

$$\bar{x} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \dots + x_n w_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

donde:

- las x representan valores concretos o la media de cada grupo.
- las w son los pesos asociados a cada una de las x, siendo concretamente en este caso el número de personas al que hace referencia dicho valor (la n).

En los casos en que fue necesario calcular la desviación típica conjunta cuando se conocían las desviaciones de los grupos por separado, se utilizó la raíz cuadrada de la varianza promediada, cuya fórmula es la siguiente:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1)}$$

donde:

- n_i = tamaño de la muestra de cada grupo.
- s_i^2 = varianza de cada grupo.

Conociendo estos valores ponderados, se dispuso al fin de información suficiente para calcular una sola “d” de cada estudio y, finalmente, la “d” ponderada de todos los estudios.

2. Cuando se dieron varias medidas en un mismo estudio con distinta escala, en primer lugar se calculó una “d” por cada conjunto de comparaciones con escala común y, en segundo lugar, se calculó una única “d” ponderada para cada estudio, teniendo en cuenta los diferentes tamaños muestrales, para que los estudios que incluyeron a más número de personas ejercieran un peso mayor específico en dichos análisis. Concretamente, utilizando el software Comprehensive Meta Analysis, la fórmula para dicho cálculo fue la siguiente (Hedges y Olkin, 1985):

$$d_+ = \frac{\sum w_i d_i}{\sum w_i}$$

donde:

- d_i = diferencia media tipificada de cada conjunto de comparaciones medidas con la misma escala.
- w_i = inversa de la varianza de cada tamaño de efecto.

Finalmente, se pasó del mismo modo al cálculo de una sola “d” para el total de los estudios.

5°. Estudio de la influencia de posibles variables moderadoras.

Una vez calculada la eficacia media de los programas de formación continua y constatado el hecho de que los resultados eran **heterogéneos**, se estudió si algunas características de los trabajos previamente codificadas estaban funcionando como posibles **variables moderadoras**, utilizando para ello el programa Comprehensive Meta Analysis.

3. RESULTADOS.

3.1. Descriptivos.

A continuación, en la tabla 4.1 se presentan las características metodológicas, sustantivas y extrínsecas de los estudios que cumplieron los criterios para ser incluidos en el estudio meta-analítico. La primera columna presenta los diferentes ítems; la segunda, el porcentaje de los casos incluidos en cada categoría; y la tercera, también el porcentaje de casos, pero esta vez obviando los valores perdidos.

ÍTEM	CON V.P. (N=14)	SIN V.P.
CARACTERÍSTICAS METODOLÓGICAS		
0. Tipo de estudio:		(N=11)
▪ Cuasiexperimental	63.64%	63.64%
▪ Experimental	36.36%	36.36%
1. Grupo de comparación:		(N=11)
▪ No hay grupo de comparación	54.55%	54.55%
▪ Inactivo	9.09%	9.09%
▪ Activo	36.36%	36.36%
2. Criterios de selección de la muestra (inclusión/exclusión):		(N=11)
▪ No especificados	36.36%	36.36%
▪ Especificados	63.64%	63.64%
3. Azar:		(N=11)
▪ Pre-experimentales y cuasiexperimentales sin control de vvee; y experimentales con asignación y/o selección aleatoria inadecuados	54.55%	54.55%
▪ Pre-experimentales y cuasiexperimentales con control de variables extrañas; y experimentales con asignación y	45.45%	45.45%

selección aleatoria adecuados		
4. Diseño:		(N=11)
▪ Pre-experimental; experimental con un momento de medida	54.55%	54.55%
▪ Cuasi-experimental con [2-29] medidas	9.09%	9.09%
▪ Experimental con ≥ 2 medidas	36.36%	36.36%
5. Muestra:		(N=11)
▪ $n=[12-40]$	27.27%	27.27%
▪ $n>40$	72.73%	72.73%
6. Mortalidad global:		(N=6)
▪ $\geq 20\%$	27.27%	50%
▪ $0 < N < 20\%$	27.27%	50%
▪ Perdidos	45.46%	---
7. Mortalidad diferencial:		(N=2)
▪ $0 < N < 20\%$	9.09%	50%
▪ 0%	9.09%	50%
▪ Perdidos	81.82%	---
8. Exclusiones posteriores a la agrupación de la muestra a las distintas condiciones:		(N=4)
▪ $0 < N < 20\%$	18.18%	50%
▪ 0%	18.18%	50%
▪ Perdidos	63.64%	---
9. Seguimiento:		(N=9)
▪ No se da seguimiento	54.55%	66.67%
▪ < 6 meses	27.27%	33.33%
▪ Perdidos	18.18%	---
10. Momentos de medida:		(N=11)
▪ Sólo posterior	18.18%	18.18%
▪ Previo y posterior	81.82%	81.82%
11. Medidas que aparecen en todos los momentos de registro:		(N=11)
▪ Falta 1 (siempre que se mida más de 1 variable)	9.09%	9.09%
▪ Todas aparecen en todos los momentos	90.91%	90.91%
12. Variables dependientes normalizadas:		(N=11)
▪ Ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado	81.82%	81.82%
▪ Al menos uno es objetivo o normalizado	18.18%	18.18%
13, 14 y 15. Enmascaramiento del evaluador, usuario y formador		(N=0)
▪ Perdidos	100%	---
16. Homogeneidad de la intervención o del proceso de registro cuando no hay intervención:		(N=11)
▪ Distinto	27.27%	27.27%
▪ Igual	72.73%	72.73%
17. Definición de los constructos:		(N=11)
▪ No se define ninguno	0.8%	0.8%
▪ Se define al menos uno conceptual y/o empíricamente	93.4%	93.4%
▪ Todos definidos conceptual y empíricamente	5.8%	5.8%
18. Métodos estadísticos para inferir los valores perdidos:		(N=7)
▪ Ninguno	63.64%	100%
▪ Perdidos	36.36%	---
19. Tamaño de efecto:		(N=11)
▪ Se especifica (o índices derivables)	100%	100%
20. Índice de calidad:		(N=11)
▪ (5-7]	36.36%	36.36%
▪ (7-9]	63.64%	63.64%
21. Índice estadístico calculado:		(N=11)
▪ Descriptivos	9.09%	9.09%
▪ Comparación grupos	90.91%	90.91%
22. Los resultados mostraron diferencias estadísticamente significativas:		(N=11)
▪ Sí	100%	100%

23. Índice de variabilidad facilitado: ▪ Desviación típica	100%	(N=11) 100%
CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo: ▪ [0-50) ▪ [50-100) ▪ [100-150) ▪ ≥ 150	45.46% 18.18% 9.09% 27.27%	(N=11) 45.46% 18.18% 9.09% 27.27%
25. Número de grupos en el estudio: ▪ 1 ▪ 2 ▪ 4	45.45% 45.45% 9.1%	(N=11) 45.45% 45.45% 9.1%
26. Exclusiones tras medidas posteriores: ▪ No ▪ Sí ▪ Perdidos	9.09% 27.27% 63.64%	(N=4) 25% 75% ---
27. Rango de edad especificado: ▪ No ▪ Sí	54.55% 45.45%	(N=11) 54.55% 45.45%
28. Media de edad: ▪ [30-35) ▪ [35-40) ▪ Perdidos	18.18% 18.18% 63.64%	(N=4) 50% 50% ---
29. Periodo de estudio: ▪ ≤ 6 meses ▪ Perdidos	72.73% 27.27%	(N=8) 100% ---
30. Intensidad del tratamiento/registro ▪ 18 horas/semana ▪ 3 días ▪ 1 día ▪ 2 días ▪ 40 horas/semana ▪ 2 horas/día ▪ Perdidos	9.09% 9.09% 27.27% 9.09% 9.09% 9.09% 27.27%	(N=8) 12.5% 12.5% 37.5% 12.5% 12.5% 12.5% ---
31. Unidades de intervención o de registro ▪ Grupal	100%	(N=11) 100%
32. Área formativa ▪ Prevención de riesgos laborales ▪ Formación ▪ Clínica	9.09% 27.27% 63.64%	(N=11) 9.09% 27.27% 63.64%
33. Campo de formación: destinatarios ▪ Clínica ▪ Formación ▪ Industria ▪ Policía	63.64% 9.09% 18.18% 9.09%	(N=11) 63.64% 9.09% 18.18% 9.09%
34. Tipo de publicación: ▪ Revista	100%	(N=11) 100%

Tabla 4.1. Características descriptivas de los estudios incluidos en el meta-análisis.

De los 11 artículos en estudio, la mayoría presentó las siguientes **características metodológicas**: diseño pre-experimental, sin grupo de comparación ni control de variables extrañas explicitado (por ejemplo, no se da ningún dato sobre enmascaramiento), con medición en dos momentos de medida; sin grupo de comparación; criterios de selección de la muestra explicitados; muestras superiores a 40 personas; mortalidad y exclusiones posteriores a la asignación de la muestra a las distintas condiciones no explicitadas; sin seguimiento; las diferentes medidas aparecen en todos los momentos de registro; aunque no suelen estudiarse variables dependientes

normalizadas u objetivas, al menos un instrumento es un cuestionario o autoinforme estandarizado; homogeneidad en la intervención; aunque no todos los constructos en juego son definidos conceptual y empíricamente, al menos uno sí es definido de una u otra manera; no se utilizan métodos estadísticos para inferir valores perdidos; el tamaño de efecto se especifica en todos los casos o, al menos, se aportan datos suficientes como para realizar el cálculo; el índice de calidad oscila entre 7 y 9 puntos; como índice estadístico, se calculan diferencias entre medias; se dan en todas las ocasiones diferencias estadísticamente significativas en las comparaciones realizadas en el sentido en que se esperaban; y el índice de variabilidad que se aporta es la desviación típica.

En relación a las **características sustantivas**, el número de participantes no suele pasar de 50 y el rango y la media de edad no se suelen especificar; respecto al tratamiento, el periodo de estudios no suele pasar de los seis meses, la intensidad de tratamiento es variada y se trabaja siempre en grupos; respecto al contexto de intervención, en el área clínica es donde más acciones formativas realiza; y en todas las ocasiones, los estudios se encontraron en revistas científicas.

3.2. Eficacia.

A continuación, en la tabla 4.2, se presentan los resultados obtenidos en el estudio de la eficacia de los programas de formación continua recogidos.

La primera columna diferencia entre dos estudios distintos: el de las diferencias encontradas entre el momento previo a la intervención y justo después de finalizada ésta, y el de las diferencias entre el mismo momento previo y un tiempo después del final de la intervención (medida tomada durante el seguimiento).

La segunda columna, la **k**, es el número de estudios que se incluyen; la **d₊** es el nivel global medio de eficacia encontrado; las siguientes dos columnas corresponden al límite inferior (**Li**) y límite superior (**Ls**), con un nivel de confianza del 95%.

La antepenúltima columna (**Q_{wj}**) muestra la prueba de homogeneidad intra-grupo. Cuanto mayor sea este valor, mayor es la heterogeneidad existente entre los tamaños de efectos de los distintos estudios. La penúltima columna (**gl**) muestra los grados de libertad (k-1).

Por último, la **p** es la significación de la prueba de homegeneidad. Trabajando con un nivel de confianza del 95%, un valor mayor o igual a 0.05 conllevaría la homogeneidad de los tamaños de efecto de los estudios, mientras que un valor menor a 0.05 se interpretaría como que las diferencias encontradas entre los distintos tamaños de efecto son significativas y, por tanto, existe heterogeneidad.

ESTUDIOS	k	d₊	Li	Ls	Q_{wj}	gl	p
1. Pre-post	14	0.434	0.392	0.477	167.037	13	0.000
2. Pre-segu.	2	0.514	0.373	0.655	14.138	1	0.000

Tabla 4.2. Nivel global medio de eficacia en el posttest y en el seguimiento.

Un total de 2221 personas participaron en las diferentes acciones formativas que se recogieron en los 11 estudios, que pasaron a ser 16 independientes pues, en varias

ocasiones, en un solo artículo se presentaron diversos grupos que participaban en diferentes acciones formativas.

De esos 16 estudios independientes, en 14 la medida posterior a la intervención se tomó justo al finalizar ésta, mientras que en los 2 restantes la medida se tomó durante el seguimiento.

Los tamaños de efecto de todos los estudios se obtuvieron en el sentido positivo, como se esperaba; es decir, con cambios a mejor tras intervenir en comparación con el momento anterior, o con diferencias más amplias en el grupo de intervención que en el control entre antes y después del programa.

Al medir justo después de la intervención, según los resultados obtenidos en la tabla 4.2, la eficacia de la formación resultó ser significativa [$d_+ = 0.434$; $\rho = 0.000$]; tras el periodo de seguimiento, la eficacia global fue incluso mayor [$d_+ = 0.514$; $\rho = 0.000$].

Atendiendo ahora al estadístico Q, se puede concluir que existe un alto grado de heterogeneidad en los tamaños de efecto de los diferentes estudios, sobre todo en la comparación entre antes de la acción formativa e inmediatamente después de finalizada ésta [$Q = 167.037$; $\rho = 0.000$]. Es por ello que a continuación se pasa a estudiar las posibles variables que puedan estar influyendo en esta variabilidad.

3.3. Variables moderadoras.

Una vez se comprobó la eficacia de los programas de formación continua y que el tamaño de efecto difería significativamente entre los distintos estudios, se tomaron las características previamente estudiadas (ver tabla 4.1), con la intención de detectar posibles variables que estuvieran afectando a la eficacia de las acciones formativas.

Sin embargo, para este fin sólo se tomaron aquellas variables que mostraron varios casos en diferentes condiciones, puesto que no tiene sentido ni es posible buscar diferencias cuando la variable es una constante. Además, a partir de ahora sólo se trabajará con los datos obtenidos justo finalizada la intervención y no los estudios de seguimiento, por recoger el número mayor de estudios (14 vs 2) y haber obtenido un grado de heterogeneidad mayor ($Q_{wj} = 167.037$ vs $Q_{wj} = 14.138$).

En la tabla 4.3 se muestran los resultados de las variables cualitativas, mientras que en la 4.4 se presenta el estudio de las variables cuantitativas (Ato y Vallejo, 2007; Camacho, López y Arias, 2006-07).

En la tabla 4.3, la primera columna presenta la variable en estudio y la segunda, cada una de las condiciones; la tercera columna es el número de estudios que se incluyó en cada condición (**k**); la columna de las d_+ presenta los tamaños de efecto globales de todos los estudios incluidos en una sola condición.

La antepenúltima columna (Q_B) hace referencia a la prueba de homogeneidad inter-grupos. Cuanto más alejado de “0” esté este valor, mayor será la heterogeneidad entre los grupos, lo que indicaría que las distintas condiciones de la variable en estudio

presentan tamaños de efecto diferentes, y esto implicaría que parte de la variabilidad encontrada es explicada por esta variable. Estas diferencias son estadísticamente significativas cuando la significación (ρ , en la última columna), es menor que 0.05 (trabando con un nivel de confianza del 95%).

La penúltima columna son los grados de libertad, correspondientes al número de condiciones o grupos de comparación menos 1.

Por otro lado, la tabla 4.4 presenta, en la primera columna, la variable en estudio; en la segunda, se recoge el número de estudios incluidos; las dos siguientes columnas hacen referencia a la significación del modelo; una Q_R elevada implica un mayor ajuste del modelo; una ρ menor que 0.05 implica que dicho ajuste es estadísticamente significativo; entre paréntesis, se muestran los grados de libertad.

Del mismo modo, las últimas dos columnas presentan la suma de cuadrados ponderada del error (Q_E), los grados de libertad entre paréntesis (gl) y la significación (ρ).

VARIABLE	CONDICIONES	k	d ₊	Q _B	gl	ρ
CARACTERÍSTICAS METODOLÓGICAS						
Tipo de estudio	Cuasiexperimental	7	0.309	12.327	1	0.000
	Experimental	7	0.482			
Grupo de comparación	No hay	6	0.294	13.958	2	0.001
	Inactivo	1	0.488			
	Activo	7	0.482			
Criterios de selección	No especificados	4	0.675	8.257	1	0.004
	Especificados	10	0.418			
Azar	Pre y cuasi sin control de vvee; experimentales inadecuados	6	0.294	13.956	1	0.000
	Pre y cuasi con control de vvee; experimentales adecuados	8	0.482			
Diseño	Pre; exp con un momento de medida	6	0.294	13.958	2	0.001
	Cuasi con [2-29] medidas	1	0.488			
	Experimental con ≥ 2 medidas	7	0.482			
Muestra	n=[12-40]	5	0.403	0.077	1	0.781
	n>40	9	0.436			
Mortalidad global	≥20%	5	0.177	55.017	1	0.000
	0<N<20%	3	0.956			
Exclusiones tras la asignación	0<N<20%	2	0.867	5.015	1	0.025
	0%	4	0.459			
Seguimiento	No se da seguimiento	9	0.459	4.479	1	0.034
	< 6 meses	3	0.343			
Momentos de medida	Sólo posterior	4	0.455	2.161	1	0.142
	Previo y posterior	10	0.385			
Variables dependientes normalizadas	Al menos uno cuestionario o autoinforme estandarizado	9	0.412	28.637	1	0.000
	Al menos uno objetivo o normalizado	5	1.035			
Homogeneidad de la intervención	Distinto	4	0.498	0.536	1	0.464
	Igual	10	0.43			
Definición de los constructos	Al menos uno	13	0.493	41.321	1	0.000
	Todos	1	0.095			
CARACTERÍSTICAS SUSTANTIVAS						
Del sujeto						
Exclusiones tras medidas posteriores	No	3	0.54	6.341	1	0.012
	Sí	3	0.956			

Rango de edad especificado	No	9	0.632	11.726	1	0.001
	Sí	5	0.406			
Del contexto						
Área formativa	PRL	1	1.037	25.561	2	0.000
	Formación	3	0.434			
	Clínica	10	0.356			
Campo de formación: destinatarios	Clínica	10	0.356	27.031	3	0.000
	Formación	1	0.452			
	Industria	2	0.801			
	Policía	1	0.118			

Tabla 4.3. ANOVA de las variables moderadoras cualitativas.

VARIABLE	k	Q _R (gl)	p	Q _E (gl)	p
CARACTERÍSTICAS METODOLÓGICAS					
Índice de calidad	14	3.181 (1)	0.074	163.857 (12)	0.000
CARACTERÍSTICAS SUSTANTIVAS del sujeto					
Número de participantes en cada grupo	14	6.129 (1)	0.013	160.909 (12)	0.000
Número de grupos en el estudio	14	5.452 (1)	0.019	161.586 (12)	0.000
Media de edad	4	17.64 (1)	0.000	1.137 (2)	0.566

Tabla 4.4. Análisis de regresión de las variables moderadoras cuantitativas.

Las variables donde se encontraron resultados significativos fueron las siguientes:

De entre las **características metodológicas**:

- El **tipo de estudio**: los diseños experimentales presentaron mayor eficacia que los cuasi-experimentales.
- El **grupo de comparación**: la mayor eficacia se encontró cuando había grupo de control inactivo, con un tamaño de efecto muy cercano al obtenido por los estudios con grupo de control activo.
- Los **criterios de selección**: cuando no se especificaron, la eficacia fue mayor.
- El **azar**: los diseños pre-experimentales y cuasiexperimentales con control de variables extrañas, y los experimentales con selección y asignación aleatorias adecuadas resultaron ser más eficaces que los diseños pre-experimentales y cuasiexperimentales sin control de variables extrañas y experimentales sin selección y/o asignación aleatorias adecuadas.
- El **diseño**: los diseños cuasiexperimentales con mínimo 3 medidas y máximo 30, y los diseños experimentales con más de dos medidas, fueron más eficaces que los pre-experimentales o experimentales con un solo momento de medida.
- La **mortalidad global**: los estudios que presentaron mayor grado de mortalidad (más del 20%) fueron menos eficaces que aquellos que mostraron un nivel medio de mortalidad (menos del 20%).
- Las **exclusiones tras la asignación**: los estudios que presentaron un nivel medio de exclusión tras la asignación de la muestra a los diferentes grupos fueron más eficaces que los que no presentaron ninguna exclusión.

- El **seguimiento**: presentaron mayor eficacia los estudios sin seguimiento que aquellos que tuvieron un seguimiento de corta duración.
- Las **variables dependientes normalizadas**: cuando al menos una de las variables dependientes medidas fue objetiva o normalizada, la eficacia fue mayor que cuando a lo sumo fueron cuestionarios o autoinformes estandarizados los instrumentos con los que se recogió el dato.
- La **definición de los constructos**: cuando faltó algún constructo por definir, la eficacia de las intervenciones fue mayor que cuando todos fueron definidos.

Respecto a las **características sustantivas del sujeto**:

- Las **exclusiones tras medidas posteriores**: los estudios que presentaron exclusiones una vez comenzada la intervención fueron más eficaces que aquellos en los que no se dieron exclusiones tras medidas posteriores.
- El **rango de edad**: el tamaño de efecto global de los estudios donde no se había especificado el rango de edad fue mayor que el de los estudios donde sí se había explicitado este dato.
- El **número de participantes en cada grupo**: esta variable mostró una relación inversa con el tamaño de efecto, por lo que cuantas más personas conformaron cada grupo, menos eficaz resultó ser la intervención.
- El **número de grupos en el estudio**: cuantos más grupos conformaron el estudio, mayor fue la eficacia de la intervención.
- La **media de edad**: esta variable es la que mejor ajuste presenta de entre las variables cuantitativas, si bien es cierto que sólo 4 de los 14 estudios explicitaron el dato. Se encontró una relación directa con el tamaño de efecto, por lo que cuanto más edad de media tuvieran los participantes a las acciones formativas, más eficaces resultaron ser éstas.

Respecto a las **características sustantivas del contexto**:

- El **área formativa**: el estudio que mostró mayor eficacia fue el encuadrado en el área de prevención de riesgos laborales; los del área de formación fueron los segundos más eficaces, quedando en último lugar la formación en clínica.
- El **campo de formación**: acorde con los resultados obtenidos en la variable anteriormente comentada, los estudios más eficaces fueron los dirigidos a la industria, seguidos de la formación y posteriormente la clínica.

4. DISCUSIÓN.

La muestra de estudios tomada para este meta-análisis parece ser representativo del total de estudios encontrados en el sentido en que las características que mayoritariamente presentan son coincidentes.

Sin embargo, un aspecto diferenciador es el descenso en el número de valores perdidos debido principalmente a la no inclusión de estudios teóricos, aunque aún quedó la codificación de algunos ítems sin especificar, como fueron la mortalidad, la realización o no de enmascaramiento, las exclusiones posteriores a la asignación de la muestra a los distintos grupos y la edad de los participantes.

Además, al comparar la tabla 3.1 con la 4.1, se vislumbra que algunos otros ítems mostraron ciertas variaciones al comparar los 11 estudios incluidos en el meta-análisis con el total de estudios sobre formación continua encontrados en la literatura:

- Los criterios de selección de la muestra fueron más especificados que no especificados entre los 11 trabajos incluidos en el meta-análisis, mientras que en el total de estudios encontrados (121), este dato se dio al contrario. Esta diferencia puede deberse a que los estudios que pudieron incluirse en el meta-análisis presentarían más información que los que no cumplieron los criterios, en vista de que uno de ellos era precisamente el nivel de disponibilidad de los datos.
- En líneas generales, los estudios sólo presentaron medida posterior mientras que, entre los 11 estudios analizados, la mayoría presentaron medidas previas y posteriores. Esto se debe a que precisamente el hecho de tener como mínimo un par de medidas fue un requisito indispensable para ser incluido en el meta-análisis.
- Por la misma razón, todos los estudios incluidos presentaron tamaño de efecto o, en su defecto, los datos necesarios para calcularlo, cuando en el total de estudios encontrados esto apenas ocurría.
- Por último, y en la misma línea, el porcentaje de estadísticos de comparación superó a los descriptivos, cuando en el total de estudios esto se daba al contrario. De nuevo, el hecho de presentar una aportación mínima de datos se hizo necesario para que el estudio se incluyera en el meta-análisis.

Antes de pasar a comentar los resultados encontrados, un aspecto a discutir es si la muestra obtenida de estudios acerca de formación continua es adecuada o el número de estudios incluidos es excesivamente bajo como para poder generalizar los resultados; otra cuestión es que sólo se tomaron datos provenientes de fuentes primarias y secundarias y no de fuentes informales, por lo que se estarían obviando los estudios no publicados.

Aunque la búsqueda realizada fue lo más amplia y exhaustiva posible, está comprobado que la inclusión en un meta-análisis sólo de estudios publicados puede sobrestimar la magnitud del efecto e incluso variar su dirección (Rosenthal, 1979). Es lo que se denomina **sesgo de publicación**: la tendencia a publicar aquellos estudios en los que se encontró un cambio y no publicar aquello que no presentó relación estadística entre las variables en estudio. De hecho, según los resultados obtenidos tras la aplicación de la escala en el ítem 22 (ver capítulo anterior) se pudo comprobar que, de los 43 estudios de formación continua en que se obtuvo el dato, en el 74.4% (casi tres

cuartos del total) mostraron diferencias estadísticamente significativas; y los 14 estudios incluidos en el meta-análisis presentaron resultados significativos a favor de la intervención.

Una aproximación matemática para determinar el número de resultados nulos (no significativos) no recuperados necesarios para que la significación global del estudio meta-analítico pudiera deberse al sesgo de publicación es la siguiente (Orwin, 1983):

$$N = \frac{k(\bar{d} - d_c)}{d_c - d_{fs}}$$

donde:

- k = número de estudios.
- \bar{d} = tamaño de efecto medio.
- d_c = tamaño de efecto criterio. Según Cohen (cit. en Sánchez-Meca y Ato, 1989), 0.2 se interpretaría como un valor “pequeño” de este índice, 0.5 como “medio” y 0.8 como “grande”.
- d_{fs} = tamaño de efecto medio para los estudios no recuperables.

Cuanto mayor sea la N, más positivo será el resultado, puesto que más trabajos no significativos tendrían que haberse obviado para que la significación encontrada se debiera al sesgo de publicación. Por tanto, la conclusión de un meta-análisis es tanto más sólida cuanto mayor es el número de estudios no significativos y no publicados necesarios para refutarla.

Para un tamaño de efecto criterio de 0.2 (que podría considerarse el menos estricto puesto que, de los tres valores propuestos, es el que aporta una N mayor) y un tamaño de efecto medio para los estudios no recuperables de 0 (suponiendo que fuera nulo), la fórmula en el caso del meta-análisis aquí presentado se resolvería del siguiente modo.

$$N = \frac{14(0.434 - 0.2)}{0.2 - 0} = 16.38$$

Esto implica que con tal de que no se hayan encontrado 16 trabajos sobre formación continua no publicados con resultados no significativos, las conclusiones del meta-análisis descrito podrían tener como causa el sesgo de publicación. Este resultado negativo es debido, por un lado, a que el tamaño de efecto encontrado no es excesivamente grande y, por otro lado y fundamentalmente, a que se incluyó en el meta-análisis un número de estudios excesivamente bajo.

Otro aspecto que podría discutirse es la adecuación de los **criterios de inclusión** utilizados a la hora de recopilar los estudios que formaron parte del posterior análisis.

Algunos investigadores afirman que, para obtener resultados útiles y no conclusiones erróneas de un meta-análisis, los estudios a incluir han de limitarse a aquellos que presenten un diseño experimental (en el que se comparen grupo control y grupo de intervención) con un alto nivel de calidad, principalmente porque la inclusión de estudios con baja calidad provocan una sobreestimación del efecto (Chacón, Sanduvete, Sánchez y Sánchez-Meca, 2007a):

Sin embargo, restringir el uso del meta-análisis exclusivamente a estas circunstancias hace que, en ámbitos de aplicación como la psicología, pierda utilidad puesto que muy pocos estudios presentan estas características, tal como se pudo constatar en el capítulo anterior, ya que los diseños experimentales en absoluto son los que preponderan y la calidad no suele ser especialmente elevada:

Respecto al número de estudios con **diseño experimental**:

- Según el ítem 0 referido al tipo de estudio, sólo presentó este diseño el 7.4% de los trabajos.
- Según el ítem 1, sólo hubo grupo de comparación en el 29.4% de los estudios empíricos recogidos.
- Según el ítem 3, sólo en el 26% de los casos se utilizó una técnica de azar adecuada en la selección de la muestra y/o su asignación a los distintos grupos.
- Según los ítems 13, 14 y 15, en casi ningún caso se utilizó el enmascaramiento, técnica de control de variables extrañas propia de los diseños experimentales, en ninguna de sus modalidades (a usuarios, evaluadores y formadores). En un gran número de ocasiones, ni tan siquiera se mencionó (se presentó un alto porcentaje de no especificación).

Respecto a la **calidad** de los estudios encontrados, algunos datos pueden resultar esclarecedores:

- Según el ítem 9, sólo en el 13% de los estudios empíricos recogidos, el seguimiento se realizó durante al menos un año.
- Según el ítem 10, sólo en el 50% de los estudios empíricos encontrados se daba al menos dos medidas a lo largo del tiempo.
- Según el ítem 12, sólo en el 33.4% de los estudios empíricos recogidos al menos una variable dependiente estudiada era objetiva o normalizada.
- Según el ítem 17, sólo en el 5.8% de los casos fueron definidos empírica y conceptualmente todos los constructos en estudio.
- Según el índice de calidad calculado (ítem 20), sólo en el 3.4% de los casos el estudio superó el 9, cuando la puntuación mayor posible de obtener fue 19.

Si a esto se le añade que sólo en el 19% de los estudios empíricos con más de una medida se aportó información suficiente como para calcular el tamaño de efecto

(ítem 19) se puede concluir que, tomando únicamente los estudios experimentales con un adecuado nivel de calidad, la muestra obtenida sería mínima.

Por tanto, para no desaprovechar las posibilidades del meta-análisis en el ámbito de la psicología, se recomienda no excluir los estudios de baja calidad y/o con un diseño distinto al experimental.

Para estudiar la influencia del nivel de calidad o del tipo de diseño en los resultados una posibilidad, tal como se ha hecho en esta ocasión, es incluir ambas variables como posibles moderadoras.

En definitiva, lo expuesto podría resumirse en dos ideas: que la inclusión de estudios con distinto diseño y diferentes niveles de calidad está justificada; y que, a pesar de esto, el número de estudios encontrados fue tan bajo que las conclusiones a las que se ha llegado no pueden ser generalizadas sin que queden dudas de si el resultado se debió al sesgo de publicación.

Teniendo en cuenta por tanto el hecho de que estos resultados habrían de tomarse con cierta precaución, en resumen puede decirse que las intervenciones en formación continua resultaron ser eficaces tanto a corto como a largo plazo, que los tamaños de efecto de los estudios resultaron ser heterogéneos entre sí y que la mayoría de las variables estudiadas resultaron ser moderadoras, en el sentido esperado.

Así, por ejemplo, respecto al **tipo de diseño**, Shadish y Ragsdale (1996) encontraron diferencias entre los resultados obtenidos con los estudios experimentales en comparación con otros diseños, aunque no especialmente significativas como cabría esperar. En este estudio se encontró algo semejante: los diseños experimentales fueron los que obtuvieron un mayor tamaño de efecto, aunque no muy alejado del obtenido por los diseños cuasiexperimentales con un buen nivel de calidad.

Por otro lado, no se encontró la relación esperada entre algunas variables; por ejemplo, el **grado de calidad** no correlacionó significativamente con el tamaño de efecto, cuando en la literatura se encuentran ocasiones en las que la calidad del diseño afecta negativamente, ya que los diseños de alta calidad suelen controlar en mayor medida los posibles efectos de las expectativas del experimentador, por lo que los índices del tamaño de efecto suelen ser más bajos que en los estudios de baja calidad (Sánchez-Meca, Rosa y Olivares, 1999).

Por último, algunas variables que explicaban en parte la variabilidad entre los tamaños de efecto resultaron ser significativas, pero en un sentido diferente al encontrado.

Concretamente, respecto a los **criterios de selección**, se halló que la eficacia fue mayor cuando no fueron explicitados; en relación a la **definición de los constructos**, se encontró que la eficacia fue mayor cuando faltó alguno por definir que cuando todos fueron definidos; y respecto al **rango de edad**, el tamaño de efecto global de los estudios donde no se había especificado este dato fue mayor que el de los estudios donde sí se había explicitado. Tras estos resultados, parece ser que la falta de especificación de algunos datos, aunque sí afecta a la calidad del reporte y a la hora de

incluir los estudios en revisiones sistemáticas y meta-análisis, no necesariamente afecta a la eficacia de la formación.

En esta misma dirección, cuando hubo un número no muy elevado de **exclusiones tras la asignación de los participantes a los distintos grupos** y/o una vez **comenzada la intervención**, la eficacia fue mayor que cuando no hubo exclusiones en absoluto; y cuando no hubo **seguimiento**, la eficacia fue mayor que cuando se dio un seguimiento de corta duración. Esto puede que sea debido a que, con valores moderados, estas variables no afectan negativamente a la eficacia de la formación, obteniendo incluso resultados muy positivos. Sería interesante estudiar qué ocurriría con valores extremos.

La variable **número de participantes en cada grupo** mostró una relación inversa con el tamaño de efecto, por lo que cuantas más personas conformaron cada grupo, menos eficaz resultó ser la intervención. Aunque en principio pueda parecer un resultado contradictorio o no esperado, realmente tiene sentido puesto que, aunque es verdad que la formación en grupo tiene ciertas ventajas respecto a la individual, también es cierto que, con un número elevado de participantes, parte de estas ventajas se pierden: el nivel de participación descende, al profesor le es más difícil adaptarse al nivel de conocimiento presentado por el grupo, la atención se dispersa, etc.

Una última cuestión que pudo estar afectando a todos estos casos no esperados fue el hecho de que, en ocasiones, algunas condiciones incluyeron un número muy bajo de estudios, siendo en ocasiones sólo 1.

A partir de las variables moderadoras encontradas, principalmente de carácter metodológico, podrían proponerse las siguientes **recomendaciones** para aumentar la eficacia de los programas de formación continua:

- Utilizar un **diseño experimental**.
- Utilizar un **grupo de comparación**, ya sea activo o inactivo.
- Evitar la **mortalidad** de los participantes en el estudio.
- Utilizar en la medida de lo posible **variables dependientes normalizadas**, teniendo en cuenta fundamentalmente la validez del dato recogido.
- Tomar una **muestra** de participantes de tamaño mediano, ni excesivamente grande ni excesivamente pequeña.
- No centrar la formación exclusivamente en personas jóvenes puesto que los **mayores**, a pesar de la experiencia, también tienen necesidad de reciclarse, suelen presentar un alto grado de motivación en la formación y, generalmente, sacan partido de las acciones formativas.

5. CONCLUSIONES.

Las principales **ventajas** que presentan los meta-análisis sobre otras técnicas son las siguientes (Sánchez-Meca, 2000):

- **Eficiencia:** se pueden sintetizar grandes cantidades de información sin mucha dificultad para obtener conclusiones no excesivamente complejas, con lo que pueden resultar útiles.
- Encontrar **contradicciones** en los resultados no sólo no es un problema, sino que da la posibilidad de obtener conclusiones más ricas, ya que se puede explicar la heterogeneidad hallada con los moderadores.
- Al detectarse **variables moderadoras**, se están tomando en consideración aspectos en los que incidir para aumentar la eficacia de las intervenciones, tal como en este caso se presentó en el apartado de discusión.
- **Potencia estadística mayor:** al trabajar con diversos estudios, el tamaño de la muestra aumenta y, con ello, se incrementa la potencia estadística, definida como la capacidad de encontrar significación en el modelo cuando realmente es significativo, disminuyendo de este modo el error tipo II o β , que sería no encontrar significación cuando realmente el modelo era significativo.
- **Alto grado de precisión:** el meta-análisis tiene capacidad para detectar efectos pequeños, imperceptibles quizá para otros tipos de análisis, pero no por ello irrelevantes a nivel práctico y/o social.
- **Estimación del tamaño del efecto:** los resultados se obtienen a nivel de magnitud y no únicamente a nivel de significación como ocurre, por ejemplo, con otro tipo de pruebas de comparación de grupos.
- **Replicabilidad:** al ser un análisis con un alto grado de rigor científico, cumple con las normas de replicabilidad de la investigación, por lo que las conclusiones a las que se llega con un meta-análisis pueden ser contrastadas con otro para su confirmación o refutación.
- **Mayor certeza:** Al basarse en una metodología cuantitativa objetiva, las conclusiones obtenidas se sostienen con mayor seguridad.

Por todas estas ventajas es tan interesante la realización de meta-análisis. Sin embargo, como se vio en este caso, muy pocos estudios pudieron formar parte de éste. Por esta razón, como ya se ha comentado, desde aquí se aboga por la flexibilidad en los criterios de inclusión de los estudios (aunque sin llegar a la laxitud), por el fomento de su uso y el aumento de su utilidad en ámbitos para los que quizá inicialmente no se consideró un método adecuado.

Pero no todo son puntos a favor. Algunas **limitaciones** que presenta este análisis se enumeran a continuación (Sánchez-Meca, 2000):

- **Dependencia:** es el sesgo que se da cuando un solo estudio aporta más de una estimación de efecto porque, como de las mismas personas se incluyen distintos resultados, no se da el supuesto de independencia de los errores. Para evitar esto, en

este trabajo se promediaron los distintos valores de un solo estudio, con lo que se logró que cada uno de ellos aportara únicamente una estimación de efecto.

- **Representatividad:** los estudios localizados pueden no representar correctamente a la población de estudios debido al sesgo de publicación. En este estudio concreto es muy probable que se esté incurriendo en este problema, tal como se argumentó anteriormente. Por esta razón se recomendó tomar las conclusiones obtenidas con precaución.
- **Inclusión de información errónea:** datos e información defectuosa o deficiente en los estudios primarios pueden afectar a la calidad del meta-análisis. Al no tenerse acceso a más información que la aportada en los artículos, no se puede en este caso constatar si ha podido estar influyendo este factor en los resultados.
- **Heterogeneidad:** estudios diferentes no deberían ser integrados. Es probable que en este caso se haya tocado este punto débil porque, por ejemplo, se incluyeron todas las acciones formativas encontradas, independientemente de la variable dependiente que se midiera o del ámbito de intervención concreto. Para resolver cuestiones como la última comentada, se consideró el contexto como una posible variable moderadora, explicativa de la heterogeneidad encontrada.

Para solventar estos puntos débiles en los que se incurrió, algunas **mejoras** a proponer para el futuro son las siguientes:

Tratando de alcanzar una mayor representatividad de los estudios incluidos en el meta-análisis y con la finalidad de hacer estudios más homogéneos, se hace necesario **ampliar el número de estudios a incluir**. Algunas propuestas para conseguirlo son ampliar el número de bases de datos en la búsqueda inicial para recoger más trabajos publicados y obtener datos de otras fuentes para recoger además trabajos no publicados. De este modo, habría menos probabilidad de estar incurriendo en el sesgo de publicación.

Respecto a la posibilidad de estar recogiendo información errónea, se proponen principalmente dos mejoras, una para los investigadores que realizan el meta-análisis y otra para los profesionales que ejecutan y/o redactan los trabajos primarios:

- Una carencia en este trabajo fue la falta de **estudios de fiabilidad** respecto a la codificación y el análisis de las estimaciones del efecto. Como mejora, se propone la realización de dichas pruebas de fiabilidad, para lo que haría falta que al menos un investigador más realizara la codificación y el análisis.
- Para los investigadores de los estudios primarios, se recomienda que aporten el mayor número posible de **datos relevantes** para que dicho estudio pueda incluirse en el meta-análisis y para poder hacer además el análisis de variables moderadoras.

En suma podría afirmarse que, por los resultados obtenidos, se recomienda la realización de acciones formativas con diseños experimentales (con asignación aleatoria y grupo control), con control de posibles variables extrañas influyentes, sin un nivel elevado de mortalidad global, sin un número muy elevado de exclusiones tras la asignación, con al menos una variable dependiente normalizada u objetiva, con los

constructos definidos conceptual y operativamente, y con un número moderado de participantes de cualquier edad.

En definitiva, la eficacia de los programas de formación encontrada en este estudio puede considerarse como una prueba más de la importancia de la formación para la mejora del desempeño en el puesto de trabajo de los empleados y para el fomento de su desarrollo personal.

2ª PARTE:

**INNOVACIONES METODOLÓGICAS APLICADAS
A CADA UNA DE LAS FASES DE EVALUACIÓN**

La parte hasta ahora presentada ha hecho referencia a las cuestiones más teóricas y generales que en este trabajo van a tratarse: en primer lugar, se hizo una recogida de los modelos más comúnmente señalados en la literatura acerca de evaluación de programas en general y, más concretamente, de evaluación de la formación continua.

En el segundo capítulo, tras constatar la idea generalizada de que los programas de formación adolecen de cierto rigor metodológico, se presentó la elaboración de una escala de calidad de los estudios primarios. Tras el paso por ciertas fases, la última versión realizada de dicho instrumento fue aplicada a los estudios referidos a la formación continua que se encontraron en la bibliografía, de tal manera que se recogió una prueba empírica de que realmente la calidad metodológica de los estudios oscilan entre valores medios y bajos. A partir de estos resultados, se trataron de proponer mejoras que pudieran contribuir al mayor éxito de estos programas de formación.

Finalmente, en el capítulo 4, con la intención de aportar una prueba más de la posible relación entre calidad metodológica y éxito en el programa, se realizó un estudio meta-analítico con los pocos trabajos que aportaron suficientes datos para ello. De este modo se pudo constatar, aunque no con alto grado de seguridad por la muestra tan pequeña, que las acciones formativas resultaron ser eficaces, que se encontró heterogeneidad en los datos y que variables de tinte metodológico como el tipo de diseño o el grado de mortalidad influyeron en la eficacia de las intervenciones.

A partir de ahora, se tratará de presentar innovaciones metodológicas en las distintas fases de evaluación de manera más aplicada, haciendo referencia concretamente a tres instituciones públicas con las que nuestro grupo de investigación HUM-649 denominado “Innovaciones metodológicas en evaluación de programas” (grupo oficial de investigación del Plan Andaluz de Investigación de la Junta de Andalucía) ha firmado convenios de colaboración y contratos de investigación con los que nos comprometíamos a llevar a cabo la evaluación de sus programas de formación. Dichos organismos son **Diputación de Sevilla**, el **Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla (FORPAS)** y el **Instituto Andaluz del Deporte de la Consejería de Turismo, Comercio y Deporte de la Junta de Andalucía (IAD)**.

Diputación de Sevilla es un organismo que funciona a nivel provincial y tiene una inmensa disparidad de funciones, por lo que sus más de 14.000 trabajadores cumplen muy diferentes tareas y sus usuarios son también realmente diversos. Así, por ejemplo, cubre parte de los servicios comunitarios, realiza formación ocupacional (en especial, existen muchos talleres para menores de edad que dejaron los estudios reglados), tiene su propio cuerpo de bomberos, colabora con la policía, etc. La Sección de Formación Continua de Recursos Humanos de esta institución tiene como objetivo que sus empleados y empleadas mejoren el desempeño en sus puestos de trabajo. La Universidad de Sevilla, a través del grupo de investigación HUM-649, lleva colaborando con este organismo ininterrumpidamente desde el año 1999.

FORPAS es un centro que forma parte de la Universidad de Sevilla, cuya función principal es la formación de su Personal de Administración y Servicios (Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999). Los profesionales que reciben formación en este caso presentan un perfil más homogéneo

que en Diputación de Sevilla ya que, por ejemplo, todos trabajan en Sevilla capital y todos se dedican a la administración y servicios de la Universidad. La formación que reciben principalmente está centrada en cuestiones informáticas. Además, el número de potenciales usuarios es considerablemente menor. El grupo de investigación HUM-649 lleva colaborando en la evaluación de los programas de formación de este organismo desde el año 2000.

El **IAD** es una institución que funciona por toda Andalucía cuya función es el fomento y desarrollo de la actividad deportiva en la comunidad autónoma. Propone mejorar de manera continuada los servicios que se ofrecen a la ciudadanía, para lo cual intenta mejorar la eficiencia en el trabajo de sus empleados y empleadas, aumentar la productividad de los mismos y cubrir necesidades y mejorar la oferta de acciones formativas destinadas a los sectores relacionados.

Como se puede apreciar las tres organizaciones, a pesar de ser muy dispares en muchos aspectos, presentan una idea común: tratar de mejorar el desempeño de sus trabajadores en sus ocupaciones, con la intención de que dicha mejora redunde en un aumento en la calidad del servicio prestado al usuario.

Como recordatorio, cabe señalar que el modelo utilizado para la estructuración del trabajo a partir de ahora responde al proceso de evaluación no lineal descrito por Anguera y Chacón (en prensa, b), Alvira, (1991) y el Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla (1999) que puede esquematizarse tal como aparece en la siguiente figura:

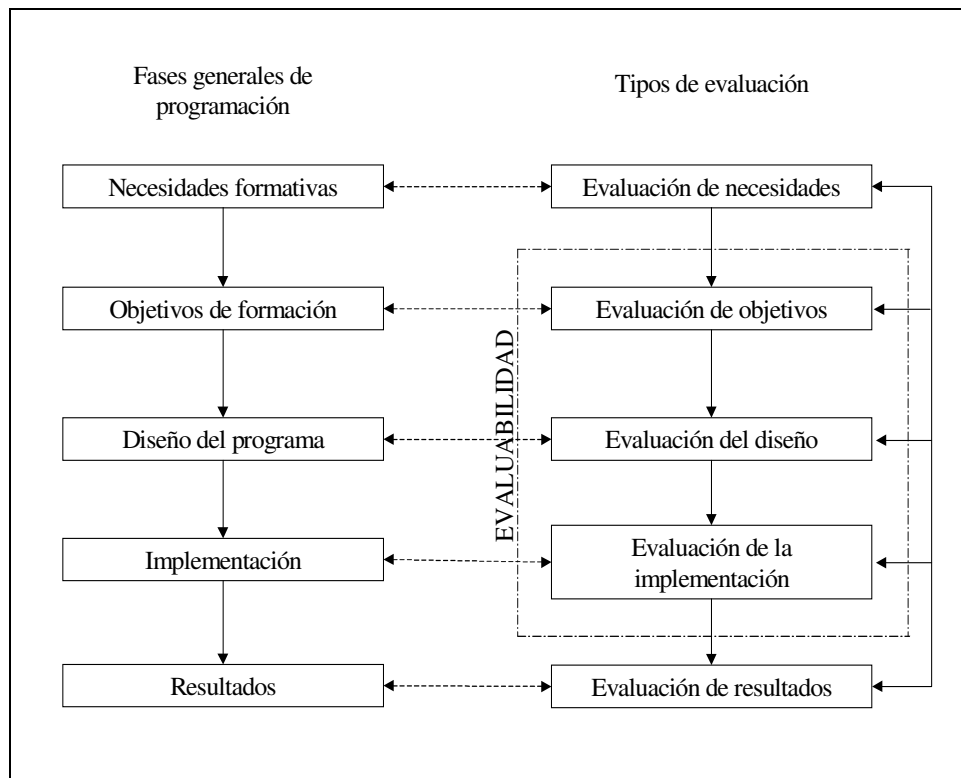


Figura 1. Esquema no-lineal de evaluación (Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999).

De este modo, las distintas fases de evaluación de un programa pueden encuadrarse en tres momentos: **antes** de la intervención, **durante** ésta y **al final**:

- **Antes** de la intervención:
 - La **detección de necesidades** supone el punto de partida de cualquier programa, pues trata de determinar qué aspectos han de ser modificados.
 - Los **objetivos** enumeran lo que se pretende conseguir con el programa.
 - El **diseño** es el plan de trabajo que posteriormente se va a llevar a cabo.
- **Durante** la intervención:
 - La **implementación** es la puesta en práctica del programa previamente diseñado.
- **Después** de la intervención:
 - Los **resultados** suponen la medición de lo que se ha obtenido con la intervención. Es la forma de determinar empíricamente si se lograron los objetivos marcados, si se alcanzaron metas no previstas y/o con qué relación coste-beneficio.

En general, como ya se ha comentado, prepondera la evaluación orientada hacia los resultados (Cascio, 1989), con lo que se trata de dar pruebas de los beneficios sobre todo económicos de los programas, pero no se obtiene mucha información acerca del proceso. Por esta razón, se opta por utilizar este modelo; para tratar de recoger datos en todo momento, lo que proporcionará la posibilidad de realizar cambios en el momento de la detección del problema y aumentar así la eficacia.

Como se verá posteriormente, no en todas las fases se trata con la misma intensidad el proceso de evaluación de las tres instituciones en estudio. Esto es debido a que, por cuestiones generalmente políticas y organizativas, no ha sido posible llevar a cabo el desarrollo de todas las etapas de evaluación con la misma profundidad, así que servirán como estandarte en cada momento aquella institución que muestre un mayor nivel de consolidación. Concretamente, **Diputación de Sevilla** será centro de atención en las fases de **detección de necesidades**, **evaluación del diseño**, **satisfacción** y **conocimiento**; **FORPAS** presenta su mayor nivel de desarrollo en **evaluación de la satisfacción**, de **conocimientos** y de **transferencia**; por último, el **IAD** destaca por su **evaluación de necesidades**, de **satisfacción**, **conocimiento** y **transferencia**.

CAPÍTULO 5. EVALUACIÓN ANTES DE LAS ACCIONES FORMATIVAS: DETECCIÓN DE NECESIDADES

1. Introducción	211
2. Primera fase: pre-evaluación	215
2.1. Objetivos	215
2.2. Método	215
2.2.1. Muestra	215
2.2.2. Instrumentos	216
2.2.3. Procedimiento	217
2.3. Resultados	217
2.4. Discusión.....	218
3. Segunda fase: evaluación principal	219
3.1. Objetivos	219
3.2. Método	219
3.2.1. Muestra	219
3.2.2. Instrumentos	220
3.2.3. Procedimiento	223
3.3. Resultados	223
3.4. Discusión.....	224
4. Tercera fase: post-evaluación	227
4.1. Objetivos	227
4.2. Método	227
4.2.1. Muestra	227
4.2.2. Instrumentos	228
4.2.3. Procedimiento	228
4.3. Resultados	229
4.4. Discusión.....	231
5. Diseño del plan de formación.....	233
5.1. Objetivos	233
5.2. Método	233

5.2.1. Muestra.....	233
5.2.2. Instrumentos	234
5.2.3. Procedimiento	234
5.3. Resultados	235
5.4. Discusión.....	235
6. Estudio comparativo de técnicas de escalamiento	236
6.1. Objetivos	236
6.2. Método	236
6.2.1. Muestra.....	236
6.2.2. Instrumentos	237
6.2.3. Procedimiento	237
6.3. Resultados	250
6.4. Discusión.....	253
7. Conclusiones.....	254

CAPÍTULO 5. EVALUACIÓN ANTES DE LAS ACCIONES FORMATIVAS: DETECCIÓN DE NECESIDADES.

1. INTRODUCCIÓN.

A continuación se pasa a describir, tras una breve definición de **detección de necesidades** y siguiendo un esquema de evaluación formativa, el modelo teórico que se utilizará en esta primera fase de evaluación que aportará, entre otros aspectos, la estructura del capítulo dividido fundamentalmente en tres fases: pre-evaluación, evaluación principal y post-evaluación. En cada una de ellas, se describirá el proceso seguido en dos de las tres instituciones previamente descritas, concretamente Diputación de Sevilla y el IAD (Instituto Andaluz del Deporte), prestando especial atención al nivel de operacionalización y a los instrumentos utilizados. Las razones por las que en esta fase no se comentará el proceso desarrollado en FORPAS es que por casuística organizativa se realizó únicamente la pre-evaluación.

Posteriormente, se pasará a realizar un estudio con el objetivo de determinar el proceso de escalamiento más parsimonioso. Por último, se propondrán mejoras para esta fase de evaluación, centradas en un aumento de operativización y de transparencia.

La **detección de necesidades** es, después de la evaluación de resultados, la fase más comúnmente utilizada en evaluación de programas. Puede ser definida como un procedimiento sistemático de identificación de necesidades, priorización y toma de decisiones de cara a la planificación de un programa para la superación de las mismas (Witkin y Altschuld, 1995). En general, una necesidad es considerada como una discrepancia entre el estado actual de un grupo de personas y su estado deseado (Kaufman, 1988). La detección de necesidades consiste en el análisis de estas discrepancias y una posterior priorización de las acciones futuras, de tal manera que se solucionen antes las necesidades más preocupantes, básicas y urgentes (véase por ejemplo de Diego, Fraile y Boada, 2004).

Esta fase previa a la intervención es sumamente importante pues sirve, por un lado, de justificación del programa que va a llevarse a cabo (no tendría sentido intervenir si no hubiera nada que cambiar); por otro lado, es útil para disponer de datos con los que poder comparar, con posterioridad, la información obtenida tras la intervención como posible evidencia de que el programa provocó cambios.

En el proceso de detección de necesidades, pueden distinguirse tres **fases** secuenciales (Chacón, Lara y Pérez, 2002c, adaptado de Witkin y Altschuld, 1995):

1. **Pre-evaluación:** en esta fase de carácter exploratorio se pretende ordenar y analizar la información útil que la organización ha ido recopilando a lo largo del tiempo, identificar las áreas en las que más incide la necesidad, fijar límites y establecer fuentes de recogida de más información. En definitiva, se determinan los criterios para la toma de decisiones final. El hecho de que desde el primer momento se justifiquen los criterios que posteriormente guiarán la toma de decisiones, aporta un alto grado de transparencia y objetividad al proceso y lo dota de validez.

2. **Evaluación principal:** consiste en la recogida de datos a través de las tres fuentes principales de obtención de información: datos de archivo, procesos de comunicación y procesos analíticos.
3. **Post-evaluación:** constituye el puente de enlace que conecta los datos recogidos y los planes de acción que se determinan finalmente. En esta tercera fase se lleva a cabo el establecimiento de prioridades y criterios para adoptar soluciones, la ponderación de las distintas vías alternativas posibles teniendo en cuenta los aspectos positivos y negativos de cada una, la comparación entre los recursos necesarios y los disponibles (en la actualidad y en el futuro a corto o medio plazo) y la formulación de planes de acción para tratar de solventar las necesidades detectadas y consideradas las de más urgente solución.

Tras un estudio bibliográfico (Chacón, Sanduvete y Holgado, en elaboración), se llegó a la conclusión de que los instrumentos metodológicos más utilizados en la detección de necesidades eran los cuestionarios (usados en el 43.07% de los estudios encontrados), seguido de la entrevista (19.61%), las técnicas grupales (9.1%), la revisión de literatura (8.06 % de los casos) y la búsqueda en archivos (4.57%).

Respecto al análisis de los datos que más a menudo se realizan, en la mayoría de las ocasiones se presentan a un nivel descriptivo (60.61%), seguido de análisis inferenciales en un 24.47% de las ocasiones (regresión, Chi cuadrado, ANOVA, McNemar, etc.), estudios métricos (7.2%), análisis multivariante (6.52%) y meta-análisis (1.2%).

Generalmente, la detección de necesidades ha estado desarrollada principalmente a tres niveles de análisis: a nivel organizacional, del trabajo o la tarea y, finalmente, individual o personal. Estos niveles son los representados en el modelo de Kirkpatrick (1999) referido a resultados y ya comentado en el tema 1, pero a la inversa ya que la evaluación de la satisfacción parte del nivel individual, la evaluación del conocimiento se centra en la tarea, y las evaluaciones de transferencia e impacto se relacionan con el nivel organizacional (Horton, cit. en Phillips y Phillips, 2001). Como podrá verse en la evolución a lo largo del tiempo, esta visión fue la inicialmente utilizada en las dos instituciones mencionadas, ya que principalmente eran los representantes sindicales, los responsables de formación y los jefes de sección quienes decidían qué acciones formativas eran necesarias para mejorar el desempeño laboral de los empleados a su cargo; sin embargo, más adelante el proceso se fue “democratizando”, dando la palabra a todos los potenciales usuarios, desde los primeros momentos hasta las fases de mayor grado de toma de decisiones, aunque se ha de admitir que la última palabra siempre queda en manos de un grupo de personas liderado por el mayor responsable. Se intenta que, a nivel técnico, el proceso sea transparente y participativo, para aportar una información fidedigna, que sea utilizada por los mandatarios a la hora de tomar las decisiones finales.

Al dar la palabra a todos los potenciales usuarios, se ha dado fuerza a una idea que no siempre ha sido bien recibida por todos los responsables y por la organización en general, y es que la formación no sólo ha de centrarse en la mejora del desempeño en el puesto de trabajo a nivel instrumental concreto en relación con las funciones específicas del puesto de trabajo, sino que también ha de buscar la formación personal a nivel holístico, con lo que es probable que aumente la motivación de las personas por la

realización de los cursos, esto a su vez redunde en un aumento de la satisfacción personal e indirectamente esto provoque una mejora en el puesto de trabajo; así, por ejemplo, tras la conducta agresiva de una persona que atiende a los usuarios a través de una ventanilla, puede esconderse una situación de estrés provocada por el descontento con las condiciones laborales o con la vida personal y, en estas circunstancias, el mínimo cambio a positivo puede conllevar grandes consecuencias en la conducta y el estado de ánimo de la persona.

Los objetivos concretos de este capítulo son los siguientes:

- Describir el proceso de detección de necesidades a lo largo del tiempo tanto en Diputación de Sevilla como en el Instituto Andaluz del Deporte.
- Proponer mejoras en las distintas fases de la detección de necesidades para un futuro desarrollo.
- Como mejora en la fase de post-evaluación, hacer un estudio comparativo entre las distintas técnicas de escalamiento para determinar aquella que, siendo más sencilla, aporta unos resultados más precisos.

A continuación, en la tabla 5.1 se describen las distintas fases de detección de necesidades en Diputación de Sevilla durante 12 años (Chacón, Pérez y Lara, 2000a y b; Chacón, Pérez, Lara y Holgado, 2001; Chacón, Lara y Pérez, 2002a; Chacón, López, Sanduvete y Holgado, 2003; Chacón, Sanduvete y Holgado, 2005a, 2005b; Chacón, Sanduvete, Holgado y Sánchez, 2007); en la tabla 5.2, el desarrollo en el Instituto Andaluz del Deporte durante 2 años (Chacón, Gómez, Sánchez y Sanduvete, 2006; Chacón, Sánchez, Holgado y Sanduvete, 2007; Chacón, Sanduvete, Holgado, Rodríguez y Gómez, 2007).

AÑO	DIFUSIÓN	MUESTRA	INSTRUMENTOS	PROCEDIMIENTO	RESULTADOS
1995-1998	Cargos electos	Cargos electos sin criterio de elección	Métodos de difusión masiva	Recogida de información no sistemática y criterios de priorización desconocidos	Frecuencia de cada acción formativa
1999-2002	Muestreo estratificado aleatorio	Pre: participantes plan anterior Ppal: muestreo de Responsables Pos: comisión, Representantes	Pre: pregunta cuestionario satisfacción Ppal: cuestionario semi-cerrado Pos: reuniones	Pre: frecuencia Ppal: frecuencia Pos: toma de decisiones en función de los resultados	Frecuencia, nivel de dificultad, modalidad, objetivos, contenidos, destinatarios, utilidad, área formativa y proveniencia (desde 2004, además, priorización)
2003-2004-2005	Todos (en 2003, a través de Responsables de Formación)	Pre y Pos como estudio anterior Ppal: todos	Pre y Ppal como estudio anterior Pos: cuestionario priorización	Pre y Ppal como estudio anterior Pos: Método de los intervalos sucesivos de Thurstone (Barbero, 1993)	
2006		Pre y Ppal como estudio anterior		Pre y Ppal como estudio anterior	Frecuencia, nivel, destinatarios,

		Pos: Todos		Pos: método de Dunn-Rankin (Arce, 1994)	utilidad, proveniencia, priorización
--	--	-------------------	--	--	--------------------------------------

Tabla 5.1. Resumen del proceso de detección de necesidades en Diputación de Sevilla.

AÑO	DIFUSIÓN	MUESTRA	INSTRUMENTOS	PROCEDIMIENTO	RESULTADOS
2005-2006	Todos	<p>Pre: participantes plan anterior y cursos planificados pero no realizados</p> <p>Ppal: muestra incidental de implicados con el deporte en Andalucía y usuarios de la web</p> <p>Pos: usuarios deportivos con quienes se contó en la fase principal</p>	<p>Pre: pregunta cuestionario satisfacción y registro de acciones formativas diseñadas en 2005</p> <p>Ppal: cuestionario semi-cerrado</p> <p>Pos: cuestionario priorización</p>	<p>Pre (sólo en 2006): Categorización de las respuestas, frecuencia y revisión del grado de implementación del programa de formación diseñado en 2005</p> <p>Ppal: frecuencia</p> <p>Pos: en 2005, Método de los intervalos sucesivos de Thurstone (Barbero, 1993). En 2006, método de Dunn-Rankin (Arce, 1994)</p>	Frecuencia, nivel de dificultad, modalidad, objetivos, contenidos, destinatarios, utilidad, área formativa, proponente, recursos, procedencia geográfica, solicitantes y priorización

Tabla 5.2. Resumen del proceso de detección de necesidades en el Instituto Andaluz del Deporte.

Como puede verse, en Diputación de Sevilla (tabla 5.1) se diferenciaron cinco estudios, ya que se consideró como uno sólo aquellos años en los que el proceso apenas varió; de este modo, el primer estudio englobó desde 1995 a 1998; el segundo, desde 1999 a 2002; el tercero del 2003 al 2005 en conjunto; y el quinto y último, 2006. Se describió la difusión del proceso de evaluación (cómo se informó sobre la detección de necesidades y sobre el plan de formación a los potenciales participantes); el método (muestra, instrumentos y procedimiento); y los resultados obtenidos. A partir de 1999, se siguió el modelo adaptado de Witkin y Altschuld (Chacón, Lara y Pérez, 2002c), por lo que aparecen diferenciadas las tres fases en las que se dividió la detección de necesidades: pre-evaluación, evaluación principal y post-evaluación.

Respecto al IAD (tabla 5.2), se presentan como un único estudio la descripción de la detección de necesidades de los años 2005 y 2006, siguiendo la misma estructura utilizada para la descripción del proceso en Diputación de Sevilla. A pesar de haberse fusionado los dos años en un solo estudio, se dieron algunas mejoras en el 2006 con respecto al 2005, que se comentarán más adelante.

Pero, como ya se ha comentado previamente, la última palabra acerca de las acciones formativas que componen los planes de formación del año siguiente no la tienen los técnicos, sino las Comisiones de Formación de cada una de las organizaciones. Esta es la razón por la que, en este capítulo, aparece tras la fase de post-evaluación un apartado referido al diseño del plan, donde se explican brevemente las posibles discrepancias entre los resultados obtenidos con la detección de necesidades y

las acciones formativas que finalmente conforman los planes, sus causas y posibles mejoras para que no se den estas discrepancias en futuras ocasiones.

A continuación, pasarán a describirse más detalladamente cada uno de los apartados en las distintas fases, se tratará de presentar la evolución en ambos organismos, y se propondrán mejoras que tratarán de implementarse en futuras detecciones de necesidades.

2. PRIMERA FASE: PRE-EVALUACIÓN.

2.1. Objetivos.

En esta primera fase, se pretendieron cubrir dos objetivos:

1. Realizar un acercamiento de forma exploratoria y sistemática a las **acciones formativas** detectadas en el transcurso del **año anterior**, y aquellas diseñadas en los planes de formación actuales pero no llevadas a cabo.
2. Elaborar un **cuestionario** de detección de necesidades que sirviera como instrumento de recogida de información para la fase siguiente.

Con el **primer objetivo**, se trató de detectar en cierto modo qué acciones formativas eran consideradas de interés por aquellas personas que habían participado en el plan del año anterior; y qué acciones formativas, habiendo sido propuestas e incluidas en el plan de formación, no se habían llevado a cabo pues, tras estudiar las razones de esta exclusión, se sopesaría la posibilidad de incluirla en el plan posterior.

Con el **segundo objetivo** el proceso lograba coherencia, ya que las necesidades propuestas por los participantes del plan anterior pasarían, como acción formativa a cubrir dicha necesidad, a formar parte del cuestionario con el que se recogerían los datos en la evaluación principal.

2.2. Método.

2.2.1. Muestra.

A lo largo del tiempo, la información acerca de la realización de una detección de necesidades fue llegando a más implicados ya que, entre los años 1995 y 1998 en Diputación sólo eran informados los cargos electos; entre 1999 y 2002 se realizó un muestreo estratificado aleatorio (con lo que no todos los pueblos adheridos al plan fueron informados; y del 2003 en adelante se dirigió el proceso de difusión a todos los potenciales usuarios, concretamente en 2003 a través de los responsables de formación y, en los años posteriores, directamente a nivel individual. Del mismo modo se hizo en el IAD desde el inicio de su evaluación en el 2005.

Y no sólo se trataba de que llegara la información, sino que también se abrió la posibilidad de participación a todos los implicados, cuando inicialmente estaba restringido a unos pocos. Así, concretamente entre los años 1995 y 1998 en Diputación donde aún no se distinguía el proceso en las tres fases de evaluación propuestas, únicamente participaron determinados cargos electos que fueron elegidos sin ningún criterio explícito y, más adelante, los Responsables de Formación, figura creada para representar a los trabajadores a su cargo, con la finalidad de que cada uno en su pueblo o su área fuera informando del proceso de detección de necesidades.

Concretamente, en Diputación de Sevilla hasta 1997, el número de alcaldes de los Municipios adheridos al plan de formación y de otros cargos electos sin criterio de elección fue en aumento, participando 33 en 1995, 35 en 1996 y 51 en 1997; y en 1998, participaron 74 Responsables de Formación.

Entre los años 1999 y 2002, se realizó un muestreo estratificado del total de Municipios: se distinguieron 4 estratos en función del número de empleados (de 1 a 5 empleados; de 26 a 50; de 51 a 75, y de 75 en adelante). Se eligieron al azar 32 Municipios (8 de cada grupo).

Sin embargo, a partir del año 2003 y en ambas instituciones, la muestra en la fase de pre-evaluación estuvo compuesta por aquellas personas que, habiendo participado en el plan de formación del año anterior, respondió a la pregunta de qué acciones formativas consideraría que necesitaba para mejorar el desempeño en su puesto de trabajo, por lo que no se hizo muestreo; todos las participantes en el plan pudieron responder, en el 2003 a través del Responsable de Formación y, a partir de 2004, directamente sin utilizar dicha figura como intermediario.

Además, en aquellas ocasiones en las que se planificó la realización de una acción formativa y posteriormente no se llevó a cabo, se incluyó en esta fase en la detección de necesidades del año posterior para decidir si era adecuado que se ejecutara en el siguiente plan.

2.2.2. Instrumentos.

Entre los años 1995 y 1998, se utilizaron métodos de difusión masiva como los dípticos, los pósters y los folletos. Además, junto con la carta enviada a los Municipios para comunicarles la posibilidad de adherirse al plan agrupado, se citaba a los Alcaldes a reuniones en las que se trató de decidir acerca de qué acciones formativas ejecutar.

A partir del año 1999 y en ambas organizaciones, se introdujo en el cuestionario de satisfacción que se recogía tras la finalización de cada una de las acciones formativas una pregunta abierta en la que se solicitaba a los participantes la enumeración de “Otras acciones formativas que consideraran necesarias”, lo que sería una recogida de necesidades expresadas. Y a partir del año 2006, se solicitó más información acerca de las acciones formativas propuestas: el grado de necesidad, la modalidad, la utilidad, el puesto de trabajo del proponente y el lugar de trabajo del participante. Esta información fue la misma que se recogió posteriormente en la evaluación principal. En el apartado referido a instrumentos de la siguiente fase se amplía información acerca de cada uno de estos componentes (ver tabla 5.4).

También se llevó a cabo un análisis documental de las acciones formativas que se habían realizado en años anteriores con el que se estudió, además de sus resultados en satisfacción una vez finalizadas, el número de personas que, aun habiéndolas solicitado, no pudieron realizarlas por falta de plazas (lo cual recogería las necesidades comparadas) y aquellas acciones que, aun habiendo sido diseñadas y planificadas, finalmente no se llevaron a la práctica.

2.2.3. Procedimiento.

Entre los años 1995 y 1998, la recogida de información en Diputación de Sevilla no fue sistemática y, aunque los criterios de priorización no fueron totalmente conocidos, se explicitó que el punto de partida fueron las peticiones personales recibidas y un análisis documental con el que se determinó qué acciones formativas tuvieron una alta demanda en planes de formación anteriores y se concretó que, de entre todas las acciones recogidas por estos procedimientos, se eligió un número concreto en función de las previsiones económicas, a través del estudio de las ofertas del mercado.

En el año 2005, la evaluación en el IAD comenzó en julio. Por falta de tiempo, fue imposible realizar una pre-evaluación aparte de la evaluación principal, por lo que los resultados se pasan a describir en el siguiente apartado.

A partir de 1999 en Diputación de Sevilla y en el año 2006 en el IAD, con las acciones formativas más solicitadas en función de los resultados obtenidos en esta fase, se elaboró un cuestionario semicerrado que se utilizaría para recoger información en la próxima fase: la evaluación principal.

Las acciones formativas demandadas previamente pasaron a formar parte del cuestionario utilizado en la siguiente fase del siguiente modo:

1. Se incluyeron aquellas acciones de las que se recogieron todos los datos que se solicitaban. Se excluyeron aquellas acciones con un bajo grado de especificación.
2. Cuando se incluyeron varias veces la misma acción pero con distinta denominación, se fusionaron en una sola.
3. Cuando el contenido de varias acciones formativas era complementario, se agruparon en solo una.

2.3. Resultados.

Las necesidades recogidas en esta fase fueron del tipo “**experimentadas, sentidas o percibidas**”, puesto que las personas respondieron porque fueron preguntadas a través de un cuestionario.

Las acciones formativas solicitadas se encuadraron en distintas áreas de conocimiento, estructura que se mantuvo en la elaboración del cuestionario para la evaluación principal. En 1995 en Diputación de Sevilla y en el año 2005 en el IAD, estas áreas formativas no venían dadas, por lo que se dedujeron a posteriori, tras la recogida y el análisis de las solicitudes. Se incluyeron en el cuestionario, para cada una

de las áreas, el número de acciones que fueron solicitadas con mayor frecuencia en proporción con el total de acciones formativas recogidas en cada una de ellas.

A modo de resumen, en la tabla 5.3 se recoge el número de acciones formativas que fueron solicitadas en esta fase de pre-evaluación a través del cuestionario de satisfacción, que fue el método de recogida que aportó mayor grado de información específica.

ORGANISMO	AÑO	NÚMERO DE ACCIONES FORMATIVAS
Diputación	1995-98	No hubo pre-evaluación propiamente dicha
	1999	27
	2000	17
	2001-2002	82
	2003	57
	2004	117
	2005	218
	2006	285
IAD	2005	No se dio pre-evaluación
	2006	149

Tabla 5.3. Número de acciones formativas recogidas en la fase de pre-evaluación a través de la pregunta abierta del cuestionario de satisfacción.

En líneas generales, el número de acciones formativas fue aumentando a lo largo del tiempo. Este hecho puede explicarse porque el uso de los cuestionarios de satisfacción (donde se presentaba la pregunta acerca de otras acciones formativas a realizar) se fue consolidando, e indudablemente porque cada año los planes de formación fueron recogiendo un mayor número de participantes.

Cuando el número de acciones formativas no fue muy elevado, aparecieron todas en el cuestionario de evaluación principal; sin embargo, cuando pasaron de 60, se mostraron únicamente las que en cada área presentaron mayor frecuencia.

2.4. Discusión.

Algunas mejoras que se llevaron a cabo a lo largo del tiempo en esta primera fase de pre-evaluación fueron las siguientes:

Respecto a los **participantes**, el proceso se fue haciendo cada vez más democrático. En los primeros años de Diputación, se realizó un muestreo de tal manera que no todos los Municipios adheridos pudieron participar; y las personas que pudieron expresarse fueron representantes, ya fueran Alcaldes u otros cargos electos, ya fueran los responsables de formación. En siguientes años, sin embargo, pudieron responder en esta fase todos los participantes en las acciones formativas a través de una pregunta realizada en el cuestionario de satisfacción. De este modo, se empezó potenciando la representatividad de los distintos estratos (cargos electos, responsables de formación y demás empleados) y, finalmente, se fomentó la participación de todos los implicados.

En relación a los **instrumentos** utilizados, se pasó de métodos de difusión masiva con los que era difícil tener certeza de que realmente habían hecho llegar la información a un método de recogida poco costoso y natural, ya que el cuestionario de satisfacción cuya utilización estaba ampliamente aceptada entre los participantes de las

distintas acciones formativas permitió, a partir de un único ítem incluido, la recogida de información acerca de las necesidades de formación detectadas.

Respecto al **análisis de datos**, encuadrando cada acción formativa solicitada en un área de conocimiento se logró organizar la información obtenida de tal manera que fue útil tanto para la elaboración del cuestionario que posteriormente se usara en la evaluación principal, como para el estudio de los resultados en dicha fase.

Los resultados obtenidos en esta primera fase sirvieron de base para la implementación de la siguiente. La **utilidad** de este modo de proceder radica en que así se dotó de coherencia al proceso y se dejó claro que los pasos seguidos no se estaban haciendo de forma arbitraria. Concretamente, cuando se recogió un amplio número de acciones formativas, en esta fase se trató de dejar claros los criterios por los que algunas acciones formativas solicitadas se incluyeron en el cuestionario de la evaluación principal y otras quedaron fuera.

El proceso se adecuó a un **modelo teórico**, concretamente el propuesto por Witkin y Altschuld (1995) y adaptado por Chacón, Lara y Pérez (2002c). De este modo, en los primeros años en Diputación y en el 2005 en el IAD, no se realizó una pre-evaluación organizada y exhaustiva que se distinguiera claramente de la evaluación principal, a pesar de que la división entre estas fases es interesante porque, de este modo, se recoge información de años previos que sirve de base para los siguientes. No tiene mucho sentido solicitar información nueva sin antes haber analizado la ya recogida.

3. SEGUNDA FASE: EVALUACIÓN PRINCIPAL.

3.1. Objetivos.

Esta segunda fase en el proceso de detección de necesidades persiguió principalmente dos objetivos:

1. **Detectar necesidades** formativas del mayor número de implicados/as posible en el proceso formativo.
2. Elaborar un nuevo **cuestionario** de priorización de necesidades para la fase de post-evaluación.

3.2. Método.

3.2.1. Muestra.

En Diputación, durante los años 1995 a 1998 no se realizó un proceso de evaluación por fases por lo que, para obtener más información en este apartado y los que le siguen, se puede recurrir a la lectura de la fase anterior, donde se explicitaron estos puntos.

Durante los años 1999 a 2002, participaron en esta fase de evaluación principal los responsables de formación de aquellos Municipios que fueron elegidos a través de un muestreo estratificado y por cuotas, suponiéndose que estos responsables habían recogido previamente las opiniones de los empleados a los que representaban.

A partir de 2003 en ambas organizaciones, en cambio, se optó por dar la oportunidad a todos los implicados a que participaran, a través de un representante en este primer año 2003 y de manera directa en los demás años.

3.2.2. Instrumentos.

Desde 1999 y en ambas organizaciones, se comenzó a utilizar para la recogida de información cuestionarios semicerrados que, como se comentó en el apartado anterior, mostraban como propuesta las acciones formativas más solicitadas en la pre-evaluación (además de apartados en blanco por si algún participante quería añadir alguna que no estuviera).

Tal como se presentó en la introducción, los cuestionarios son los instrumentos de evaluación más utilizados con diferencia en detección de necesidades; y hay que tener en cuenta que optar por un instrumento y descartar los otros tiene tanto ventajas como inconvenientes que sería interesante valorar (Chacón, Lara y Pérez, 2002c). Algunos de sus principales **puntos fuertes** son que aportan información homogénea, permiten en poco tiempo encuestar a un gran número de personas, permiten cuantificar e interpretar los datos mediante análisis estadísticos y proporcionan datos que se pueden resumir y narrar fácilmente. Entre sus **inconvenientes** se encuentran que impiden profundizar en las respuestas de los encuestados, es posible que se logre una baja tasa de respuesta, su elaboración es costosa y pueden exigir unos determinados requisitos en las personas que lo cumplimentan.

Ningún instrumento será perfecto en este sentido pero, para tratar de eliminar los puntos débiles, se intentó construir el cuestionario con la mayor rigurosidad que fue posible.

Según la literatura existente sobre evaluación de necesidades formativas (Salanova y Grau, 1999), la validez de un cuestionario para la detección de necesidades formativas viene marcada por el hecho de que el instrumento permita estructurar, en unas áreas formativas concretas, una serie de acciones con unos destinatarios específicos y con una serie de mejoras relacionadas con el desempeño en sus puestos de trabajo. Todos estos componentes van a formar parte del cuestionario. Veamos cada uno de ellos de forma más detallada.

1. Áreas formativas.

El primer paso que se llevó a cabo en la elaboración del cuestionario fue el establecimiento de las distintas áreas formativas presentes en el plan de formación de Diputación de Sevilla y Municipios, Consorcios y Mancomunidades adheridos.

En Diputación de Sevilla, la conformación de las mismas se realizó en base al III Acuerdo para la Formación Continua en la Administración Pública (A.F.C.A.P.)

(Instituto Nacional de Administración Pública, 2001). Las áreas resultantes fueron “A. Dirección y Gerencia Pública”, “B. Económico Presupuestaria”, “C. Específico de Determinados Colectivos”, “D. Idiomas/Lengua”, “E. Información y Atención al Ciudadano”, “F. Jurídico – Procedimental”, “G. Nuevas tecnologías. Información y Comunicación”, “H. Prevención de Riesgos Laborales”, “I. Recursos Humanos”, “J. Unión Europea” y “K. Urbanismo y Medio Ambiente”.

En el IAD, las distintas áreas formativas fueron conformadas en el 2005 tomando como base las que previamente aparecían. A posteriori, tras estudiar las acciones formativas que habían sido solicitadas, se incluyó algunas áreas para que todas las solicitudes pudieran ser encuadradas. En definitiva, las áreas formativas fueron “A. Gestión y Administración Deportiva”, “B. Medicina Deportiva”, “C. Infraestructura”, “D. Turismo y Deporte”, “E. Actividad Física Deportiva” y “F. Formación de Formadores”. En el año 2006 se incluyó un área más: “G. Títulos de Postgrado”.

El siguiente paso fue identificar las características de aquellas acciones formativas que entraran a formar parte de cada una de las áreas mencionadas anteriormente. Para mayor claridad en la exposición, en la tabla 5.4 se presenta un ejemplo de las cuadrículas que conformaban el cuestionario.

Nombre de la acción formativa:	Nivel: Básico/ medio/ avanzado	Modalidad: Presencial/ distancia
Destinatarios:		
Utilidad/Mejora en tu Puesto de Trabajo (explicitar número de afectados si es posible):		
Proponente, lugar y puesto de trabajo		

Tabla 5.4. Ejemplo de una viñeta del cuestionario de detección de necesidades.

2. Nombre de la acción formativa.

Con objeto de conseguir unos resultados claros y estructurados, fue necesario el encuadre de las acciones formativas en cada área de conocimiento; con ello, se consiguió dar coherencia al informe de resultados final, lo que facilitó en gran medida relacionar el proceso de evaluación de necesidades formativas con la programación del plan de formación del año posterior.

El procedimiento utilizado para decidir qué acciones formativas incluir en el cuestionario se basó en el estudio de los resultados obtenidos en la pre-evaluación de tal manera que, como ya se comentó en el apartado referido a esta primera fase de evaluación, se tomaron las acciones formativas con mayor frecuencia de manera proporcional a las solicitadas en cada área. Y además, en cada área de formación se incluyeron un par de cuadros en blanco como el presentado en la tabla 5.4, por si algún participante quisiese incluir alguna acción formativa no presentada en la pre-evaluación.

4. Nivel.

La finalidad de la recogida de este componente fue facilitar el diseño por parte de los formadores de cada una de las acciones formativas, al haberse obtenido información sobre el punto de partida desde el que los futuros participantes necesitaban empezar (básico, medio o avanzado).

5. Modalidad.

Con este apartado, se recogió información acerca de la preferencia de los solicitantes respecto al formato de impartición de la acción formativa solicitada: de manera presencial o con la modalidad a distancia; así, en un primer momento, la modalidad presencial era la más común y consistía en asistir a clases, estudiar un material en papel y realizar actividades que posteriormente eran entregadas; y más adelante, con el avance en las nuevas tecnologías, se fue imponiendo el formato de cursos on-line, a través de Internet, ganándose así en agilidad e interactividad.

6. Destinatarios.

Con este apartado, se trató de establecer a qué destinatarios concretos iban dirigidas las acciones formativas solicitadas, con la intención de tener información, a la hora de diseñar las acciones y sus contenidos, acerca de las necesidades concretas de los trabajadores afectados.

7. Utilidad para la mejora del desempeño en el puesto de trabajo tras la acción formativa.

En este apartado se recogieron las necesidades a cubrir y las mejoras que, según la opinión del solicitante, conllevarían las acciones formativas implementadas. Como dato concreto, se preguntó por el número aproximado de afectados por la necesidad como posible criterio de priorización, ya que puede considerarse como más urgente aquella necesidad que afecte a más personas; sin embargo, hay que decir que este criterio no sería el único a tener en cuenta, ya que la gravedad de las consecuencias, por ejemplo, también sería algo a considerar a la hora de priorizar unas necesidades sobre otras.

8. Proponente y lugar de trabajo.

Uno de los objetivos que se plantearon ambas organizaciones fue descentralizar las acciones formativas acercándolas a los lugares en los que la necesidad fue manifestada en mayor medida, para facilitar la asistencia a aquellas personas que realmente necesitaban dicha formación. Esta es la razón por la que fue tan importante la recogida de este componente; fue esencial para tratar de crear una zonificación de las acciones formativas en función de la demanda.

Además, se dio al solicitante la oportunidad de que diera sus datos de identificación por si fuera de interés contactar con él para aclarar algún aspecto o solicitar ampliación de información de cara a la planificación de dicha acción formativa.

Esta es la última versión de la información obtenida de cada acción formativa con el cuestionario de necesidades. En versiones anteriores, se preguntaba también por otros dos aspectos:

9. Objetivos y contenidos.

Estos componentes se habían incluido en el cuestionario para que los solicitantes explicaran los objetivos y contenidos principales que debían cubrir las acciones

formativas propuestas, para que estuvieran lo más adaptadas posibles a las necesidades reales surgidas en el puesto de trabajo.

Sin embargo, estos dos elementos se eliminaron del cuestionario en su última versión porque, aunque un empleado tenga la capacidad de discernir una necesidad que le impide realizar correctamente su labor, a priori no tiene por qué tener suficientes conocimientos para valorar qué debe aprender concretamente para cubrir dicha necesidad. Más bien, esto sería tarea de los formadores.

3.2.3. Procedimiento.

En los primeros años en Diputación, este instrumento elaborado básicamente se entregaba en papel por correo ordinario, y en papel era recogido una vez cumplimentado a través de fax o en mano en la Sección de Formación.

Con el avance de las nuevas tecnologías, sin embargo, a partir del año 2005, en esta organización se primó el acceso al cuestionario a través de Internet y la recogida de la información automáticamente en bases de datos. Sólo en algunas ocasiones, cuando esta vía no era factible por alguna razón, se utilizaron los medios tradicionales y el correo electrónico.

En el IAD, desde el primer año (2005) se primó el uso del correo electrónico para envío y recogida de los cuestionarios. Así, en primer lugar se recopiló el mayor número posible de direcciones de correo electrónico de los implicados en deportes en la comunidad autónoma andaluza; en segundo lugar, se envió por correo electrónico el cuestionario a estas personas; y por último, se recogieron las necesidades formativas demandadas también por esta vía.

En ambas organizaciones, una vez recogida la información, se pasó al análisis de datos para obtener los resultados que a continuación se pasan a comentar.

3.3. Resultados.

En esta fase del proceso de detección de necesidades, al igual que en la anterior, se recogieron necesidades “**experimentadas, sentidas o percibidas**”, puesto que las personas respondieron al ser preguntadas a través de un cuestionario.

En la tabla 5.5 que a continuación se presenta se muestra, para cada año y en ambas organizaciones, el número de acciones formativas que fueron solicitadas en la evaluación principal y el número de las que finalmente conformaron el plan de formación.

Concretamente, la primera columna muestra el organismo del que se habla (Diputación o IAD); la segunda, el año en el que se realizó el estudio de detección de necesidades; en la tercera columna, se muestra el número de acciones formativas distintas solicitadas; y en la última columna, con la intención de posibilitar una comparación con el dato anterior, se presenta el número de acciones formativas que finalmente conformaron el plan de formación del año siguiente.

ORGANISMO	AÑO	ACCIONES SOLICITADAS	ACCIONES QUE CONFORMARON EL PLAN
Diputación	1995	Dato no disponible	27
	1996	96	67
	1997	184	112
	1998	109	76
	1999	82	78
	2000	82	96
	2001	82	55
	2002	117	63
	2003	72	61
	2004	117	89
	2005	191	119
	2006	286	97
	IAD	2005	217
2006		303	59

Tabla 5.5. Número de acciones formativas solicitadas y las que finalmente conformaron el plan a lo largo del tiempo en las distintas organizaciones.

De manera general puede decirse que, a lo largo de los años, el número de acciones formativas solicitadas fue en aumento, al igual que el número de acciones que conformaron el plan de formación.

Hay que tener en cuenta, sin embargo, que el número de acciones formativas que pasaban a ser parte del plan no dependió de la participación en la detección de necesidades, sino de otros factores como la subvención recibida.

Se recogió toda la información de cada una de las acciones formativas y se contabilizó la frecuencia.

Esta información se utilizó finalmente para crear el cuestionario para la fase de post-evaluación; concretamente se presentó un listado, dividido por áreas de formación, de todas las acciones que fueron solicitadas.

3.4. Discusión.

En esta fase de evaluación principal, se realizaron algunas **mejoras** a lo largo del tiempo:

Respecto al **modelo teórico** utilizado (Chacón, Lara y Pérez, 2002c, adaptado de Witkin y Altschuld, 1995) cabe destacar que, a lo largo del tiempo, se fue logrando un mayor ajuste. Así en los primeros años tanto en Diputación de Sevilla como en el IAD, no se realizó fase de pre-evaluación y, posteriormente sí se hizo.

Respecto a los **destinatarios**, al igual que en la fase de pre-evaluación, el proceso fue volviéndose más participativo a lo largo del tiempo ya que se pasó de que sólo pudieran responder altos cargos a que pudieran dar respuesta todos pero a través de los responsables de formación y, finalmente, a que pudieran responder de manera individual quienes quisieran.

Además, anteriormente no se consideraba de mismo valor una respuesta dada por un alto cargo y una dada por un administrativo, por ejemplo. Desde el año 2005, todas las opiniones fueron igualmente consideradas.

Respecto al **instrumento** utilizado, a lo largo del tiempo se fue concretando la información recogida, consiguiendo un cuestionario más sencillo. Un instrumento que requiera demasiado tiempo para su cumplimentación puede provocar un bajo número de participantes y se llegó a la conclusión de que los destinatarios no necesariamente tenían que conocer las respuestas a todos los apartados. Por esta razón, se excluyó la recogida de información acerca de los objetivos y los contenidos de las acciones formativas. Además, no siempre se procesaba ni utilizaba la información que se recogía con estos apartados, hecho que era un sinsentido.

Respecto al **procedimiento**, cabe destacar que la utilización de manera común de los instrumentos que Internet ofrece permitió una cumplimentación, envío y recogida de información más cómodos y ágiles y una codificación más clara y fiable (al recogerse los datos automáticamente en una base, la probabilidad de cometer errores disminuye en comparación con la codificación manual). Sin embargo, hay que tener en cuenta que los medios tradicionales no se eliminaron completamente puesto que no todos los empleados utilizan con soltura los ordenadores e Internet y, además, no todos los puestos tienen acceso fácil a este medio (véase lo paradójico, por ejemplo, de una persona que quiera solicitar el curso “Internet y correo electrónico” en su nivel inicial y que únicamente puede hacerlo por esta vía, cuyo funcionamiento desconoce).

Los **momentos de recogida** de información fueron ampliados. En los primeros años, la información se recogía en momentos puntuales; sin embargo, desde 1999 hasta hoy, la recogida de información se hace de manera continua, de tal manera que cualquier persona puede expresar su necesidad de formación en el justo momento en que la detecta. Así, por ejemplo, en la página web de Formación Continua de Diputación (<http://www.dipusevilla.es/formacioncontinua>) hay siempre disponible una ficha de necesidades que puede cumplimentarse y enviarse en cualquier momento.

Respecto a la **estructuración de los resultados**, para facilitar la toma de decisiones final y dotar de coherencia al estudio de detección de necesidades en relación a los planes de formación resultantes, las acciones formativas se mostraron en todo momento diferenciadas en función del área de formación al que se adscribían.

Respecto a la **utilidad** del proceso y de los resultados obtenidos con éste, se logró una recogida de datos exhaustiva en la que cualquier implicado tuvo la oportunidad de participar directamente, se detectó cada vez un mayor número de acciones formativas y se trató de que el proceso fuera abierto y transparente, con plazos de recogida de información amplios, debidamente publicitados y rigurosamente aplicados para que sirvieran como fuente científica a la hora de tomar las decisiones finales.

Haciendo una valoración global se puede decir, tras ver las mejoras alcanzadas, que se fue llegando a un estudio de necesidades cada vez más útil y participativo. Sin embargo, aún quedarían **mejoras por hacer** y algunas de las que están planificadas son las siguientes:

- Respecto al tipo de **instrumento** utilizado, en los últimos años el cuestionario ha adquirido especial relevancia en la detección de necesidades. Sin embargo, se está pensando en utilizar como complemento otros instrumentos de recogida de información cualitativa como pueden ser los grupos de discusión o entrevistas. Se organizarían por zonas para facilitar la participación de todos los Municipios, Consorcios y Mancomunidades en el caso de Diputación y todas las provincias de Andalucía en el caso del IAD. Se intentará de esta forma lograr los siguientes objetivos:
 - Hacer más humano el proceso de evaluación, que se vuelve frío y distante cuando la cumplimentación únicamente se hace con métodos a distancia, donde los usuarios no tienen contacto directo con los evaluadores ni con el personal de la sección de Formación Continua en ningún momento.
 - Provocar en los participantes la sensación de que son escuchados de manera inmediata y de que tienen la posibilidad de participar de manera directa en la creación de los planes de formación.
 - Fomentar la inclusión de acciones formativas más novedosas, animando a los participantes en las entrevistas y grupos de discusión a que innoven y no se limiten a solicitar nuevas ediciones de acciones formativas que ya se realizaron en planes anteriores.
 - Perfilar no sólo las acciones formativas que se quieren incluir, sino también los objetivos y contenidos que han de tener en caso de que los solicitantes conozcan esta información, e incluso los posibles formadores que pudieran encargarse de su diseño e implementación.
 - Utilizar más de un tipo de instrumentos para solventar así los puntos débiles que presenta su utilización por separado y para tener la posibilidad de triangular los datos.
- Respecto al **dato** solicitado a los participantes, hasta ahora se les ha preguntado acerca de cursos y no de necesidades pero, ¿realmente una persona que tiene una carencia a la hora de llevar a cabo su trabajo sabe qué habilidades tiene que adquirir para cubrirla? Quizá no tiene mucho sentido solicitar a los usuarios lo que no tiene por qué saber a priori, ya que puede que no tenga suficientes conocimientos para valorar qué debe aprender o a qué nivel se encuentra respecto a una acción formativa. Por ello, se está barajando la posibilidad de preguntar a través del cuestionario únicamente por carencias a cubrir, y que sean los expertos quienes propongan y diseñen las acciones formativas para eliminar o disminuir al menos dichas carencias.
- En referencia a las **vías de comunicación**, en Diputación la participación vía web ha ido en aumento, principalmente porque para los usuarios se ha convertido en un medio cotidiano ya que, por ejemplo, para solicitar las acciones formativas que quieren realizar o para confirmar la asistencia a dichas acciones, han de usar Internet. En el IAD, al ser algo más novedoso, se tratará de potenciar la participación por esta vía.

- Respecto al **feedback** dado a participantes y/o interesados, se espera que se logre crear una base de datos con la que, a medida que vaya llegando información, se vaya analizando de tal manera que, a través de Internet, se puedan visualizar con actualizaciones instantáneas el conjunto de acciones formativas solicitadas y su frecuencia.

4. TERCERA FASE: POST-EVALUACIÓN.

4.1. Objetivo.

El objetivo que se pretende alcanzar en esta fase es obtener una priorización del total de necesidades formativas delimitadas en las fases anteriores por cada área de conocimiento.

Ocurre que los recursos son limitados, y no siempre se puede atender a todas las necesidades que fueron detectadas. Ante este problema, una solución es hacer un estudio de cuáles de las carencias detectadas son más urgentes de solventar, y actuar en consecuencia, para tener la seguridad de que realmente se incide en lo más problemático y ninguna necesidad importante queda sin atender.

Por ello, en esta fase de post-evaluación se consiguió obtener una ordenación de todas las acciones formativas solicitadas desde la considerada más importante a la menos trascendental.

Esta fase de post-evaluación se realizó por primera vez en Diputación en 1999; el IAD ha contado con esta fase en su detección de necesidades desde el 2005.

4.2. Método.

4.2.1. Muestra.

En esta fase de post-evaluación, la muestra con la que se contó en Diputación desde 1999 (primer año en que se llevó a cabo esta fase) hasta 2005 estuvo compuesta por los miembros de la Comisión de Formación, formada concretamente por Representantes Sindicales (de CCOO, UGT y CIS-CSIF), los técnicos y el Jefe de la Sección de Formación Continua del Área de Recursos Humanos, y los Responsables de Formación de los distintos pueblos y ciudades. En el 2006, sin embargo, se abrió la posibilidad a que respondieran todos los implicados.

En el IAD, la muestra en esta fase estuvo compuesta por aquellas personas que habían participado en las fases anteriores.

Concretamente, el número de participantes en ambas organizaciones a lo largo del tiempo a partir de 2004 (que fue cuando comenzaron a utilizarse los cuestionarios como instrumento exclusivo de recogida de información) se presenta en la tabla 5.6:

ORGANISMO	AÑO	Nº DE PARTICIPANTES (N)
Diputación	2004	9
	2005	21
	2006	374
IAD	2005	5
	2006	31

Tabla 5.6. Número de participantes en la fase de post-evaluación desde que se implementó el cuestionario como instrumento de recogida de información.

Un aspecto a resaltar tras observar los datos es que a medida que pasó el tiempo el número de participantes fue siendo mayor, encontrándose en algunos momentos un aumento considerable; por ejemplo, en el IAD se pasó de 5 participantes en 2005 a 31 en 2007; y aún más llamativo es el cambio de los 21 participantes en 2005 en Diputación a los 374 en 2006.

4.2.2. Instrumentos.

Desde el año 1999 hasta el 2002, en Diputación de Sevilla se recogió la información a través de reuniones con la Comisión de Formación; en el 2003, se utilizó un instrumento híbrido entre el grupo de discusión y el cuestionario; a partir del 2004, al igual que en el IAD, se utilizó únicamente un cuestionario de priorización elaborado a partir de los resultados obtenidos en las fases anteriores. Concretamente, en el 2005 se eligió un número total de acciones formativas a incluir y, a partir de ahí, se tomó de cada área de conocimiento la proporción respecto al total de acciones formativas; en el año 2006, se presentaron para su priorización todas las acciones formativas recogidas, divididas en las distintas áreas de formación. A modo de ejemplo, se presenta en la tabla 5.7 cómo apareció representado en el cuestionario de priorización del año 2005 el área B “Económico-Presupuestaria” de Diputación. En la columna de la izquierda se presentó el nombre de todas las acciones formativas solicitadas, mientras que en la columna de la derecha se dejaron espacios para que los participantes enumeraran las distintas acciones de 1 (mayor necesidad) a 4 (menor necesidad).

ÁREA B: “ECONÓMICO-PRESUPUESTARIA”	
Acciones formativas	Orden (de 1 a 4)
Gestión tributaria y recaudación en la Administración Local	
Presupuesto y contabilidad en la Administración Local	
Contabilidad financiera	
Análisis de balances	

Tabla 5.7. Ejemplo de una viñeta del cuestionario de priorización.

Una diferencia sustancial surgida a lo largo del tiempo fue que, en los años 2004 y 2005, se solicitó a los participantes que priorizaran todas las acciones formativas que aparecían. Sin embargo, en el 2006 (y en ambas organizaciones) se solicitó únicamente la priorización de 10 acciones formativas por cada área formativa.

4.2.3. Procedimiento.

Entre los años 1999 y 2002, la información se recogió a través de reuniones. Los participantes eran convocados con al menos 15 días de antelación, señalándoles el orden del día (priorización de las acciones formativas recogidas) y el lugar, fecha y hora en la que se celebraba.

En el año 2003, se recogió información a través de un cuestionario al que se accedía vía internet, pero se explicó a los participantes el modo de cumplimentación en una reunión realizada en una sala informatizada; tras esta explicación, rellenaron “in situ” el instrumento y lo enviaron.

A partir del año 2004, los cuestionarios fueron enviados en papel y también dispuestos en la página web y se pudieron cumplimentar tras introducir el DNI, para estar seguros de que los participantes realmente trabajaban en Diputación o en el Ayuntamiento de algún Municipio adherido. En el IAD, los cuestionarios fueron enviados por correo electrónico.

Respecto al análisis de datos utilizado para obtener la priorización, desde el año 2003 al año 2005 (en ambas organizaciones), se llevó a cabo el método de los intervalos sucesivos, derivado de la ley de juicios categóricos de Thurstone (Barbero, 1993). En el año 2006, sin embargo, se cambió al método de Dunn-Rankin (Arce, 1994). En el apartado 6 de este capítulo, se describen estos procedimientos y se trata de estudiar las diferencias entre ellos y algunos otros para decidir cuál sería más apropiado para obtener los resultados que se necesitan.

4.3. Resultados.

Dos tipos de necesidades fueron recogidas en esta última fase de detección de necesidades:

- a) Necesidades “**experimentadas, sentidas o percibidas**” cuando priorizaron los trabajadores a partir del cuestionario propuesto.
- b) Necesidades “**normativas**” cuando priorizaron los representantes o responsables de formación, al ser expertos en la materia (aunque los criterios que guiaron sus decisiones no siempre estuvieron explicitados).

De las reuniones realizadas entre 1999 y 2002, se obtenía la priorización de las acciones formativas recogidas previamente y el diseño del plan de formación del año posterior.

A partir del año 2003 y en ambas organizaciones, se distinguió entre la fase de priorización en la que se ordenaban todas las acciones formativas de mayor a menor relevancia (divididas en áreas formativas) y la fase del diseño definitivo del plan de formación que se verá en el siguiente apartado, y en la que las decisiones políticas eran tenidas en consideración (aunque, como se verá se intentó que, al menos, los criterios que se siguieron para tomar las decisiones fueran explicitados).

Respecto a los resultados obtenidos concretamente en esta fase de priorización, en la figura 5.1 que a continuación se presenta puede verse parte de la priorización obtenida en el año 2004 tras llevar a cabo el método de los intervalos sucesivos de Thurstone (Barbero, 1993), en una de las áreas de formación de Diputación de Sevilla, concretamente el área B “Económico-presupuestaria”, y que en el punto 6 se explicará con mayor detalle. Por ahora sólo se comentará que la acción formativa C fue la más

valorada, mientras que la priorizada como menos relevante fue la B (la más cercana a 0).

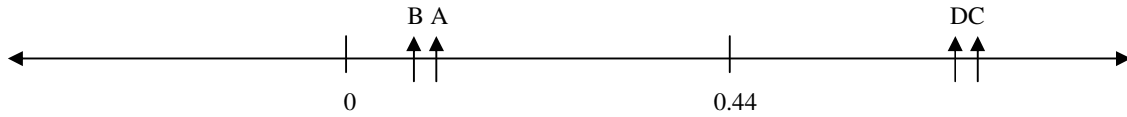


Figura 5.1. Orden de priorización de las acciones formativas del área de conocimiento B “Económico-presupuestaria” en el año 2004 en Diputación de Sevilla.

El método de Dunn-Rankin (Arce, 1994) logra también la priorización de estímulos (en este caso acciones formativas). Por poner un ejemplo, en la figura 5.2 que a continuación se presenta aparece, en un formato distinto al anterior, la ordenación de necesidades formativas para el área A “Gestión y administración deportiva” del IAD en el año 2006.

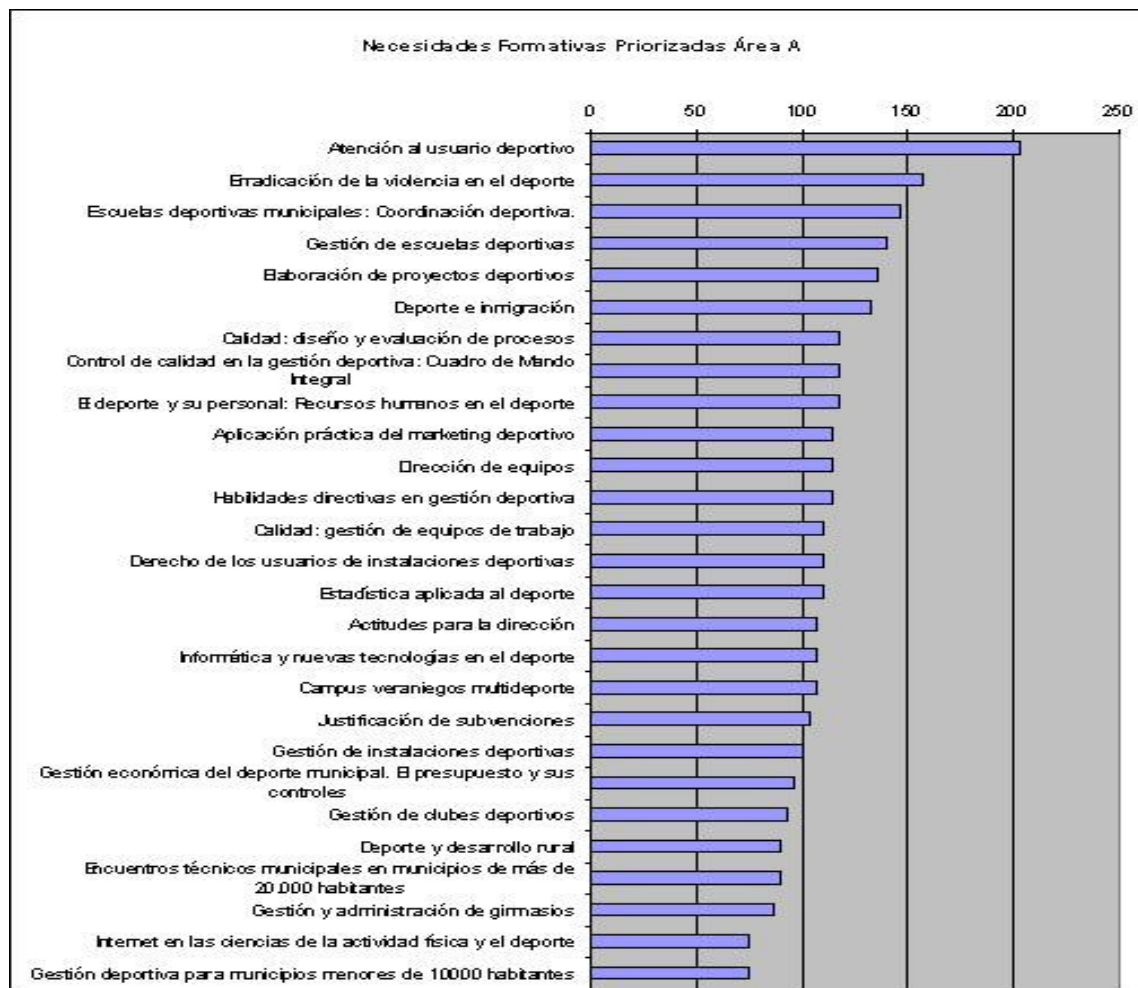


Figura 5.2. Orden de priorización de las acciones formativas del área de conocimiento A “Gestión y administración deportiva” en el año 2006 en el IAD.

Como puede verse, la acción formativa priorizada como más relevante fue “Atención al usuario deportivo” seguida de “Erradicación de la violencia en el deporte”, mientras que las consideradas menos relevantes fueron “Internet en las ciencias de la

actividad física y el deporte” y “Gestión deportiva para municipios menores de diez mil habitantes”.

4.4. Discusión.

Algunas **mejoras** que se han ido logrando a lo largo del tiempo en esta fase de post-evaluación, son las siguientes:

Respecto a los **participantes**, se puede decir que el proceso se fue haciendo más participativo porque en los primeros años sólo podían dar su opinión altos cargos, responsables de formación y representantes sindicales y, sin embargo, en el último año todos los implicados fueron considerados potenciales usuarios. Esto hizo que el número de participantes en Diputación aumentara considerablemente del año 2005 donde se obtuvo respuesta de 21 personas al 2006, con 374 participantes.

Sin embargo, este aumento en la participación posiblemente no se debió únicamente al aumento de la población, sino también a los cambios en el **instrumento**. El esfuerzo que se necesitaba para cumplimentar el cuestionario disminuyó considerablemente cuando, del año 2005 al 2006, se pasó de solicitar la priorización de todas las acciones formativas a la ordenación de únicamente 10 de cada área formativa, y sólo de aquellas áreas que resultaran de interés. Por ejemplo, quizá a una persona que trabaja en Servicios Sociales con menores no le interese el área de “Urbanismo y medio ambiente” por ejemplo y, además, no tenga suficiente conocimiento acerca de la materia como para poder opinar acerca de la urgencia de las necesidades en dicha área.

Otra razón del aumento de participación pudo ser el cambio en la **vía de entrega y recogida** del instrumento. En Diputación, a pesar de que se usó Internet en todo momento, su utilización fue distinta: en los años 2004 y 2005 las personas, tras cumplimentar el cuestionario, tuvieron que entregarlo por fax, correo electrónico o en mano en la Sección de Formación Continua. Sin embargo, en el 2006, y gracias a la colaboración de INPRO (Informática Provincial), se logró crear un cuestionario on-line más sofisticado donde, tras su cumplimentación (simplemente marcando con el ratón en distintos apartados), se enviaba automáticamente, de manera que supuso un esfuerzo mucho menor que en años atrás. A esto se unió el hecho de que los trabajadores de Diputación accedían con mayor frecuencia a la página web ya que todo el que quisiera solicitar un curso o confirmar asistencia tenía que hacerlo por dicho medio, con lo que más personas se encontraron con el aviso de que podían opinar acerca de la priorización para el plan del año siguiente. En el IAD, sin embargo, se mantiene el envío por correo electrónico, que puede ser una de las causas de que el aumento en la participación no fuera tan abismal.

Concretamente en el IAD, otra causa por la que pudo aumentar la participación del año 2005 al 2006 fue el hecho de que, al ser una fase de nueva implantación, las personas aún no estaban familiarizadas con el proceso, no habían sido concienciadas de la utilidad al aportar la información que se solicitó o sencillamente no habían sido informadas. En el segundo año, el aumento fue considerable.

Otra mejora a lo largo del tiempo en Diputación fue el aumento en la **transparencia** del proceso. Así, entre 1999 y 2002, la recogida de información se hacía

a través de reuniones; aunque también este instrumento puede tener ventajas, algunos inconvenientes saltan a la vista. Por ejemplo, al hacerse a puertas cerradas y no permitirse la participación de todos los implicados, se corría el riesgo de que se tomaran decisiones injustas, sin criterios claros o que se manipularan los datos obtenidos; además, no quedaba constancia por escrito de lo decidido y si se hacía no era documentación disponible para cualquier interesado. Con el uso de los cuestionarios, la participación se abrió a todos y posibles quejas por no estar de acuerdo con las decisiones tomadas eran fácilmente refutables al quedar constancia de lo solicitado, ya fuera en papel o en una base de datos.

Respecto al **instrumento** utilizado, aparte de haberse simplificado como ya se comentó anteriormente, mejoró al diferenciarse en el proceso de detección de necesidades las tres fases de evaluación, puesto que el cuestionario de priorización se elaboró a partir de los resultados obtenidos en la fase anterior de evaluación principal, con lo que se dotó de coherencia al proceso y se fue fiel al modelo seguido.

Respecto a la **codificación**, también se mejoró en Diputación en el año 2006 ya que, hasta entonces, se había hecho de manera manual, y en esta ocasión se logró que los datos fueran recogidos directamente en una base de datos. Con esto, se logró una mayor rapidez en la respuesta esperada por los implicados y una mayor fiabilidad, ya que no daba lugar a errores de codificación.

Respecto al **análisis de datos**, se logró obtener un sistema de priorización objetivo, basado en los datos obtenidos, donde no cabía la opinión personal y el subjetivismo.

A pesar de las mejoras ya logradas en esta fase de post-evaluación, siempre se puede avanzar y, a este respecto, se espera implementar en un corto periodo de tiempo las **mejoras para el futuro** que a continuación se detallan:

- Respecto al **feedback** a los implicados de la información recogida, se espera avanzar un poco más en el campo de las nuevas tecnologías y tratar de crear, como ya se comentó en la fase de evaluación principal, una base de datos que permita la visualización automática e inmediata a través de Internet de los datos recogidos, con lo que los usuarios podrían informarse de manera actualizada de la situación de priorización en la que se encuentran las acciones formativas.
- También respecto a la **difusión de los resultados**, se pretende dejar disponible en la página web de ambas organizaciones el informe de detección de necesidades, que recoge de manera pormenorizada los objetivos que se trataron de lograr en cada fase, el método utilizado y los resultados obtenidos. Se tratará de que sea un documento adaptado a las personas que participaron en el proceso o que pudieran estar interesadas, por lo que se intentará que sea sencillo, legible, evitando los tecnicismos pero sin perder por ello objetividad y concreción.
- En las mejoras a realizar en la fase de evaluación principal, ya se comentó que se trataría de complementar la información recogida con **otros instrumentos** que aportaran datos cualitativos. En esta fase, también se quieren implementar, concretamente mesas técnicas de comités de expertos en las distintas áreas, para el desarrollo de los contenidos de las acciones formativas solicitadas. De este modo, se lograría un mayor ajuste con el modelo teórico. En el IAD, en el año

2006, se realizó una prueba piloto a este respecto y los resultados fueron satisfactorios.

- En referencia al **análisis de datos** realizado, desde el año 2003 se utilizó concretamente el método de intervalos sucesivos de Thurstone (Barbero, 1993), con lo que se consiguió priorizar las acciones formativas en función de la importancia otorgada. De este modo, se aportó un indicador más para tomar la decisión final, aparte de la frecuencia recogida en la fase de evaluación principal. En el año 2006, sin embargo, se cambió al método de Dunn-Rankin (Arce, 1994) que se consideró, siendo de cálculo más sencillo, igual de preciso en los resultados. A este respecto, como se comentó previamente, en el apartado 6 de este capítulo se hace un estudio comparativo de las distintas técnicas de escalamiento con la intención de obtener datos fundamentados para tratar de determinar qué método de priorización sería el más adecuado para el caso concreto que nos ocupa.

5. DISEÑO DEL PLAN DE FORMACIÓN.

Las tres fases que según el modelo adaptado de Witkin y Altschuld (Chacón, Lara y Pérez, 2002c) conforman el proceso de detección de necesidades han sido ya explicadas. Sin embargo, ya se comentó que el diseño del plan realmente no termina aquí, sino que las decisiones finales son tomadas por los encargados de la formación en ambas organizaciones.

Sin embargo, el equipo técnico de evaluación trató de mantener la transparencia incluso en este último periodo, donde existe más cabida a la subjetividad. Para ello, básicamente se trataron de explicitar los criterios por los que se tomaron las decisiones finales.

En definitiva, a continuación se presentan algunos detalles del modo de actuación en este último momento del proceso de detección de necesidades, donde se tomaron las decisiones finales acerca de qué acciones formativas conformarían el plan de formación.

5.1. Objetivos.

En esta fase, el objetivo fue diseñar el plan de formación definitivo en función de los recursos disponibles. De manera resumida, se puede decir que esta decisión final se basó en los resultados obtenidos en las tres fases de detección de necesidades pero que, además, se siguieron otros criterios que llevaron a modificaciones de los resultados encontrados.

5.2. Método.

5.2.1. Muestra.

Las personas participantes en este diseño final fueron, en Diputación de Sevilla, la denominada Comisión de Formación, conformada por un representante de cada Sindicato, el Jefe de la Sección de Formación Continua y los técnicos de dicha sección. Se considera que este grupo de personas en su conjunto tiene suficiente conocimiento sobre formación, sobre las necesidades de los trabajadores de Diputación de Sevilla y sobre qué acciones formativas pueden tener éxito y cuáles no, como para tomar decisiones acertadas partiendo de los datos previamente recogidos.

En el IAD, el primer año (2005) guardó semejanza con Diputación respecto a la muestra en esta fase de diseño del plan final. Sin embargo, en el año 2006 colaboraron profesionales habituales y de contrastada eficiencia y reconocida valía en concretas áreas formativas de interés.

5.2.2. Instrumentos.

El punto de partida fueron los resultados obtenidos en la fase de post-evaluación; es decir, el listado de todas las acciones formativas solicitadas con el orden de priorización asignado.

5.2.3. Procedimiento.

En primer lugar, se planificó una serie de reuniones; los participantes fueron convocados y se les habían hecho llegar los resultados encontrados en la fase de post-evaluación.

El primer aspecto a tener en cuenta era el número de acciones formativas que podrían realizarse al siguiente año, en función del presupuesto. Se decidía entonces tomar un número proporcional de acciones formativas de cada área; así, por ejemplo, en la tabla 5.8 se puede ver el número de acciones formativas que fueron solicitadas en cada área, el porcentaje respecto al total que ello suponía y el número de acciones formativas que habría en cada área en el plan definitivo teniendo en cuenta que, de manera estimada, los recursos podrían cubrir aproximadamente la realización de 60 acciones formativas.

Área formativa	Nº cursos solicitados	% que representa del total	Nº cursos en el plan final
A. Gestión y administración deportiva	103	34%	20
B. Medicina deportiva	49	16%	10
C. Infraestructura deportiva	21	7%	4
D. Turismo y deporte	5	2%	1
E. Actividad física y deportiva	104	34%	20
F. Formación de formadores	21	7%	4
TOTAL	303	100%	59

Tabla 5.8. Número de acciones formativas solicitadas en cada área y número de acciones en el plan definitivo del IAD para el año 2007.

En principio, podría parecer que las decisiones estarían ya tomadas. Sin embargo, en este punto es cuando entran las opiniones de los responsables y expertos

con la intención de aumentar la eficiencia del programa a la hora de distribuir los recursos. Principalmente, dos fueron los criterios seguidos a la hora de modificar los resultados obtenidos con las fases anteriores:

- **Fusión** de distintas acciones formativas bajo un mismo nombre. Así, por ejemplo, en el área A en el 2006, se fusionaron las siguientes acciones formativas en el IAD: “Dirección de equipos”, “Habilidades directivas en gestión deportiva”, “Calidad: gestión de equipos de trabajo”, “Actitudes” y “Gestión de instalaciones deportivas”. La acción resultante fue denominada “Dirección de equipos”. De este modo se daba lugar a la realización de una acción formativa con el contenido de varias de tal modo que, con menor coste, se daba la posibilidad de responder al mismo número de solicitudes.
- **Cambio de área temática.** En Diputación, donde las áreas formativas estaban más consolidadas, no ocurría esto muy a menudo; en el IAD fue un criterio seguido más comúnmente. Consistió en modificar una acción formativa de área con la intención de que los contenidos tuvieran un mejor ajuste con el marco.

Aun con estos cambios, ha de decirse que se siguió el orden estricto de priorización obtenido en la fase de post-evaluación.

Pero, en ocasiones, lo más difícil no fue tanto la decisión de qué acciones formativas incluir, sino qué objetivos y contenidos eran adecuados impartir en base a la necesidad presentada por los potenciales destinatarios. Es por ello que, como experiencia piloto, en el IAD durante el año 2006 se realizaron técnicas nominales de grupos de discusión y de grupo focal donde se intentó que unos expertos en cada una de las áreas perfilaran un esquema de contenidos para cada acción formativa.

5.3. Resultados.

Los resultados finales consistieron en el conjunto de acciones formativas que finalmente conformarían el plan cada año, además de un diseño concreto para cada una en el caso del IAD en el año 2006.

5.4. Discusión.

Respecto a las mejoras realizadas en esta fase concreta de diseño final, un punto fundamental a destacar es que, a pesar de incluirse cambios en los resultados encontrados en la fase de post-evaluación, los criterios seguidos estuvieron explícitamente definidos, con lo que esto supuso de transparencia.

También respecto a los criterios, cabe destacar que fueron siendo menos y más concretos cada año, con lo que se puede decir que los planes finales se ajustaron cada vez más a los resultados obtenidos en la fase de post-evaluación.

En el diseño final, apenas aparecieron acciones formativas que fueran incluidas a posteriori, sino que todas provinieron del estudio de detección de necesidades previamente realizado. Esta mejora se fue logrando a lo largo del tiempo pues, por

ejemplo en el IAD, en el año 2005 alguna acción formativa fue excluida a posteriori, mientras que en el 2006 esto ya no ocurrió.

En líneas generales, teniendo en consideración que esta fase final es altamente susceptible de sufrir sesgos y subjetividades en las tomas de decisiones, se puede decir que el desarrollo a lo largo del tiempo fue positivo, llegándose cada vez más a una situación de claridad respecto a las acciones realizadas.

Sin embargo, a pesar de estas mejoras, algunas más están en fase de desarrollo. Así, por ejemplo, tras los buenos resultados obtenidos con las mesas técnicas formadas por expertos diseñadas y ejecutadas en el 2006 en el IAD, se espera poder continuar en esta línea e implementarlas también en Diputación. En esta fase, es fundamental tener en cuenta la opinión de los expertos en las diversas materias de interés, ya que son ellos los que realmente tienen capacidad para determinar qué contenidos son los apropiados para cubrir las necesidades de formación de los empleados.

6. ESTUDIO COMPARATIVO DE TÉCNICAS DE ESCALAMIENTO.

6.1. Objetivos.

Como ya se comentó en el apartado referido a la post-evaluación, se realizaron dos análisis de datos distintos para un mismo fin: la priorización del listado de acciones formativas recogidas. En los años 2004 y 2005, se llevó a cabo el método de los intervalos sucesivos, derivado de la ley de juicios categóricos de Thurstone (Barbero, 1993); sin embargo, considerando que era más sencillo que el anterior y aportaba la misma información, se realizó en el año 2006 el método de Dunn-Rankin (Arce, 1994).

Ahora bien, ¿podemos estar seguros de que este último es el método más apropiado? Para responder a esto, en este apartado se presenta la comparación de los resultados obtenidos en la priorización utilizando diferentes métodos, concretamente el método de Dunn-Rankin (Arce, 1994) y tres derivados de la Ley de juicios categóricos de Thurstone (Barbero, 1997): el método de ordenación por rangos, el método de los intervalos sucesivos y el método de los intervalos aparentemente iguales.

En definitiva, el **objetivo** de este apartado es determinar el método de priorización más parsimonioso; es decir, qué análisis tiene la mejor relación “nivel de complejidad-información obtenida” (que siendo sencillo, obtenga un nivel adecuado de información).

6.2. Método.

6.2.1. Muestra.

Se tomaron para este estudio los datos recogidos en la fase de priorización desde el 2004 al 2006 en Diputación de Sevilla. Se eligieron concretamente los resultados disponibles de esta organización para el análisis porque bastaron para estudiar las

distintas posibilidades que interesaban: si habría variación en función del número de participantes y si habría diferencias en los resultados en función del número de acciones formativas a priorizar. Como ya se comentó en el apartado referido a la postevaluación, concretamente en Diputación se recogieron 9 cuestionarios en el año 2004, 21 en el 2005 y 374 en el 2006.

6.2.2. Instrumentos.

Los instrumentos utilizados fueron los que a continuación se enumeran:

- El **cuestionario de priorización** ya comentado en el apartado referido a la postevaluación. Aparecía el listado de acciones formativas que habían sido solicitadas en la evaluación principal, organizadas por áreas formativas. Concretamente, el número de acciones formativas priorizadas y su distribución por áreas se presenta en la tabla 5.9.

AÑO	A	B	C	D	E	F	G	H	I	J	K	TOTAL
2004	4	4	65	0	2	2	25	2	4	0	9	117
2005	2	5	37	0	2	5	15	0	3	2	4	75
2006	12	19	136	4	4	33	41	7	10	4	16	286

Tabla 5.9. Acciones formativas priorizadas por cada área en Diputación de Sevilla en los distintos años.

La primera columna diferencia entre los distintos años. La última columna presenta el total de acciones formativas; las columnas intermedias, el número de acciones formativas que se priorizaron en cada área formativa, siendo éstas las siguientes: “A. Dirección y Gerencia Pública”, “B. Económico Presupuestaria”, “C. Específico de Determinados Colectivos”, “D. Idiomas/Lengua”, “E. Información y Atención al Ciudadano”, “F. Jurídico – Procedimental”, “G. Nuevas Tecnologías. Información y Comunicación”, “H. Prevención de Riesgos Laborales”, “I. Recursos Humanos”, “J. Unión Europea” y “K. Urbanismo y Medio Ambiente”.

- Los softwares **Microsoft Excel** y **SPSS 14** para el análisis de datos.

6.2.3. Procedimiento.

Hasta el 2005 se solicitó a los participantes a través de los cuestionarios, en formato papel o a través de Internet, que priorizaran las acciones formativas desde la más relevante a la menos necesaria; en el 2006, se solicitó una puntuación de 1 a 10 a únicamente diez acciones formativas por área, siendo 10 el valor más positivo. Los datos recogidos fueron exportados a Excel, programa con el que se llevaron a cabo todos los análisis de datos. Con el SPSS se calcularon posteriormente la significación de las diferencias entre los resultados obtenidos con los distintos métodos. A continuación, se presenta una descripción pormenorizada de los diferentes métodos de análisis utilizados.

El método de Dunn-Rankin.

Concretamente, los pasos a seguir en el método de Dunn-Rankin son los siguientes:

1. **Ordenación por preferencias:** las columnas representan las distintas acciones formativas y las filas, la ordenación, siendo el 1 el valor más positivo y, en el caso del ejemplo que a continuación se presenta, el 4 el más bajo al haber cuatro estímulos que priorizar. Así, por ejemplo, en la tabla 5.10 se aprecia que la primera persona priorizó como más importante el curso de Word, seguido de Openoffice, después Arc View y, finalmente, Wordperfect.

	Word	Openoffice	Arc View	Wordperfect
Sujeto 1	1	2	3	4
Sujeto 2	2	1	3	4
Sujeto 3	1	2	3	4
Sujeto 4	2	1	3	4
Sujeto 5	4	2	3	1
Sujeto 6	2	1	3	4
Sujeto 7	2	1	3	4
Sujeto 8	4	3	1	2
Sujeto 9	1	2	3	4

Tabla 5.10. Ejemplo de priorización (método de Dunn-Rankin).

2. **Transformación de los valores:** a continuación, se trata de que los números que representan las prioridades tengan la misma dirección que las prioridades en sí; es decir, si el 1 es el priorizado como más importante, ahora obtendrá el valor numérico más elevado. Para ello, cada valor es restado por el número de estímulos más 1, siendo en el ejemplo $4+1=5$. Tras esta transformación, los valores quedarían tal como se muestran en la tabla 5.11:

	Word	Openoffice	Arc View	Wordperfect
Sujeto 1	4	3	2	1
Sujeto 2	3	4	2	1
Sujeto 3	4	3	2	1
Sujeto 4	3	4	2	1
Sujeto 5	1	3	2	4
Sujeto 6	3	4	2	1
Sujeto 7	3	4	2	1
Sujeto 8	1	2	4	3
Sujeto 9	4	3	2	1
Σ	26	30	20	14
Σ/\max	0.72	0.83	0.56	0.39
%	72%	83%	56%	39%
Orden	2º	1º	3º	4º

Tabla 5.11. Transformación de valores (método de Dunn-Rankin).

3. **Cálculo del sumatorio de los valores transformados:** la décima fila de la tabla 5.11 (representada por la letra griega Σ), presenta la suma de los valores una vez transformados. Con este valor, ya se puede tomar la decisión: cuanto mayor sea, con mayor importancia los participantes en conjunto clasificaron a la acción formativa en cuestión. De esta manera, en el ejemplo presentado, Openoffice sería la acción formativa priorizada en primer lugar (como la más relevante), habiendo obtenido un

valor de 30; a continuación le seguirían, en orden de mayor a menor importancia, Word (con un valor de 26), Arc View (20) y Wordperfect (14). Las últimas tres filas presentadas en la tabla 5.11 se realizan en este método con la finalidad de que el resultado sea más comprensible de manera intuitiva.

4. **Cálculo de los valores relativos:** la antepenúltima fila de la tabla 5.11 (Σ/\max), presenta el valor obtenido para cada acción formativa al dividir el valor de la columna anterior (el sumatorio) entre el valor máximo posible. En este ejemplo, este valor máximo es 36, ya que si los nueve participantes hubieran dado prioridad 1 a una acción formativa, se hubieran sumado nueve veces el valor 4 (que corresponde al 1 tras la transformación).
5. **Cálculo del porcentaje:** en la penúltima fila de la tabla 5.11 aparece el porcentaje (%) que se obtiene al multiplicar por 100 el valor obtenido en la fila anterior, de tal manera que el valor máximo posible mostrado en esta fila sería 100 (en el caso en que todos los participantes hubieran dado prioridad máxima a la misma acción formativa).
6. **Ordenación:** finalmente, en la última fila de la tabla 5.11 se presenta cómo quedaría la ordenación de las cuatro acciones formativas en función de las prioridades que los nueve participantes dieron a cada una de ellas.

La ley de juicios categóricos de Thurstone.

Este modelo categoriza estímulos, en este caso acciones formativas, en función del número de veces que las personas asignan cada estímulo a cada una de las categorías. Algunos de los supuestos de los que parte son los siguientes:

- El atributo en estudio, en este caso la relevancia o utilidad de una acción formativa para mejorar el desempeño en el puesto de trabajo, se presenta en un **continuo subjetivo o psicológico**.
- Este continuo puede dividirse a su vez en **categorías ordenadas**.
- Cada persona participante **asignará un valor subjetivo** en dicho continuo a cada estímulo o acción formativa en este caso. Su trabajo consistirá, concretamente, en dar una categoría a cada acción formativa en función del grado de utilidad que subjetivamente le otorgue.
- La respuesta de todas las personas a un estímulo formará una **distribución de frecuencias** que se asume conforma una distribución normal definida por su media y su desviación típica.
- La media de los valores otorgados a cada estímulo se denomina **distribución discriminante**, y no es otra cosa que el lugar que el estímulo ocupa en el continuo psicológico; a esto se le denomina **valor escalar del estímulo**. La desviación típica de cada estímulo se denomina **dispersión discriminante** y nos da idea del grado de acuerdo que ha habido entre los distintos participantes a la hora de categorizar cada estímulo.

Todo esto se traduce a una serie de ecuaciones que, una vez simplificadas, posibilita la aplicación práctica de este modelo (Barbero, 1993). Concretamente, se pueden diferenciar tres procedimientos para el análisis empírico de los datos: el método de ordenación por rangos, el método de los intervalos sucesivos y el método de los intervalos aparentemente iguales que, a continuación, se pasan a describir pormenorizadamente:

Método de ordenación por rangos.

Este método es muy utilizado por su sencillez. Los pasos a seguir para su realización se muestran a continuación, ilustrados con un ejemplo que parte de los datos utilizados en el apartado anterior referido al método de Dunn-Rankin.

1. Elaboración de la matriz de datos. En la tabla 5.12 las líneas son los distintos participantes, mientras que cada columna hace referencia a las diferentes acciones formativas. Así, por ejemplo, el sujeto 1 consideró que el curso más relevante era el de Microsoft Word, seguido de Openoffice, Arc View y, finalmente, Wordperfect.

	Word	Openoffice	Arc View	Wordperfect
Sujeto 1	4	3	2	1
Sujeto 2	3	4	2	1
Sujeto 3	4	3	2	1
Sujeto 4	3	4	2	1
Sujeto 5	1	3	2	4
Sujeto 6	3	4	2	1
Sujeto 7	3	4	2	1
Sujeto 8	1	2	4	3
Sujeto 9	4	3	2	1

Tabla 5.12. Matriz de datos (método de ordenación por rangos).

2. Suma de los órdenes asignados a cada estímulo. Sencillamente, consistiría en sumar las columnas, de tal manera que el resultado quedaría tal como se presenta en la tabla 5.13:

	Word	Openoffice	Arc View	Wordperfect
Sujeto 1	4	3	2	1
Sujeto 2	3	4	2	1
Sujeto 3	4	3	2	1
Sujeto 4	3	4	2	1
Sujeto 5	1	3	2	4
Sujeto 6	3	4	2	1
Sujeto 7	3	4	2	1
Sujeto 8	1	2	4	3
Sujeto 9	4	3	2	1
$\Sigma =$	26	30	20	14

Tabla 5.13. Suma de los órdenes asignados a cada estímulo (método de ordenación por rangos).

3. Ordenación de los estímulos en función de las sumas obtenidas, teniendo en cuenta que la acción formativa que haya obtenido la suma más alta en este caso será la considerada como más prioritaria. De este modo, en el ejemplo se ha obtenido que

la acción formativa más relevante es Openoffice, seguida de Microsoft Word, Arc View y, finalmente, Wordperfect.

Método de los intervalos sucesivos.

Los pasos a seguir en el método de los intervalos sucesivos son los siguientes:

1. **Creación de una matriz de frecuencias:** en la tabla 5.14 que a continuación se presenta, aparece un posible ejemplo referido a hipotéticas acciones formativas priorizadas por sujetos en el área G, “Informática. Nuevas tecnologías” de Diputación de Sevilla, suponiendo que 10 personas las priorizaron desde nada necesaria (1) hasta muy necesaria (5). Las filas representan las distintas acciones formativas y las columnas, la valoración. Los números son la frecuencia en que cada acción formativa fue elegida en cada valoración. Así, por ejemplo, 6 personas consideraron la acción formativa referida a Word como muy necesaria.

A. FORMATIVA	1	2	3	4	5
Word	0	1	1	2	6
Openoffice	1	1	1	3	4
Arc View	1	2	6	1	0
Wordperfect	3	3	1	3	0

Tabla 5.14. Matriz de frecuencia (método de los intervalos sucesivos).

Hay que tener en cuenta que, a pesar de que aquí se han utilizado cinco categorías, podría ser un número distinto.

2. **Cálculo de la matriz de frecuencias acumuladas.** Para ello, se van sumando las casillas de izquierda a derecha, de tal manera que cada casilla contendrá el número de personas que consideraron a la acción formativa concreta por encima de un orden determinado. Así, por ejemplo, observando la tabla 5.15, podría decirse que 7 personas han valorado la acción formativa referida a Wordperfect con un 3 o una puntuación menor. Una forma de confirmar que no ha habido confusión en los cálculos es el asegurarnos de que, en la última columna, todas las casillas contienen el número total de personas participantes; en este caso, en la columna 5, todas las casillas contienen el valor 10 (número total de participantes).

A. FORMATIVA	1	2	3	4	5
Word	0	1	2	4	10
Openoffice	1	2	3	6	10
Arc View	1	3	9	10	10
Wordperfect	3	6	7	10	10

Tabla 5.15. Matriz de frecuencias acumuladas (método de los intervalos sucesivos).

3. **Cálculo de la matriz de proporciones acumuladas.** Para ello, se divide cada valor de la tabla anterior entre el número total de participantes. El valor que obtenemos es la proporción de personas que valoraron una acción formativa determinada con un valor concreto y los inferiores a éste. Haciendo referencia a la tabla 5.16, se podría decir que, por ejemplo, siendo 1 el valor total, 0.3 respuestas eligen a la acción formativa “Arc

View” con un valor de 2 o menor. En términos de porcentaje, se interpretaría como que el 30% del total de personas participantes eligieron la acción formativa “Arc View” con una valoración de 1 ó 2. Para la interpretación en porcentajes, únicamente habría que multiplicar por 100 los valores de la tabla 5.16 que a continuación se presenta:

A. FORMATIVA	1	2	3	4	5
Word	0	0.1	0.2	0.4	1
Openoffice	0.1	0.2	0.3	0.6	1
Arc View	0.1	0.3	0.9	1	1
Wordperfect	0.3	0.6	0.7	1	1

Tabla 5.16. Matriz de proporciones acumuladas (método de los intervalos sucesivos).

4. **Transformación de las proporciones en puntuaciones típicas Z.** Para ello, habría que buscar cada valor en la tabla correspondiente a las áreas bajo la curva normal. Esta distribución se caracteriza porque la media de los datos corresponde al valor 0 y la desviación típica es 1. Un valor de 0.5, por tanto, obtendrá una puntuación típica Z de 0, por lo que todo valor menor de 0.5 obtendrá una puntuación típica Z negativa y todo valor mayor que 0.5, positiva. Una proporción de 0 y 1 no será transformada a puntuación típica porque carece de significado; generalizando esta lógica a valores cercanos, tampoco se transformarán las proporciones mayores de 0.98 ni las menores de 0.02. La tabla 5.17 muestra las puntuaciones típicas obtenidas de las proporciones de la tabla 5.16:

A. FORMATIVA	1	2	3	4	5
Word	-	-1.28	-0.84	-0.25	-
Openoffice	-1.28	-0.84	-0.52	0.25	-
Arc View	-1.28	-0.52	1.28	-	-
Wordperfect	-0.52	0.25	0.52	-	-

Tabla 5.17. Puntuaciones típicas (método de los intervalos sucesivos).

5. **Cálculo de la amplitud de cada intervalo.** Consiste en restar, para cada acción formativa, el valor obtenido en una posición concreta menos el que se le dio en la posición anterior a ésta. Así, por ejemplo, la amplitud del primer intervalo para la acción formativa Openoffice sería $-0.84 - (-1.28) = 0.44$. La amplitud de todos los intervalos se presenta en la tabla 5.18:

A. FORMATIVA	2 - 1	3 - 2	4 - 3
Word	-	0.44	0.59
Openoffice	0.44	0.32	0.77
Arc View	0.76	1.8	-
Wordperfect	0.77	0.27	-

Tabla 5.18. Amplitud de intervalos (método de los intervalos sucesivos).

6. **Cálculo de la amplitud de cada intervalo concreto en la escala general:** una vez que se conoce la amplitud de los intervalos para cada acción formativa, el siguiente paso es averiguar la amplitud de cada uno de los intervalos en la escala total. Para ello, se halla la media de todas las amplitudes (una distinta por cada acción formativa). Así, en el ejemplo, la amplitud en la escala del intervalo asociado a la puntuación 2 (segundo intervalo) es:

$$\frac{0.44 + 0.76 + 0.77}{3} = 0.66$$

Para el intervalo asociado a la puntuación 3 (tercer intervalo), la amplitud es:

$$\frac{0.44 + 0.32 + 1.8 + 0.27}{4} = 0.71$$

Por último, para el intervalo asociado a la puntuación 4 (cuarto intervalo), la amplitud es:

$$\frac{0.59 + 0.77}{2} = 0.68$$

7. Cálculo de los límites de cada categoría en la escala: para ello, se sitúa el primer valor obtenido como límite superior del segundo intervalo, la suma de las dos primeras amplitudes como límite superior del tercero y la suma de las tres amplitudes como límite superior del cuarto intervalo. Así, la escala obtenida sería la que se presenta como figura 5.3:

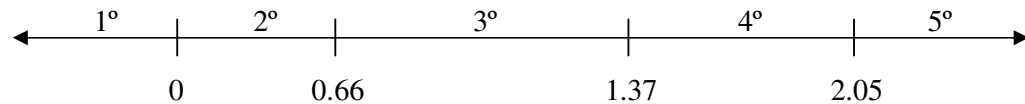


Figura 5.3. Límites de las categorías (método de los intervalos sucesivos).

En este caso, el valor del límite superior del tercer intervalo (1.37) es el resultante de sumar las dos primeras amplitudes (0.66 + 0.71) y el valor del límite superior del cuarto intervalo (2.05) resulta de sumar al 1.37 la amplitud restante (0.68).

8. Cálculo de los valores escalares de los estímulos: para conocer el lugar exacto que ocupa cada acción formativa, se calcula la mediana para cada uno de ellas, cuya fórmula es:

$$Med = L_i + \frac{I(N/2 - f_i)}{f_b}$$

siendo:

- L_i = Límite inferior del intervalo asociado a la categoría que deja por encima y por debajo al 50% de las personas de la muestra.
- I = Amplitud del intervalo asociado a la categoría correspondiente a la mediana.
- N = número de personas que respondieron al cuestionario de priorización.

- f_i = número de personas que clasificaron la acción formativa concreta por debajo de la categoría en la que se va a encontrar la mediana (aquel que deja por encima y por debajo al 50% de las personas).
- f_b = número de personas que clasificaron la acción formativa en la categoría correspondiente a la mediana.

Así, las medianas de cada una de las acciones formativas del ejemplo son:

$$Med(Word) = 2.05 + \frac{0.68(5-4)}{6} = 2.16$$

$$Med(Openoffice) = 1.37 + \frac{0.68(5-3)}{3} = 1.82$$

$$Med(ArcView) = 0.66 + \frac{0.71(5-3)}{6} = 0.9$$

$$Med(Wordperfect) = 0 + \frac{0.66(5-3)}{3} = 0.44$$

Por ejemplo, la mediana de la acción formativa “Arc View” se ha calculado del siguiente modo:

- El intervalo en el que se encuentra la mediana (más del 50% de las personas participantes, es decir, más de 5), es el de la puntuación 3, ya que su frecuencia acumulada es 9, mientras que en el intervalo anterior únicamente se acumulaban 3 personas. El 0.66 que aparece en el cálculo hace referencia al límite inferior para este intervalo 3º, como puede verse en la escala.
- El siguiente 0.71 hace referencia a la amplitud de este tercer intervalo.
- 5 es la mitad del número total de personas que priorizaron las acciones formativas (en este caso, 10 personas).
- 3 es el número de personas que puntuaron a la acción formativa con una valoración de 1 ó 2 (en los intervalos anteriores a aquél en el que se encuentra la mediana; en este caso, el de la puntuación 3).
- 6 es el número de personas que valoraron a la acción formativa “Arc View” con un 3 (en el intervalo donde se encuentra la mediana).

De este modo, la posición de cada acción formativa aparecería en la escala general como se muestra en la figura 5.4:

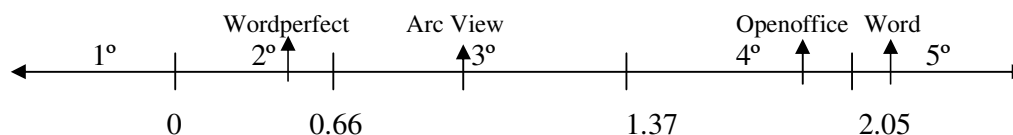


Figura 5.4. Ordenación resultante (método de los intervalos sucesivos).

La interpretación de estos datos es que “Word” ha sido la acción formativa priorizada como más importante, seguida por “Openoffice”, “Arc View” y “Wordperfect”.

Este tipo de análisis es especialmente útil cuando se trata de priorizar muchas acciones formativas con frecuencias semejantes, ya que produce unos resultados muy precisos que permiten discriminar de manera clara.

Método de los intervalos aparentemente iguales.

Con este método, al igual que con el anterior, se trata de ordenar cada una de las acciones formativas a lo largo de un continuo con la diferencia de que, en este caso, la amplitud de todos los intervalos se considera la misma.

Así, suponiendo por ejemplo que las personas colocaran cuatro estímulos en una categoría de 1 a 5, siendo la 1 la más desfavorable y la 5 la más favorable, y suponiendo que cada categoría tuviera una amplitud de 1, el continuo quedaría tal como se muestra en la figura 5.5:

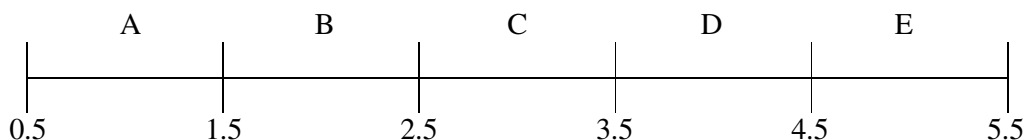


Figura 5.5. Cinco intervalos de igual amplitud (método de los intervalos aparentemente iguales).

Suponiendo que se parte de los mismos datos que en el ejemplo anterior, los pasos a seguir para realizar el escalamiento con este método son los siguientes:

1. Creación de la matriz de datos, donde cada línea es una acción formativa y cada columna, una categoría. Cada valor es la frecuencia; es decir, el número de veces que ha sido asignado cada curso a cada una de las categorías.

A. FORMATIVA	1	2	3	4	5
Word	0	1	1	2	6
Openoffice	1	1	1	3	4
Arc View	1	2	6	1	0
Wordperfect	3	3	1	3	0

Tabla 5.19. Matriz de datos (método de los intervalos aparentemente iguales).

2. Cálculo de la matriz de frecuencias acumuladas. Como ya se explicó en el método anterior, consiste en ir sumando las casillas de izquierda a derecha.

A. FORMATIVA	1	2	3	4	5
Word	0	1	2	4	10
Openoffice	1	2	3	6	10
Arc View	1	3	9	10	10
Wordperfect	3	6	7	10	10

Tabla 5.20. Matriz de frecuencias acumuladas (método de los intervalos aparentemente iguales).

3. Cálculo de la mediana, utilizando para ello la misma fórmula presentada en el método anterior:

$$Med = L_i + \frac{I(N/2 - f_i)}{f_b}$$

donde:

- L_i = Límite inferior del intervalo asociado a la categoría que deja por encima y por debajo al 50% de las personas de la muestra.
- I = Amplitud del intervalo asociado a la categoría correspondiente a la mediana. Realmente, al considerarse la amplitud de todos los intervalos igual, se trata de una constante, por lo que podría eliminarse de la fórmula sin causar por ello ninguna variación en la priorización resultante.
- N = número de personas que respondieron al cuestionario de priorización.
- f_i = número de personas que clasificaron a la acción formativa concreta por debajo de la categoría en la que se va a encontrar la mediana (aquél que deja por encima y por debajo al 50% de las personas).
- f_b = número de personas que clasificaron la acción formativa en la categoría correspondiente a la mediana.

Así, los valores de las medianas de cada una de las acciones formativas del ejemplo son:

$$Med(\text{Word}) = 4.5 + \frac{1(5-4)}{6} = 4.67$$

$$Med(\text{Openoffice}) = 3.5 + \frac{1(5-3)}{3} = 4.17$$

$$Med(\text{ArcView}) = 2.5 + \frac{1(5-3)}{6} = 2.83$$

$$Med(\text{Wordperfect}) = 1.5 + \frac{1(5-3)}{3} = 2.17$$

Equivalencia entre métodos.

Algunas equivalencias entre métodos, ya extendidamente conocidas, se presentan a continuación:

Equivalencia entre método de Dunn-Rankin – método de ordenación por rangos.

La ordenación obtenida con el método de Dunn-Rankin es exactamente la misma que la obtenida con el método de ordenación por rangos. La razón es que la lógica utilizada es la misma. La única diferencia radica en que el método de Dunn-Rankin presenta ciertas transformaciones tras la suma (se divide entre la puntuación máxima posible y, posteriormente, se multiplica por 100 para presentar el dato como porcentaje) mientras que, en el método de ordenación por rangos, se jerarquiza directamente tras la suma de las ordenaciones obtenidas por cada acción formativa.

Equivalencia entre método de los intervalos sucesivos – método de los intervalos aparentemente iguales.

El método de los intervalos sucesivos muestra una equivalencia en sus resultados casi perfecta con los obtenidos con el método de los intervalos aparentemente iguales (Barbero, 1993).

En apariencia, aunque la ordenación obtenida coincide perfectamente, los valores son muy diferentes. Sin embargo, al pasar los datos a la misma escala tras una transformación lineal (Ponsoda, 1986, cit. en Barbero, 1993), se comprueba que realmente los valores son bastante similares, como se muestra a continuación tomando el ejemplo previamente presentado. Los pasos concretos a seguir son los siguientes:

- 1. Cálculo de la ecuación de regresión que transforma los resultados obtenidos con el método de intervalos sucesivos.** Siendo X los valores obtenidos con el método de intervalos sucesivos e Y la transformación de éstos, habría que averiguar la ecuación de regresión:

$$Y = a + bX$$

Para ello se siguen los siguientes pasos:

- 2. Cálculo de la media y la desviación típica de los resultados obtenidos con ambos métodos.** Siguiendo las siguientes fórmulas,

$$\bar{X} = \frac{\sum X_i}{n}$$

$$S^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$

$$S = \sqrt{S^2}$$

se obtiene que la media, la varianza y la desviación típica de los resultados obtenidos con el método de los intervalos aparentemente iguales son, respectivamente,

$$\bar{Y} = \frac{4.67 + 4.17 + 2.83 + 2.17}{4} = 3.46$$

$$S_y^2 = \frac{4.67^2 + 4.17^2 + 2.83^2 + 2.17^2}{4} - 3.46^2 = 1.01$$

$$S_y = \sqrt{1.01} = 1$$

y la media, la varianza y la desviación típica de los resultados obtenidos con el método de los intervalos sucesivos son:

$$\bar{X} = \frac{2.16 + 1.82 + 0.9 + 0.44}{4} = 1.33$$

$$S_x^2 = \frac{2.16^2 + 1.82^2 + 0.9^2 + 0.44^2}{4} - 1.33^2 = 0.48$$

$$S_x = \sqrt{0.48} = 0.69$$

- 3. Cálculo de la a y la b de la ecuación de regresión.** Se conoce que, entre las propiedades de la media y la varianza de dos distribuciones X e Y relacionadas por una ecuación de regresión, se encuentran las siguientes:

$$\bar{Y} = a + b\bar{X}$$

$$S_y^2 = b^2 S_x^2$$

Sustituyendo, nos quedamos con un sistema de ecuaciones con dos incógnitas, a y b :

$$3.46 = a + b * 1.33$$

$$1.01 = b^2 * 0.48$$

Despejando la b de la segunda ecuación, obtenemos su valor:

$$b^2 = \frac{1.01}{0.48} = 2.1 \rightarrow b = 1.45$$

Despejando la a en la segunda ecuación y sustituyendo la b por su valor, obtenemos el valor de a :

$$a = 3.46 - b * 1.33 \rightarrow a = 3.46 - 1.45 * 1.33 \rightarrow a = 1.55$$

Por tanto, la ecuación de regresión resultante es:

$$Y = 1.53 + 1.45X$$

- 4. Obtención de los valores transformados.** Para obtener los valores que serán comparables con los calculados a través del método de los intervalos aparentemente iguales, es necesario sustituir la X en la ecuación por cada uno de los valores obtenidos con el método de los intervalos sucesivos. Los resultados de esta operación se muestran a continuación:

$$Y(\text{Word}) = 1.53 + 1.45 * 2.16 = 4.66$$

$$Y(\text{Openoffice}) = 1.53 + 1.45 * 1.82 = 4.17$$

$$Y(\text{ArcView}) = 1.53 + 1.45 * 0.9 = 2.83$$

$$Y(\text{Wordperfect}) = 1.53 + 1.45 * 0.44 = 2.17$$

- 5. Comparación de los resultados.** En la tabla 5.21 que a continuación se presenta, se muestran los resultados obtenidos para cada una de las acciones formativas con el método de intervalos aparentemente iguales (segunda columna) y con el método de intervalos sucesivos, pero tras la transformación.

ACCIÓN FORMATIVA	VALORES OBTENIDOS CON MÉTODO DE INTERVALOS APARENTEMENTE IGUALES	VALORES OBTENIDOS CON MÉTODO DE INTERVALOS SUCESIVOS (TRANSFORMADOS)
Word	4.67	4.66
Openoffice	4.17	4.17
Arc View	2.83	2.83
Wordperfect	2.17	2.17

Tabla 5.21. Resultados obtenidos con el método de intervalos aparentemente iguales y con el método de intervalos sucesivos, tras la transformación.

Como puede apreciarse, tras la transformación, las diferencias son ínfimas, lo que explica el hecho de que la ordenación de los estímulos con ambos métodos suela coincidir. Sin embargo, según la literatura (Barbero, 1993), no siempre se da esto ya que los estímulos con valores extremos suelen presentar diferencias mayores entre ambos métodos.

A continuación, utilizando como ya se explicó en el procedimiento los datos obtenidos en la fase de post-evaluación en Diputación de Sevilla durante los años 2004 a 2006, se tratará de confirmar algunas afirmaciones ya expuestas y se intentará responder a otras cuestiones. Concretamente, los puntos a tratar serán los siguientes:

- ¿Es cierto que no existen diferencias entre los resultados obtenidos con el método de Dunn-Rankin y el método de ordenación por rangos?
- ¿Existen ocasiones (por puntuaciones extremas) en que los resultados obtenidos por el método de los intervalos sucesivos y el método de los intervalos aparentemente iguales difieran (Barbero, 1993)?

- ¿Existen diferencias entre ambos pares de métodos? Se comparará Dunn-Rankin y ordenación por rangos en conjunto con intervalos sucesivos y aparentemente iguales por otro lado.
- En caso de que existan diferencias significativas entre métodos, ¿a qué se deben dichas diferencias? ¿Difieren los resultados según el número de personas que respondan? ¿Difieren en función del número de acciones formativas a priorizar? ¿O en función del número de categorías que se establezcan? ¿Influyen los datos perdidos?
- ¿Qué métodos son más precisos? Se estudiará el número de empates que presenten.
- ¿Qué métodos son más sencillos?
- Cruzando los dos criterios anteriores, ¿qué métodos son más parsimoniosos?

6.3. Resultados.

Para estudiar más pormenorizadamente las diferencias entre métodos y determinar cuál es más parsimonioso, en el **anexo VIII** (pág. xlix) se presentan las tablas que muestran los resultados obtenidos en la ordenación de las acciones formativas diferenciadas por áreas en Diputación de Sevilla desde los años 2004 a 2006.

Arriba de cada tabla, el primer dato que aparece entre paréntesis es el número de cuestionarios que se recogió (N). La primera columna presenta las acciones formativas divididas en las distintas áreas. Las cuatro restantes columnas muestran, para cada método, el orden en el que quedó cada acción formativa y, entre paréntesis, el valor numérico obtenido.

Hay que tener en cuenta que este último dato no es comparable entre métodos a simple vista, puesto que la escala que se utiliza es diferente. Por ello, a la hora de comparar los resultados, se atendió únicamente al orden.

En sombreado, aparece la casilla donde se obtuvieron resultados diferentes en cuanto a ordenación en comparación con otros métodos.

El asterisco que en ocasiones aparece entre paréntesis en la columna correspondiente al método de intervalos sucesivos indica que no fue posible calcular el valor concreto por limitaciones de dicho método; concretamente, no se pudo calcular el dato cuando la acción formativa se encontró en las categorías extremas porque los intervalos quedaban abiertos (la primera categoría no presentaba límite inferior mientras que la última no tenía el superior).

Para que estos casos no quedaran sin ordenación, cuando existía más de una acción formativa situada en la categoría más baja, se utilizó para desempatar entre ellas el cálculo de las medianas tomando como límite inferior (L_i según la fórmula) un valor menor que el límite inferior de la siguiente categoría y como amplitud (I según la fórmula) un valor menor a la amplitud más pequeña presentada por el resto de categorías; y al contrario, cuando el empate se encontró en la categoría más elevada, se

tomó como límite inferior un valor más elevado que el último límite conocido y, como amplitud, un valor más alto que la amplitud más grande.

Cuando se dieron valores perdidos en alguna acción formativa dentro de un área priorizada, se interpretó que fueron debidos a que dichas acciones no eran tan necesarias ni como para otorgarles el valor mínimo de priorización, por lo que todos los huecos fueron considerados valores mínimos; se les asignó por tanto automáticamente el valor de 1, puesto que dejar un vacío afectaba al cálculo de la priorización.

Un primer aspecto a mencionar es que, tal como se anticipó, las priorizaciones obtenidas con el **método de Dunn-Rankin** y el de **ordenación por rangos** resultaron ser exactamente iguales.

Idéntica situación se encontró al comparar los resultados obtenidos con el **método de los intervalos sucesivos** y el **método de los intervalos aparentemente iguales**. Aunque cuando los estímulos reciben puntuaciones extremas las diferencias en los resultados entre métodos suelen ser mayores, no se ha dado ni un solo caso en el que estas posibles diferencias fueran tan amplias como para provocar un cambio de orden de priorización.

Sin embargo, se detectaron algunas diferencias al comparar, por un lado, la ordenación obtenida con el método de Dunn-Rankin (coincidente con la obtenida con el método de ordenación por rangos) y, por otro lado, la ordenación tras aplicar el método de intervalos sucesivos y el método de intervalos aparentemente iguales.

Para responder a si estas diferencias fueron estadísticamente significativas, se calculó el **coeficiente de correlación de rango o rho de Spearman** (Siegel, 1991b), índice con el que se estudia el grado de correspondencia de dos variables con datos que muestran un ordenamiento. La fórmula para casos en los que una categoría es elegida en varias ocasiones es la que a continuación se presenta:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}}$$

donde:

- $\sum x^2 = \frac{N^3 - N}{12} - \sum T_x$ = suma de cuadrado corregida por las repeticiones en cada categoría en el primer conjunto de datos o primera variable.
- $\sum y^2 = \frac{N^3 - N}{12} - \sum T_y$ = suma de cuadrado corregida por las repeticiones en cada categoría en el segundo conjunto de datos o segunda variable.
- N = número de acciones formativas.
- $\sum T_x$ = suma de los factores de corrección de todas las categorías en el primer conjunto de datos o primera variable.

- $\sum T_y$ = suma de los factores de corrección de todas las categorías en el segundo conjunto de datos o segunda variable.
- $T = \frac{t^3 - t}{12}$ = factor de corrección para cada categoría.
- t = número de repeticiones en una categoría concreta.
- $d = x - y$ = diferencias, para una acción formativa, entre el rango obtenido con el primer par de métodos y el obtenido con el segundo par.

Sus valores pueden oscilar entre 0, correspondencia nula a 1, máxima correspondencia.

Los resultados obtenidos con esta prueba llevaron a concluir que la correspondencia entre ambos grupos de resultados fue muy elevada y estadísticamente significativa con un nivel de confianza del 99% [$\rho = 0.983$; $p = 0.000$]. Por tanto, se puede decir que las diferencias que se observaron entre ambos pares de métodos no fueron relevantes.

Se constató de este modo que variables como el número de personas respondientes, el número de acciones formativas a priorizar, el número de categorías que puedan elegirse o el número de valores perdidos, no marcan una diferenciación significativa en la priorización obtenida con los distintos métodos.

A pesar de que las diferencias no resultaron significativas, un simple empate entre dos acciones formativas en un momento dado puede resultar un problema si una de las dos ha de quedar fuera del plan de formación por la limitación de los recursos disponibles, por ejemplo. Por ello, se hizo un estudio para dilucidar qué métodos eran más exactos en el sentido de aportar un número menor de acciones formativas con ordenación repetida.

En el **anexo IX** (pág. lxi) se muestra el número de acciones formativas que compartieron posición con otra. En la primera columna, se distinguen los distintos años y en la segunda, las diferentes áreas de conocimiento; en la tercera columna se contabiliza el número de acciones formativas que quedaron empatadas con otras en la misma posición de priorización cuando se utilizó el método de Dunn-Rankin y el de ordenación de rangos; la cuarta y última columna presenta el número de empates cuando se realizó el método de intervalos sucesivos y de intervalos aparentemente iguales.

En la última fila, se presenta el total de empates encontrados para cada par de métodos. Los resultados mostraron que con los métodos de intervalos sucesivos y de intervalos aparentemente iguales se encontraron un mayor número de empates que con los métodos de Dunn-Rankin y de ordenación por rangos; concretamente, 172 contra 81.

Y para constatar si estas diferencias resultaron ser significativas, se realizó la prueba de rangos señalados y pares igualados de Wilcoxon (Siegel, 1991a). Se trata de

una prueba no paramétrica que responde a cuánto difiere la suma de las diferencias entre pares de datos positivas de las diferencias negativas. Los resultados obtenidos constataron que las diferencias entre los empates de los dos conjuntos de métodos fueron significativas [$z = -2.99$; $p = 0.003$], por lo que puede concluirse que los métodos de Dunn-Rankin y de ordenación por rangos son los que presentaron mayor exactitud en sus resultados, al encontrarse menos empates que con los métodos de intervalos sucesivos y de intervalos aparentemente iguales.

6.4. Discusión.

Tras los resultados obtenidos, se propone como método más parsimonioso el de ordenación por rangos, seguidos por el de Dunn-Rankin, el método de los intervalos aparentemente iguales y, por último, el método de intervalos sucesivos.

Aunque decidir qué método es más sencillo pueda parecer una opinión algo subjetiva, podría decirse que obtener los resultados por el método de ordenación por rangos es lo que menos cálculo requiere, ya que basta con sumar las priorizaciones otorgadas a cada curso y ordenar estos resultados.

Teniendo en cuenta que, junto con el de Dunn-Rankin, este método resultó ser el más exacto (con menos empates), se puede decir que el **método de ordenación por rangos** es el más **parsimonioso** de los cuatro puestos a prueba, aunque es importante recordar que, independientemente del que se use, los resultados obtenidos serán similares (las diferencias mostradas no son estadísticamente significativas).

En el otro extremo, como método **menos recomendable** para obtener una priorización, se encontraría el **método de intervalos sucesivos**, principalmente por ser el único que tiene limitaciones de uso en función de las características de los datos. Así, en los casos extremos, cuando las acciones formativas se encontraban en el primer o en el último intervalo, surgía el problema de que el límite inferior en el primer caso y el superior en el segundo eran inexistentes (los intervalos estaban abiertos por un lado), lo que conllevó que tampoco pudiera calcularse la amplitud del intervalo.

Si hubiera que elegir entre los dos métodos que quedaron intermedios, el de Dunn-Rankin y el de intervalos aparentemente iguales, se recomendaría el primero mencionado porque, aunque es cierto que ninguno de los dos requiere unos cálculos excesivamente complejos, el método de Dunn-Rankin parece ser más discriminante al presentar menor número de casos empatados.

Si bien es cierto que existen otros métodos de escalamiento aparte de los cuatro presentados, una pregunta podría ser por qué se escogieron concretamente éstos y no otros. Los criterios que se tomaron para la elección fueron los siguientes:

- Se buscaron métodos que escalaran **estímulos**. Por tanto, fueron descartados algunos otros métodos donde lo que se escalaban eran las personas que daban las respuestas.
- **Factibilidad**. Se eligieron aquellos métodos cuyo procedimiento se adecuara a las características de los datos que en este caso se iban a recoger. Así, por ejemplo, el método de comparaciones binarias y la ley del juicio comparativo (Thurstone, cit. en

Arce, 1994) no se usaron porque implicaba que las personas fueran eligiendo constantemente entre dos estímulos, lo cual hubiera sido muy cansado al haber tantos cursos, y porque vía internet hubiera supuesto la creación de un programa que no estaba hecho, que fuera presentando los cursos de dos en dos y que fuera analizando en el mismo momento en que se introducía la respuesta.

La **principal mejora** que se propone es que una segunda persona realice la codificación de los cuestionarios y posteriormente los cálculos con los distintos métodos con la finalidad de calcular la fiabilidad interjueces; un índice elevado apoyaría la idea de que los resultados encontrados son correctos.

7. CONCLUSIONES.

La detección de necesidades planteada se caracteriza por ser un **proceso mixto**, donde las tomas de decisiones **políticas** se basan en evidencias obtenidas por vía **técnica** con un grado aceptable de validez. Se trata de este modo de potenciar los siguientes aspectos:

- **Transparencia.** Se intenta que la información esté disponible en todo momento y para todos los interesados para que así se sepa por qué se toman unas decisiones concretas y se descartan otras.
- **Apertura.** Se busca que el canal de recogida de información esté abierto continuamente y que se tenga la capacidad de reaccionar ante imprevistos y de salirse de lo programado cuando esto sea imprescindible sin que esto entorpezca la consecución de los objetivos planteados.
- **Innovación.** Se trata de facilitar la inclusión de acciones formativas novedosas y actuales, animando a los participantes a que innoven y no se limiten a solicitar nuevas ediciones de acciones formativas que ya se realizaron en planes anteriores.
- **Democracia.** Si bien es cierto que las decisiones finales son políticas, es preciso intentar que el proceso de detección de necesidades permita que todos los implicados tengan la posibilidad de participar y que a todos se les consideren iguales.
- **Participación.** Se fomenta la reflexión de los trabajadores y potenciales destinatarios de las acciones formativas en relación con las necesidades a cubrir y las mejoras producidas en el puesto de trabajo. Es muy importante provocar en los participantes la sensación de que son escuchados de manera inmediata y de que tienen la posibilidad de participar directamente en la creación de los planes de formación.
- **Realismo.** Por un lado, se procura que el proceso de detección de necesidades planificado se adapte a la cantidad de recursos disponibles para su realización; por otro lado, es importante que sea un proceso cercano a los trabajadores y adaptado a ellos. Por esta razón, cada vez más se está pasando de preguntar por cursos

concretos a solicitar necesidades a cubrir ya que, quien presenta una necesidad, no tiene por qué conocer la manera de cubrirla.

- **Rigor.** Si se pretenden obtener datos fiables y válidos, es necesario seguir los distintos pasos de la evaluación de manera rigurosa, siguiendo el diseño previsto en la medida de lo posible y tratando de obtener resultados precisos con los que las decisiones a tomar carezcan de ambigüedad.
- **Exhaustividad.** Es preciso recoger una información completa, sin omitir ningún aspecto importante. Por ello, se trata de perfilar no sólo todas las acciones formativas que se quieren incluir, sino también los objetivos y contenidos que han de tener en caso de que los solicitantes conozcan esta información, e incluso los posibles formadores que pudieran encargarse de su diseño e implementación, entre otros aspectos.
- **Utilidad.** La finalidad principal de un plan de formación es la mejora del desempeño en el puesto de trabajo. Para lograrlo, el primer paso es el estudio de las necesidades reales de los trabajadores a partir del cual diseñar acciones formativas de calidad. Esta recogida de información sobre el plan sirve como medida previa con la que comparar la posterior y así poder estudiar el cambio provocado por las acciones formativas. Además, la información obtenida es útil para evaluar la transferencia, puesto que las necesidades de partida revelan los aspectos en los que hay que centrarse en dicha fase de evaluación.

En busca de estos ideales, a lo largo del tiempo se fueron poniendo en marcha ciertas **mejoras**. Destacan los siguientes aspectos:

- **La difusión del proceso y la muestra elegida.** Se pasó de recoger la opinión de los altos cargos como representación a informar y obtener opiniones directamente de cualquier empleado, independientemente de su cargo, sin intermediarios. Además, el número de participantes fue en aumento cada año, por lo que podría decirse que los datos recogidos fueron cada vez más representativos de los implicados.
- **Los instrumentos utilizados:**

El uso de las **nuevas tecnologías**, que no estaban disponibles los primeros años, permitió la aparición de nuevos métodos de contacto, como los foros y chats a través de Internet, lo que permitió una comunicación fluida con respuestas al instante.

Los **cuestionarios** recogieron cada vez una información más precisa, completa y útil, al ir incrementándose el número de ítems que versaban sobre cuestiones relevantes.

La **variedad de instrumentación** utilizada permitió que la información recogida fuera más rica, objetiva y exacta, y que se pudieran completar entre sí los resultados obtenidos a través de distintas vías, además de detectar posibles coherencias y/o discrepancias.

- **El procedimiento.** Las nuevas tecnologías agilizaron el envío y la recogida de la información.
- **El momento de recogida.** Se pasó de una recogida de información puntual, durante un breve periodo de tiempo, a dar la posibilidad de cumplimentar y entregar el cuestionario en cualquier momento.
- **La codificación de la información.** Las nuevas tecnologías permitieron la agilización de esta tarea, pues se pasó de la introducción de los datos manualmente a la recopilación automática en una base de datos.
- **El análisis de datos.** Se pasó de calcular únicamente algunos índices descriptivos a realizar el cálculo de la priorización de las acciones formativas, lo que llevó a obtener información detallada de la importancia otorgada a cada acción. Y para optimizar este proceso, en el apartado 6 de este capítulo se analizó qué método de priorización fue más parsimonioso.

Aunque la evolución de la detección de necesidades a lo largo del tiempo podría en términos generales considerarse positiva, aún quedan pendientes algunas **mejoras para el futuro**:

- **Instrumentos.** Además del cuestionario utilizado en la fase principal, se piensa utilizar como complemento otros instrumentos de recogida de información cualitativa como pueden ser los grupos de discusión o entrevistas, organizadas por zonas. En la post-evaluación, se intentará consolidar la organización de mesas técnicas de comités de expertos en las distintas áreas, para el desarrollo de los contenidos de las acciones formativas solicitadas.
- **Vías de comunicación.** A lo largo del tiempo, el uso de Internet por los participantes ha resultado algo más familiar y mejor aceptado. Se seguirá trabajando en este sentido. Por ejemplo, y para una mayor transparencia, se tratará de facilitar por esta vía el calendario de actividades de la detección de necesidades del año próximo, con la intención de tener informados a todos los implicados. Además, se intentará utilizar el envío de mensajes SMS para hacer llegar esta información a través de los teléfonos móviles de los participantes, método que actualmente está ampliamente extendido y aceptado.
- **Feedback.** Se espera crear una base de datos que permita el cálculo y la visualización inmediatos a través de Internet de los datos recogidos, con lo que los usuarios podrían informarse de manera actualizada de la situación en que se encuentran las acciones formativas que solicitaron.
- **Difusión de los resultados.** Para aumentar la transparencia del proceso, se pretende dejar disponible en la página web de las diferentes organizaciones el informe de detección de necesidades, que recoge de manera pormenorizada los objetivos que se trataron de lograr en cada fase, el método utilizado y los resultados obtenidos.

Pensando en la posibilidad de **generalizar** estas reflexiones es necesario tener en cuenta que, a la hora de llevar a cabo una detección de necesidades, el contexto organizacional y la realidad son complejos, por lo que no se puede actuar de forma

estereotipada, sin partir de la idiosincrasia de la institución sobre la que se va a intervenir. Para **extrapolar** cualquier procedimiento de evaluación, por tanto, hay que tener en cuenta las características propias de la organización. Así pues, no hay un método estándar para la detección de necesidades; la forma de actuar dependerá de la situación contextual en la que se encuentre inmerso el organismo en cuestión (como, por ejemplo, la orientación política imperante) y de sus características intrínsecas (número de empleados, dispersión geográfica, etc.).

Además, las circunstancias no siempre van a permitir llevar a cabo un procedimiento perfecto a nivel metodológico. Pero obtener cualquier dato, aunque sea mejorable, es más útil que no tener información. Es interesante tratar de mejorar poco a poco, pero sin ofuscarse porque la realidad organizacional impida que sea factible hacer todo lo que como evaluadores se desea.

CAPÍTULO 6. EVALUACIÓN DURANTE LAS ACCIONES FORMATIVAS: DISEÑO E IMPLEMENTACIÓN

1. Introducción	261
2. Método	262
2.1. Muestra	262
2.2. Instrumentos	262
2.3. Procedimiento.....	263
3. Resultados.....	264
4. Discusión.....	265
5. Conclusiones	266

CAPÍTULO 6. EVALUACIÓN DURANTE LAS ACCIONES FORMATIVAS: DISEÑO E IMPLEMENTACIÓN.

1. INTRODUCCIÓN.

Generalmente, esta fase de evaluación no es desarrollada en muchas ocasiones, puesto que es la evaluación sumativa la que impera, donde los resultados finales son lo único a tener en consideración.

Sin embargo parece lógico pensar que, si se pone en marcha un proceso de recogida de datos a lo largo de la intervención que permita detectar posibles problemas que hayan surgido y si, además, se tiene previsto una posible línea de actuación ante dichos problemas y se pone en marcha este plan alternativo, los resultados serán más óptimos que si simplemente se espera a evaluar al final de la intervención.

Tomando como marco de referencia por tanto una **evaluación formativa**, los **objetivos** de la evaluación “**durante**” en este caso concreto son los siguientes (Chacón, López y Sanduvete, 2004):

- Proporcionar al alumnado, al formador y a los responsables de la formación, información sobre el desarrollo del proceso de aprendizaje.
- Corregir durante el desarrollo de la acción formativa las deficiencias de los recursos y de la programación efectuada y, así, tratar de mejorar o remodelar sobre la marcha aquellos aspectos de los que depende el buen transcurso de la misma.
- Localizar los obstáculos principales para la implementación de la acción formativa, permitiendo así una retroalimentación continua al alumnado que facilite el aprendizaje y estimule su interés, ayudándole a superar el proceso.

Cabe señalar que no todas las intervenciones van a ser evaluables en este sentido, puesto que un requisito indispensable es que haya un diseño previo, un plan de partida que sirva de referente. Las discrepancias entre este modelo y lo que en realidad se está dando al implementarse la acción formativa son los aspectos sobre los que intervenir.

Y el siguiente requisito para que se pueda dar esta evaluación durante el proceso es la previsión de alternativas para cuando surjan los imprevistos. Un buen diseño de evaluación es flexible, puede adaptarse a las nuevas circunstancias acaecidas en un momento dado, pero deja poco a la improvisación; se toman planes ya diseñados que se mantienen en la recámara mientras no son necesarios.

Esto último marca la diferencia entre “seguimiento” y “monitorización”. Aunque en ambos casos se toman datos a lo largo del tiempo, en la monitorización la tarea acaba con esta recogida, mientras que en el seguimiento lo más importante estaría aún por hacer: la implementación de una alternativa para solventar los problemas detectados con la toma de datos.

En suma, una vez planificado el plan de formación, se hace necesario establecer mecanismos de seguimiento de dicho plan con objeto de tener, en la medida de lo posible, certeza de que los resultados alcanzados se deben a las acciones formativas previamente programadas y eficazmente implementadas. El establecimiento de un mecanismo de seguimiento permite remodelar el plan previamente definido; detectar posibles carencias o aspectos negativos; y actuar mejorándolos en el acto, sin necesidad de esperar al final de la acción formativa; e incluso detectar necesidades no identificadas en un principio.

A continuación, se presenta la evolución de esta fase de evaluación en las organizaciones en estudio, teniendo en cuenta que en Diputación de Sevilla falta su consolidación, en el IAD no llegó a implementarse y en el Centro de Formación del Personal de Administración y Servicios de la Universidad de Sevilla se planificó esta evaluación, quedando aún por implantar.

2. MÉTODO.

2.1. Muestra.

La población potencial fueron todos los asistentes a las acciones formativas presenciales de los planes de formación desde el año 2004, todos los profesores de esta modalidad y todos los coordinadores cuando los había.

2.2. Instrumentos.

Dos son los instrumentos utilizados en esta fase; el primero que se presenta, dirigido a los participantes de las acciones formativas; el segundo, para que sea cumplimentado por el profesorado y los coordinadores, con mismo formato pero contenido diferente para cada figura (Chacón, López y Sanduvete, 2004):

El **cuestionario durante** presenta seis ítems de formato abierto que hace referencia a los objetivos, contenidos, el ambiente, los formadores y los materiales y/o recursos, además de que se deja un apartado abierto por si hubiera alguna otra sugerencia (ver anexo X, pág. lxv). Con este contenido, se analiza si la metodología se adecua al desarrollo de los objetivos, si se está siguiendo la planificación prevista (espacios, tiempo, apoyos y recursos) y los posibles niveles parciales de logro y variables relacionadas con el ambiente de la acción formativa (interés de las actividades, flexibilidad en la aplicación del plan, relaciones con el alumnado o conflictos con el profesorado, por ejemplo).

Aunque este instrumento no presenta ningún estudio de validez, muestra un elevado grado de coherencia dentro del proceso, puesto que toca los mismos aspectos que el cuestionario que servirá para evaluar la satisfacción de los asistentes con la acción formativa.

Los **listados de verificación** o **checklists** (anexo XI, pág. lxix) son un conjunto de ítems de respuesta dicotómica (sí/no) con los que se presentan las diferentes tareas que tanto profesorado como coordinación habrían de ir haciendo a lo largo del proceso.

Concretamente, se trata de un listado para coordinadores y otro diferente para profesores, divididos cada uno de ellos en tres momentos: antes del inicio de la acción formativa, durante ésta y una vez finalizada.

Estas tareas son propuestas por expertos en formación continua. Así, los listados de verificación que se muestran en el anexo XI (pág. lxix) fueron creados por la Sección de Formación Continua y contienen las tareas que según los técnicos de dicha Sección son imprescindibles para el buen funcionamiento del plan de formación.

Por poner un ejemplo, y para hacernos una idea de los contenidos que se presentan con este instrumento se puede apreciar que, concretamente, entre las tareas a realizar por los coordinadores antes de que empiece la acción formativa están entregar la propuesta de programa a la coordinación con una antelación no inferior a 15 días, comunicarles los recursos didácticos necesarios para la ejecución correcta de la acción y preparar el material de apoyo a entregar al alumnado; y entre las tareas a llevar a cabo por la coordinación después de que finalice la acción formativa se encuentran entregar a la organización del plan de formación tanto las hojas de firmas del profesorado y el alumnado como el cuestionario de evaluación de la satisfacción (presentado en el anexo XII, pág. lxxiii, y descrito en el próximo capítulo).

Una descripción más pormenorizada de las tareas a realizar por formadores y coordinadores puede encontrarse en el “Manual de Formación”, elaborado por la Sección de Formación Continua (2007), disponible en <http://www.dipusevilla.es/formacioncontinua>, en el apartado de mismo nombre. Instrumento similar se elaboró en el Centro de Formación del PAS (Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 2001).

La utilidad de este instrumento no se limita exclusivamente a valorar el grado de implicación y de competencia de coordinadores y formadores sino que, además, sirve para informar a ambas figuras de manera clara y concreta acerca de las tareas que han de realizar.

2.3. Procedimiento.

En las organizaciones en las que había coordinador como en Diputación de Sevilla, el **cuestionario durante** era entregado a éste; donde no existía esta figura, se entregaba al profesorado de la acción formativa.

Este instrumento quedaba disponible a lo largo de todas las sesiones formativas para que pudiera ser cumplimentado anónimamente por cualquier participante que estuviera interesado en mostrar su opinión respecto a las dimensiones que recoge para mejorar el desarrollo del curso, dándoles la posibilidad de entregarlos en un sobre cerrado que se les proporcionaba.

Al final de cada sesión, era el mismo coordinador (o profesor en su defecto) quien se encargaba de recoger los cuestionarios y de dar respuesta inmediata a las solicitudes, problemas o necesidades que se planteaban, esto sí, con ayuda del profesorado y demás profesionales implicados en la organización de la acción formativa.

Los **listados de verificación** o **checklists** también eran entregados, junto con la documentación, a la coordinación (o al profesorado en su defecto).

Tanto coordinadores como profesores, cumplimentaban dichos listados en tres momentos: antes, durante y después de la acción formativa. Al tratarse de tareas concretas, la respuesta a este instrumento era tan sencilla como marcar con una X el “sí”, una vez realizada la tarea que se solicitaba.

Una vez llevada a cabo su cumplimentación, estos listados eran entregados a los evaluadores, quienes realizaban un análisis en conjunto de estos datos con los recogidos a lo largo de todo el proceso, con la intención de obtener una visión holística de lo sucedido en la acción formativa y relacionarlo con los resultados hallados al final de ésta.

En los casos en que fue necesario, y en busca de mejorar para futuras acciones formativas, toda la información anterior se contrastó en reuniones entre formadores, coordinadores, evaluadores y técnicos organizadores del plan de formación.

3. RESULTADOS.

Una vez que se ha planificado la acción formativa, se hace necesario establecer mecanismos de seguimiento con objeto de tener, en la mayor medida de lo posible, certeza de que los resultados alcanzados se deben a la acción formativa previamente programada y efectivamente desarrollada.

Los **cuestionarios durante** posibilitan al alumnado dar cualquier sugerencia que permita mejorar el desarrollo de la acción formativa, lo que da la posibilidad de responder inmediatamente a los posibles problemas y necesidades detectados en el transcurso de dichas acciones.

Este mecanismo permite remodelar el programa previamente definido y, a su vez, detectar necesidades no identificadas en un principio o incluso beneficios, directos o indirectos, no planificados en el proceso.

Por otro lado, los datos obtenidos con el **listado de verificación** o **checklist** posibilitan, en el mismo momento de su cumplimentación, que coordinadores y profesores sean conscientes de las tareas que les corresponden hacer antes, durante y después de la realización de la acción formativa por el bien de la organización e implementación de ésta; y al final, poniendo los datos en relación con los resultados del informe global de evaluación, permite que se analicen las tareas que han podido llevarse a cabo o no, y su posible relación con el desarrollo y resultados de la acción formativa.

4. DISCUSIÓN.

Aunque en este capítulo no se han aportado datos concretos obtenidos en algunas de las organizaciones en estudio, se ha tratado de dar una visión pragmática de la evaluación realizada en el momento en que la acción formativa está en marcha, presentando unos instrumentos concretos cuyo uso podría generalizarse a otros contextos organizacionales.

Con esta actuación, se llevan a cabo tres pasos con los que se posibilita la valoración de la coherencia interna de cada acción formativa:

1. Se hace un estudio más pormenorizado y profundo del diseño presentado; concretamente, de aspectos tales como los medios humanos y materiales necesarios, las acciones a desempeñar, los usuarios a los que va dirigida la acción formativa, la calendarización, los momentos de medida y otras particularidades de la intervención.
2. Se recogen datos del proceso en referencia a los aspectos previamente diseñados.
3. Comparando lo diseñado con los resultados encontrados durante la ejecución de la acción formativa, se detectan las discrepancias entre lo planificado y lo implementado, lo que puede traducirse a propuestas de mejora en el mismo instante y/o para futuras acciones formativas.

Las intenciones concretas que se persiguen con el uso de los instrumentos previamente descritos son las siguientes:

Con el **cuestionario durante**, se trata de “invitar” al alumnado a depositar cualquier sugerencia que permita mejorar el desarrollo de la acción formativa. En acciones formativas de un día no tiene mucho sentido su uso, pues tan breve duración puede imposibilitar el más mínimo margen de reacción para mejorar en algún aspecto concreto; sin embargo, en otras de mayor duración puede ser un instrumento eficaz para que los formadores tengan la posibilidad de mejorar la acción formativa durante el transcurso de la misma.

Con los **listados de verificación** o **checklists** no se pretende “controlar”, sino concretar y difundir las tareas y principales funciones a desarrollar antes, durante y después de la implementación de la acción formativa por parte del profesorado y la coordinación en su caso, además de detectar la repercusión que esto tiene en el resto de implicados en la acción formativa. Se pretende con ello disponer de un seguimiento de forma continua de la realización de las funciones que les corresponde y fomentar la mejora continua desde la óptica de la organización, así como la correcta implementación de la acción formativa. Hay veces que no se realizan o se realizan otras tareas simplemente porque no se conoce lo que se supone debería saberse, o porque no están explícitas en ningún documento concreto.

Por el hecho de que las evaluaciones suelen ser de carácter sumativo en lugar de evaluarse el proceso, al ser ésta una manera relativamente poco común de recogida de

datos y posterior valoración, los diversos implicados en las acciones formativas pueden dudar de la utilidad de estas listas; pueden ver en ellas simplemente “otro papel más a rellenar que no sirve de nada”. Para contrarrestar esta posible visión del instrumento, a continuación se sintetizan algunas de sus ventajas:

- El mero hecho del diseño de este instrumento puede suponer en sí un avance para aclarar posibles dudas en los procesos por parte de todos los implicados. Es una manera de aclarar las tareas de cada uno, para que no existan solapamientos o, por el contrario, incumplimiento de algún deber por no tener conocimiento de a quién correspondía su realización.
- Disponer de estas listas supone aclarar a todos los implicados, en un solo registro que además es de muy rápida cumplimentación, cómo se desarrollan las diferentes actividades y en qué secuencia.
- La utilidad de este instrumento se potencia al tenerse en cuenta que no se trata de registros aislados, sino que van a analizarse junto con toda la información recogida y relacionada, directa e indirectamente, con una acción formativa concreta.

Se dice que, al evaluar, a la vez se está interviniendo. Un posible ejemplo de cómo la utilización y publicidad de registros de tareas o funciones que se dan por hecho tiene repercusiones es el siguiente: el mero acto de que un profesor tenga que firmar en cada clase que imparte un simple registro de las posibles incidencias que han surgido en dicha clase, repercute en una clara mejora a la hora de detectar posibles problemas de organización docente.

5. CONCLUSIONES.

Cuando se habla de evaluación durante la intervención, en muchas ocasiones se trata de una **monitorización** o mera recogida de datos en lugar de un verdadero seguimiento donde, además de obtener información, se compara con el diseño previsto y se proponen mejoras inmediatas previamente planificadas, pues se trata de predecir posibles problemas que puedan surgir y proponer algunas soluciones a éstos.

Con la evaluación presentada en este capítulo, no sólo se da la posibilidad de actuar en busca de la mejora de futuras acciones formativas sino que, al detectarse las carencias “in situ”, se permite la acción inmediata para tratar de solventar las cuestiones detectadas.

Con la propuesta de instrumentos concretos, se avanza en uno de los puntos débiles encontrados tras el estudio de la bibliografía referente a la evaluación de la formación continua: la carencia de **operativización**. Aunque puede que estos instrumentos no se adapten a las necesidades de evaluación de todas las organizaciones, la descripción del proceso de elaboración y de cómo se utilizan los resultados obtenidos, da la posibilidad de generalizar a otras situaciones.

A pesar de la utilidad manifiesta de este proceso de evaluación, se considera que aún existen algunos **aspectos que mejorar**, como son los que a continuación se presentan:

El hecho de que no sea un proceso de evaluación muy común, está impidiendo que el uso de estos instrumentos se **consolide** y se **ponga en marcha** en las tres organizaciones en estudio. La situación actual es que ya no se realiza la evaluación de la formación realizada en el IAD, que en el Centro de Formación y Perfeccionamiento del PAS se trató de implementar por primera vez en el plan de formación del año 2007 y que en Diputación de Sevilla se generó en el año 2006 una situación de retroceso en esta evaluación.

Concretamente en Diputación de Sevilla, se implementó tal como se ha descrito durante los años 2004 y 2005; sin embargo, en el año 2006 dejó de utilizarse el cuestionario durante y los listados de verificación o checklists se siguieron utilizando, pero en formato listado y no como instrumento a cumplimentar y luego entregar.

La primera modificación (la anulación del uso del cuestionario durante) se debió a que, entre otras razones, se le estuvo dando un uso no apropiado a este instrumento, ya que se entregaba uno por participante y muchos quedaban en blanco, con el gasto de papel y de esfuerzo por parte de quien se ocupaba de la preparación de la documentación que esto implicaba, cuando realmente la propuesta era dejar un número menor de copias encima de la mesa o en algún otro lugar donde el alumnado tuviera acceso para que, sin necesidad de solicitar permiso, cualquiera que tuviera algo que decir pudiera dejar constancia por escrito. Esta función se ha suplido actualmente por el intercambio verbal de opiniones, modo de proceder que conlleva algunos puntos flacos como que no se recoge en ninguna parte, por lo que existe más posibilidad de que la petición sea ignorada.

La segunda modificación (el cambio de formato de lo que antes era un checklist) se debió a que los técnicos de formación de Diputación consideraron que la misma información se daba a formadores y coordinadores dando los ítems para que se marcara con una cruz en positivo o negativo, que dándolos formando un listado al que no habría que responder. Desde la unidad de evaluación se consideró que, si bien es cierto este planteamiento, el cambio de formato conllevaba una pérdida de información (qué tareas habían llevado a cabo los formadores y los coordinadores en su caso y qué les había quedado por hacer), con lo que se eliminó la posibilidad de poner en relación estos datos con los obtenidos en la evaluación final. Se intentará por tanto justificar este punto de vista y tratar de reimplantar los listados de verificación en su formato originario.

Hasta ahora, no se ha llevado a cabo un **estudio de validez** de ninguno de los dos instrumentos. Su coherencia dentro del proceso de evaluación y el contexto organizativo está justificada debido a que, como ya se ha comentado, por una parte el cuestionario durante contiene prácticamente los mismos apartados que el instrumento destinado a la evaluación de la satisfacción, lo que facilita la comparación del dato tomado durante la acción formativa y el recogido tras finalizar ésta; y, por otra parte, el listado de verificación está creado a partir del Manual de Formación elaborado por los técnicos de la Sección de Formación Continua.

Sin embargo, una mejora que se tratará de llevar a cabo en el futuro es la realización de un estudio de validez de contenido de estos instrumentos para averiguar si realmente todos los elementos que los componen son trascendentales para la mejora de la formación y si faltaría algún aspecto a considerar para tal fin.

Se ha detectado un considerable punto flaco en referencia a la utilización de estos instrumentos, y es que sólo se aplican a las acciones formativas de modalidad presencial, quedando las realizadas a través de **teleformación** sin posibilidad de una evaluación del proceso.

Para solventar esta carencia concretamente se propone, en primer lugar, crear una versión adaptada a esta modalidad; y, en segundo lugar, ponerla disponible en la plataforma de teleformación utilizada para la realización de los cursos, de tal manera que se pudiera responder en cualquier momento, y coordinadores y profesores tuvieran acceso a lo solicitado de inmediato, con lo que se posibilita la opción de actuar.

Hasta ahora sólo se intentó la aplicación de dos instrumentos, el listado de verificación y el cuestionario durante, para evaluar el proceso. Sin embargo, éstos no son los únicos **instrumentos** aptos para evaluar el proceso. Otras posibilidades son las siguientes (Anguera, Chacón, Holgado y Pérez, en prensa):

- Entrevistas: la información podría recogerse, en lugar de por escrito, de manera oral (aunque posteriormente se transcribieran), con lo que la inmediatez de la respuesta y la posibilidad de margen de movimiento están prácticamente aseguradas.
- Cuestionarios a todos los implicados: el cuestionario durante está pensado principalmente para que sea cumplimentado por los participantes a las acciones formativas. Sin embargo, otros directamente implicados en el proceso de formación como son el profesorado y la coordinación en su caso, también tienen mucho que aportar en relación a mejorar la formación. Por este motivo, una posible propuesta sería hacer una versión de este cuestionario pensando en los aspectos que podrían ser evaluados por estas dos figuras.
- Registros observacionales: siguiendo con la misma línea, con la implicación en la mejora de la formación por parte de coordinación y profesorado, otra posibilidad sería elaborar una hoja de registro de observación donde éstos pudieran anotar incidencias acaecidas como, por ejemplo, aspectos relacionados con una posible desmotivación del alumnado o un clima inapropiado, dándose la posibilidad así de actuar al respecto.

Hasta aquí llegó la descripción de la evaluación del proceso. A partir del siguiente capítulo, se presenta la evaluación final o de resultados.

CAPÍTULO 7. EVALUACIÓN DESPUÉS DE LAS ACCIONES FORMATIVAS (I). SATISFACCIÓN

1. Introducción	271
2. Método	272
2.1. Muestra	272
2.2. Instrumentos	273
2.3. Procedimiento.....	275
3. Resultados.....	276
4. Discusión.....	282
5. Conclusiones	285

CAPÍTULO 7. EVALUACIÓN DESPUÉS DE LAS ACCIONES FORMATIVAS (I). SATISFACCIÓN.

1. INTRODUCCIÓN.

En este capítulo y en el siguiente, se tratan distintas fases de la evaluación de las acciones formativas que se realizan una vez finalizadas éstas. Concretamente en este capítulo 7, se habla de la evaluación de la satisfacción de los participantes en las acciones formativas. Tal como se comentó previamente, según Kirkpatrick (1999), se trata de la fase de evaluación más superficial y más sencilla de poner en marcha, por lo que es de las más extendidas entre las organizaciones.

La evaluación de la satisfacción consiste en recabar información sobre las actitudes, impresiones y percepciones de los participantes en una acción formativa, estableciendo en qué medida les ha gustado y cómo consideran que la formación puede resultarles útil en el desempeño de su trabajo (Chacón, López y Sanduvete, 2004). Esta información permite conocer, además del nivel de satisfacción de los participantes en la acción formativa, los aspectos en los que será necesario incidir para mejorar acciones formativas similares que se emprendan en el futuro.

El **objetivo general** de esta fase de evaluación es mejorar la calidad de la formación. Este objetivo puede concretarse en otros **objetivos específicos**:

- Comprobar que la acción formativa se desarrolló con el diseño establecido al inicio del proceso. Para ello, se valoran cuestiones relacionadas con los objetivos y contenidos, la metodología, la utilidad para la formación profesional y el desarrollo personal, la valoración global, el papel del coordinador y del profesorado o los tutores.
- Conocer en qué medida los participantes consideraron que la acción formativa realizada respondió a sus necesidades y expectativas.
- Proporcionar al formador información sobre los puntos fuertes y débiles de la acción formativa en cuanto a metodología, actividades programadas y recursos utilizados, con objeto de que pueda utilizarla para mejorar las acciones de formación futuras.

En este capítulo, concretamente se presenta una **descripción del proceso** llevado a cabo y de los resultados disponibles en relación a la evaluación de la formación ejecutada en Diputación de Sevilla, en el Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla (en ambas organizaciones desde el año 2000 al 2006) y en el Instituto Andaluz del Deporte (en los años 2005 y 2006), además de presentarse posibles propuestas de mejora para el futuro (Chacón, Lara y Pérez, 2001, 2002b; Chacón, López y Pérez, 2003; Chacón, Sanduvete y Holgado, 2004, 2005c; Chacón, Gómez, Sánchez y Sanduvete, 2006; Chacón, Sanduvete, Holgado y Sánchez, 2006; Unidad de Evaluación del Centro de Formación y Perfeccionamiento del PAS, 2000, 2006b; Chacón, Sánchez, Holgado y Sanduvete, 2007; Chacón, Pérez, Sanduvete, Holgado, Sánchez y Lozano, 2007).

2. MÉTODO.

2.1. Muestra.

A continuación, en la tabla 7.1, se presenta un cuadro resumen de la evolución acaecida en la muestra en las distintas organizaciones. La primera columna presenta la organización a la que se hace referencia; la segunda, el año en el que se llevó a cabo la evaluación; la tercera columna presenta el número de acciones formativas que se realizó en dicho año; la cuarta columna, el número de personas que finalizaron las acciones formativas; la penúltima columna muestra el número de encuestas que fueron entregadas; y la última columna, el porcentaje que este valor supone del total de personas que finalizaron las acciones formativas.

ORGANIZACIÓN	AÑO	Nº EDICIONES	POBLACIÓN POTENCIAL	Nº ENCUESTAS	% RESPUESTAS
DIPUTACIÓN DE SEVILLA	2000	---	---	---	72.63%
	2001	---	2113	1635	77.38%
	2002	---	1699	1668	98.17%
	2003	89	2089	1775	84.97%
	2004	173	2804	1956	69.76%
	2005	195	3093	2338	75.59%
	2006	163	3139	2051	65.34%
CENTRO DE FORMACIÓN DEL PERSONAL DE ADMINISTRACIÓN Y SERVICIOS	2000	114	1512	215	14.22%
	2001	131	1839	938	51.01%
	2002	154	1664	1274	76.56%
	2003	175	2450	1925	78.57%
	2004	163	3099	1683	54.31%
	2005	149	2494	1244	49.88%
	2006	164	2999	1564	52.15%
IAD	2005	---	2264	892	39.4%
	2006	---	3283	1674	50.99%

Tabla 7.1. Evolución de la participación en la evaluación de la satisfacción.

Algunas casillas se presentan en blanco debido a que no se tuvo acceso a la información. Con los datos que pudieron recogerse es posible destacar, como característica significativa, que en todos los índices se presentó una tendencia ascendente o al menos de mantenimiento, caracterizada por su marcada irregularidad a lo largo del tiempo.

En la figura 7.1 se presenta la evolución en el **porcentaje** de participación de las tres organizaciones a lo largo de los años.

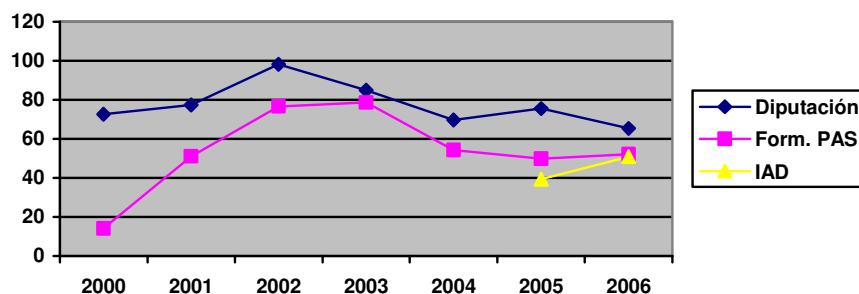


Figura 7.1. Evolución del porcentaje de participación en la evaluación de la satisfacción con las acciones formativas.

Algunos aspectos en este sentido que tendrían que ser puntualizados son los siguientes:

Los datos presentados para **Diputación** en el 2003 y el 2004 sólo hacen referencia a las acciones formativas de modalidad presencial, razón por la que puede parecer que existe un salto entre el número de participantes estos años y los del 2005 y 2006.

Como puede verse, en esta organización el porcentaje de personas que, habiendo finalizado la acción formativa, aportó su opinión respecto a su satisfacción con ésta no mantiene un patrón regular, si bien es cierto que no podría decirse que empeorara en gran medida a lo largo del tiempo.

El bajo número de participantes presentado en el **IAD** en el año 2005 se debe a que no toda la evaluación se realizó con el cuestionario que se muestra en el anexo XII (pág. lxxiii) debido a que el equipo de formación comenzó la colaboración con el IAD una vez iniciado el plan. Las personas que respondieron al cuestionario de satisfacción existente previamente no fueron contabilizadas. Esta circunstancia es una de las posibles razones por las que respondieron más personas en el 2006 que en el 2005.

En el **Centro de Formación del Personal de Administración y Servicios**, el porcentaje de participación presenta una tendencia claramente ascendente hasta el año 2004, momento en que se presenta un declive y a partir del cual el porcentaje de participación se estabiliza.

2.2. Instrumentos.

El principal instrumento que se utilizó en esta fase de evaluación fue un **cuestionario de satisfacción** (ver anexo XII, pág. lxxiii), que los participantes de las acciones formativas cumplimentaban.

Los puntos que se evalúan con este instrumento son la acción formativa, el profesorado y la coordinación, aunque ésta última figura sólo es evaluada en Diputación cuando la acción formativa es realizada en modalidad presencial, ya que en el Centro de Formación del PAS, en el IAD y en modalidad de teleformación en Diputación se prescinde de ella.

Concretamente, este instrumento de evaluación consta de 12 ítems que deben responderse con una escala de valoración de 5 puntos, siendo “1” la mínima puntuación (muy en desacuerdo) y “5” la máxima puntuación (muy de acuerdo). Estos ítems se agrupan en función de las variables a evaluar: los tres primeros ítems tratan sobre los objetivos y contenidos del curso; los seis siguientes se refieren a la metodología y el ambiente; y los tres últimos, a la utilidad y valoración global de la acción formativa. A continuación, aparecen dos preguntas en formato abierto en las que se recoge información acerca de aspectos a mejorar de la acción formativa en cuestión y propuestas de acciones formativas para mejorar el desarrollo del trabajo de los participantes, con la posibilidad de concretar nivel, modalidad, destinatarios, utilidad, objetivos y contenidos de la acción formativa propuesta. Por último, el cuestionario se completa con cuatro ítems que evalúan la actuación del coordinador y seis ítems que evalúan la actuación de cada ponente que también se responden con la misma escala de valoración de 5 puntos. Además, aparece un apartado en formato abierto en el que se recogen otros comentarios acerca de la coordinación y el profesorado.

Un estudio de la calidad métrica y dimensionalidad de los ítems que hacen referencia concretamente a la acción formativa mostró índices apropiados (Holgado, Chacón, Losada, Pérez y López, 2003). Concretamente, presentó:

- Buena fiabilidad y un alto nivel de precisión de medida.
- Un buen índice de discriminación, lo que significa que el instrumento diferencia con precisión aquellos sujetos que están realmente satisfechos con la acción formativa de los que no lo están.

Además, tras un estudio realizado con correlaciones policóricas, se obtuvo evidencia empírica de la validez del modelo de medida utilizado; se constató el ajuste del modelo teórico que sirvió de base para estructurar el contenido de este instrumento con los datos empíricos obtenidos (Holgado, Chacón, Barbero y Sanduvete, 2006).

Este cuestionario descrito es la última versión existente pero no la única; lleva utilizándose desde el año 2005. La evolución por la que pasó este instrumento hasta llegar a la versión actual, con la que poco a poco fue ganando en concreción y especificación, fue la siguiente:

- Hasta el 2002 el cuestionario, mucho más breve que el actual, constaba de 6 ítems de cinco puntos de valoración referidos a la acción formativa, concretamente a la duración, los materiales, la documentación, la coordinación, la utilidad del curso y la valoración global; y de 3 ítems más de valoración del profesorado, concretamente la claridad de exposición, la capacidad de respuesta y la evaluación global. A partir del 2003, se ampliaron los aspectos a valorar ante la necesidad de ahondar más profundamente en otras cuestiones relacionadas con el curso (que pasó de 6 a 13 ítems) y el profesorado (que se amplió de 3 a 6).
- En el 2003, se utilizaron dos cuestionarios algo diferentes en función de la modalidad de impartición (presencial o a distancia). En el año 2004, se logró aunar la información en un solo instrumento aplicable a cualquiera de las modalidades; la única diferencia radicaba en que, al no haber coordinador a distancia, se eliminó en estos casos la evaluación de esta figura. Además, el

coordinador era evaluado con un solo ítem general, mientras que en el 2004 se pasó a una evaluación más pormenorizada a través de 4 ítems.

- En el año 2004, el ítem referido a las necesidades de formación existentes no pedía más que el título de la acción formativa; para el 2005, se concretó la información necesaria, ampliando a nombre, nivel, modalidad, destinatarios, utilidad, objetivos y contenidos.

2.3. Procedimiento.

En la modalidad presencial, el cuestionario es entregado a los asistentes de las acciones formativas al finalizar las mismas. Se cumplimenta de manera individual, voluntaria y anónima. Los encargados de su distribución, recogida y custodia hasta el momento de la entrega en la Sección de Formación son los coordinadores en Diputación y los mismos profesores en el Centro de Formación del PAS y en el IAD.

En la modalidad por teleformación, el procedimiento varía en función de la organización de la que se trate:

- En Diputación de Sevilla, los cuestionarios son expuestos en la plataforma que utiliza el alumnado para la realización de la acción formativa varios días antes de la finalización de ésta. Se envía a todos los participantes un correo electrónico solicitándoles su participación a través de la cumplimentación y envío del instrumento, insistiendo en que es fundamental para la mejora de futuras acciones formativas. Finalmente, el personal de INPRO (Informática Provincial) recibe las respuestas y las envía en un archivo a la Sección de Formación Continua, donde se imprimen.
- En el IAD, los participantes entregan las respuestas en la última sesión presencial y es el profesorado quien se encarga de hacerlos llegar a la Unidad de Evaluación.
- En el Centro de Formación del PAS, donde el proceso está bastante más automatizado, los participantes responden vía web y la información pasa a una base de datos que directamente realiza el tratamiento de la información y genera un informe con los estadísticos descriptivos y las figuras pertinentes.

En Diputación y el IAD donde la información obtenida con los cuestionarios es tratada de manera no automática, los datos son introducidos manualmente y analizados con el programa estadístico SPSS o con Microsoft Excel. A partir de este proceso, se elabora un breve informe de evaluación por cada acción formativa, además de una memoria anual con los datos en conjunto. En ambos casos, se trata de poner en relación los datos cuantitativos con los cualitativos que el mismo cuestionario aporta, con la finalidad de tratar de vislumbrar las posibles razones que llevaron a cada persona a estar o no satisfecha con los aspectos estudiados.

En las tres organizaciones, los resultados pueden ser consultados en las páginas web correspondientes. Así, en <http://www.diputaciondesevilla.es/formacioncontinua>, en el apartado “Evaluación acciones formativas” se presentan los resultados obtenidos en cada una de las ediciones impartidas en los planes de formación de Diputación de Sevilla desde el 2002; en <http://www.juntadeandalucia.es/turismocomercioydeporte>, el IAD presenta los informes globales, tanto de satisfacción como de las demás fases de

evaluación; por último, en <http://www.forpas.us.es/cursos/>, el Centro de Formación del PAS presenta también los resultados de evaluación diferenciados por ediciones.

La difusión de esta información a través de Internet permitió que, de manera fácil y rápida, los interesados pudieran consultar los resultados obtenidos en las acciones formativas.

3. RESULTADOS.

Como aproximación descriptiva, se presentan en las tablas 7.2, 7.3 y 7.4 las medias que se obtuvieron en cada uno de los ítems en las diversas organizaciones en estudio en referencia al curso en general, a la coordinación y al profesorado respectivamente.

Las tres tablas mencionadas siguen el mismo patrón de presentación: la primera columna presenta el ítem completo, mientras que en las columnas posteriores se posicionan las distintas organizaciones en los diferentes años en estudio. El primer dato que aparece en cada casilla hace referencia a la media obtenida; el dato que se presenta abajo entre paréntesis es el número de respuestas obtenidas (N).

En la tabla 7.2, los datos de Diputación de Sevilla sólo se aportan desde el 2003, momento en el que se comenzó a utilizar para la recogida de información el cuestionario actual con algunas modificaciones.

En la tabla 7.3, sólo aparece la evaluación de la coordinación en Diputación de Sevilla porque en las otras dos organizaciones esta figura no se requiere. Se dispone de estos datos desde el 2004, momento en que se pasó de evaluar a la coordinación con un único ítem general a evaluarla con los cuatro presentados en dicha tabla.

En la tabla 7.4 referida a la evaluación del profesorado, los valores relacionados con el Centro de Formación y Perfeccionamiento del PAS se presentan desde el año 2000 al 2005 en su conjunto y el 2006 aparte, por ser el modo en que se encuentran los datos disponibles.

Capítulo 7. Evaluación después de las acciones formativas (I). Satisfacción

Objetivos y contenidos	PAS 2000	PAS 2001	PAS 2002	PAS 2003	PAS 2004	PAS 2005	PAS 2006	DIP. 2003	DIP. 2004	DIP. 2005	DIP. 2006	IAD 2005	IAD 2006
1. A su juicio se han alcanzado los objetivos previstos	3.43 (215)	3.99 (938)	3.88 (1274)	3.99 (1925)	4.05 (1683)	3.97 (1244)	3.92 (1564)	4.09 (1384)	3.99 (1603)	4.03 (2314)	4.03 (2038)	3.97 (676)	3.99 (1471)
2. Los temas se han tratado con la profundidad que permitía la duración del curso	3.5 (215)	4.01 (938)	3.87 (1273)	3.97 (1925)	4.01 (1683)	3.92 (1244)	3.9 (1564)	4.02 (1386)	3.94 (1600)	3.97 (2314)	3.98 (2033)	3.95 (681)	4 (1472)
3. La duración del curso ha sido adecuada a los objetivos y contenidos	2.83 (213)	3.28 (938)	3.2 (1273)	3.37 (1925)	3.44 (1683)	3.34 (1244)	3.46 (1564)	3.57 (1377)	3.48 (1587)	3.61 (2307)	3.72 (2019)	3.62 (674)	3.48 (1453)
Metodología y ambiente	PAS 2000	PAS 2001	PAS 2002	PAS 2003	PAS 2004	PAS 2005	PAS 2006	DIP. 2003	DIP. 2004	DIP. 2005	DIP. 2006	IAD 2005	IAD 2006
4. La metodología se ha adecuado a los objetivos y contenidos	3.52 (212)	4.18 (938)	4.01 (1274)	4.03 (1925)	4.1 (1683)	4.04 (1244)	3.99 (1564)	4.18 (1381)	4.07 (1601)	4.08 (2306)	4.11 (2027)	3.94 (677)	3.93 (1465)
5. La metodología ha permitido la participación activa de los asistentes	3.66 (208)	4.35 (938)	4.2 (1271)	4.05 (1925)	4.19 (1683)	4.15 (1244)	4.11 (1564)	4.43 (1373)	4.3 (1599)	4.28 (2315)	4.33 (2039)	4.1 (672)	4.07 (1455)
6. La acción formativa ha permitido compartir experiencias profesionales con otros compañeros	3.55 (208)	4.01 (938)	3.83 (1273)	3.72 (1924)	3.78 (1683)	3.77 (1244)	3.73 (1564)	4.14 (1376)	4.1 (1599)	3.95 (2319)	4.04 (2028)	3.83 (677)	3.78 (1470)
7. La acción formativa ha sido realista y práctica	3.53 (213)	4.21 (938)	4.08 (1273)	4.04 (1923)	4.11 (1683)	3.98 (1244)	3.93 (1564)	4.29 (1380)	4.17 (1601)	4.18 (2311)	4.21 (2035)	3.99 (677)	4.06 (1473)
8. La documentación entregada ha sido de calidad	3.43 (213)	4.17 (938)	3.93 (1243)	3.96 (1925)	3.92 (1683)	3.72 (1244)	3.82 (1564)	4.02 (1362)	4.07 (1603)	4.06 (2323)	4.07 (2017)	3.64 (662)	3.71 (1402)
9. Las condiciones del aula, mobiliario y ambiente de aprendizaje han sido buenas	3.58 (213)	4.25 (938)	4.08 (1274)	4.05 (1924)	4.26 (1683)	4.08 (1244)	4.1 (1564)	4.02 (1385)	4.05 (1594)	4.22 (2309)	4.13 (2032)	4.03 (676)	3.99 (1468)
Utilidad, valoración global	PAS 2000	PAS 2001	PAS 2002	PAS 2003	PAS 2004	PAS 2005	PAS 2006	DIP. 2003	DIP. 2004	DIP. 2005	DIP. 2006	IAD 2005	IAD 2006
10. El curso recibido es útil para mi puesto de trabajo	4.05 (215)	4.41 (938)	4.21 (1274)	4.27 (1925)	4.21 (1683)	4.12 (1244)	4.01 (1564)	4.32 (1387)	4.23 (1592)	4.28 (2312)	4.22 (2038)	3.99 (674)	3.96 (1466)
11. El curso recibido es útil para mi formación personal	3.88 (215)	4.32 (938)	4.25 (1274)	4.29 (1925)	4.31 (1683)	4.26 (1244)	4.16 (1564)	4.48 (1387)	4.44 (1604)	4.44 (2323)	4.38 (2044)	4.26 (677)	4.28 (1475)
12. El curso merece una valoración global buena	3.79 (215)	4.32 (938)	4.21 (1274)	4.23 (1925)	4.29 (1683)	4.16 (1244)	4.16 (1564)	4.39 (1385)	4.29 (1600)	4.31 (2303)	4.29 (2036)	4.18 (677)	4.2 (1472)

Tabla 7.2. Resultados globales de las preguntas de formato cerrado del cuestionario de satisfacción en las distintas organizaciones a lo largo del tiempo.

<i>Coordinación</i>	DIP. 2004	DIP. 2005	DIP. 2006
1. El coordinador ha estado disponible durante la realización del curso	4.69 (1620)	4.75 (1818)	4.62 (1911)
2. Ha llevado correctamente el control de asistencia	4.76 (1629)	4.77 (1819)	4.64 (1907)
3. Ha solucionado debidamente los imprevistos del curso	4.68 (1611)	4.7 (1813)	4.58 (1893)
4. Valoración óptima del coordinador/a	4.71 (1618)	4.72 (1818)	4.59 (1902)

Tabla 7.3. Resultados de las preguntas de formato cerrado del cuestionario de satisfacción referidas a la coordinación.

<i>Profesorado</i>	PAS 2000-05	PAS 2006	DIP. 2003	DIP. 2004	DIP. 2005	DIP. 2006	IAD 2005	IAD 2006
1. El profesor domina la materia	4.5 (10171)	4.22 (1553)	4.55 (2902)	4.51 (3623)	4.64 (3924)	4.57 (3868)	4.43 (3543)	4.45 (7948)
2. Tiene capacidad para transmitir con claridad sus ideas	4.28 (10174)	4.04 (1553)	4.4 (2914)	4.29 (3629)	4.44 (3933)	4.46 (3859)	4.12 (3536)	4.17 (7941)
3. Las enseñanzas transmitidas se adecuan a los objetivos y contenidos del curso	4.22 (10175)	3.96 (1553)	4.34 (2897)	4.25 (3617)	4.4 (3930)	4.38 (3857)	4.11 (3532)	4.17 (7940)
4. Es ameno en sus intervenciones	4.15 (10178)	3.99 (1553)	4.31 (2903)	4.14 (3619)	4.32 (3922)	4.35 (3860)	3.91 (3535)	3.97 (7943)
5. Motiva y despierta el interés en los asistentes	4.11 (10179)	3.96 (1553)	4.31 (2904)	4.13 (3624)	4.32 (3927)	4.37 (3856)	3.95 (3532)	3.96 (7942)
6. El profesor merece una buena valoración global	4.2 (10166)	4.08 (1553)	4.43 (2898)	4.3 (3590)	4.46 (3912)	4.47 (3850)	4.14 (3483)	4.23 (7881)

Tabla 7.4. Resultados de las preguntas de formato cerrado del cuestionario de satisfacción referidas al profesorado.

A continuación, se presentan estos datos en distintas figuras en función de las diferentes organizaciones y diferenciando además entre la evaluación de los cursos en sí, de la coordinación y del profesorado.

Centro de Formación del PAS: cursos

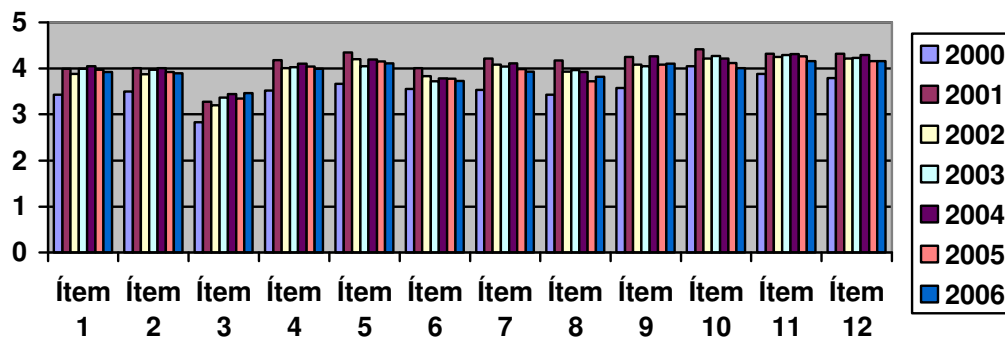


Figura 7.2. Medias obtenidas en la evaluación de la satisfacción con los cursos en el Centro de Formación y Perfeccionamiento del PAS.

En la figura 7.2, se presentan las valoraciones obtenidas en el Centro de Formación y Perfeccionamiento del PAS en relación a los 12 ítems referidos a las acciones formativas. Tras una visual general, destaca el hecho de que las valoraciones más bajas se obtuvieron en el ítem 3 referido a si la duración fue acorde con los objetivos y contenidos.

Comparando entre años, en líneas generales puede decirse que en el 2000 se obtuvieron las valoraciones más negativas; en el 2001, las más positivas.

De manera global, se puede decir que las valoraciones fueron bastante positivas, estando todas por encima de 4 o al menos rondando este valor, a excepción del ya comentado ítem 3 referido a la duración y las puntuaciones también ya comentadas del año 2000 en general.

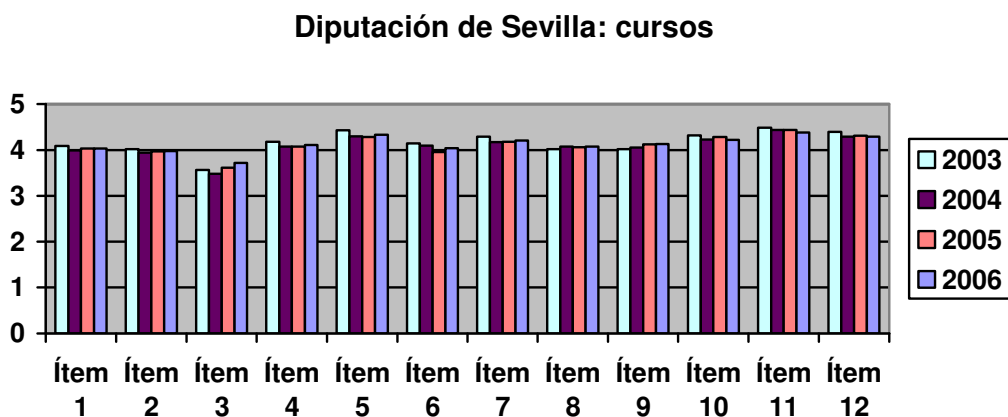


Figura 7.3. Medias obtenidas en la evaluación de la satisfacción con los cursos en Diputación de Sevilla.

La figura 7.3, donde se muestran las valoraciones obtenidas en los cursos de Diputación, presenta algunas similitudes con los resultados encontrados en el Centro de Formación del PAS. Así, el ítem 3 referido a la duración de la acción formativa es el valorado más negativamente; y, en general, el resto de ítems son positivamente valorados, estando sus medias por encima de 4 o rondando este valor.

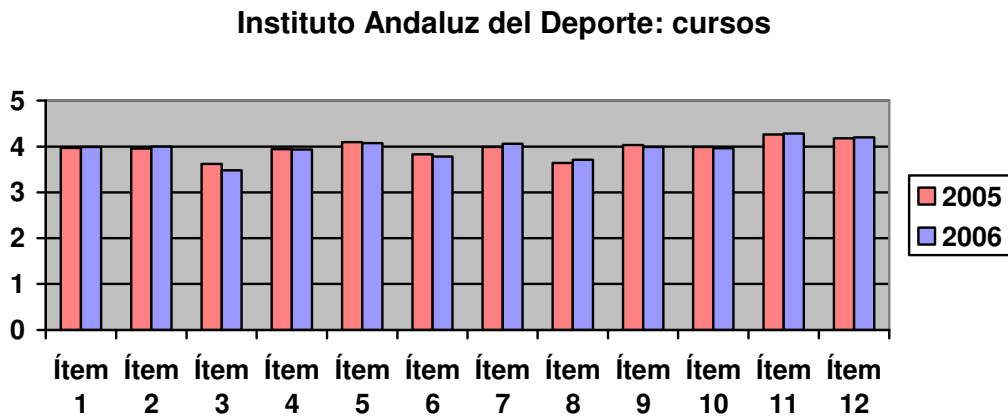


Figura 7.4. Medias obtenidas en la evaluación de la satisfacción con los cursos en el Instituto Andaluz del Deporte.

En la figura 7.4, se presentan los valores que obtuvieron los cursos impartidos en el IAD. El primer aspecto a destacar es que no existe gran diferencia entre los resultados obtenidos en el 2005 y el 2006. Y estas valoraciones siguen un patrón similar a la encontrada en el Centro de Formación y Perfeccionamiento del PAS y en Diputación: bastante positivas, rondando el 4 de media, aunque algo más bajas en el ítem 3 referido a la duración de la acción formativa y el ítem 8 referido a la documentación entregada.

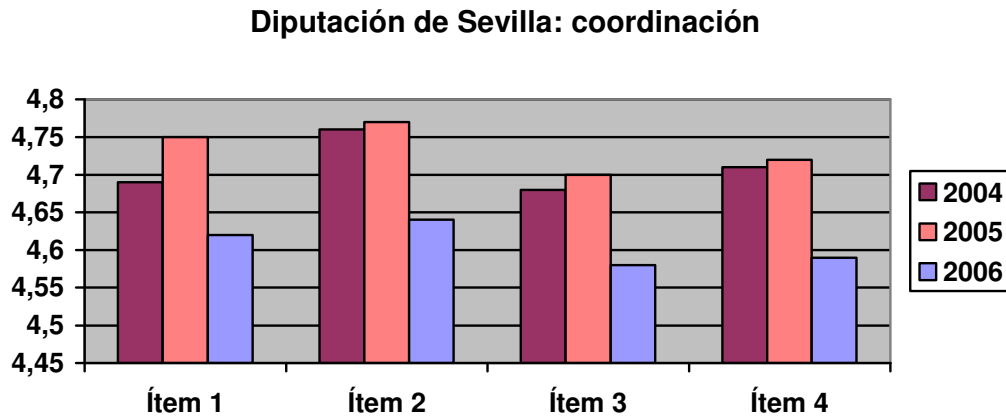


Figura 7.5. Medias obtenidas en la evaluación de la satisfacción respecto a la coordinación en Diputación de Sevilla.

La figura 7.5 hace referencia a las valoraciones otorgadas a la coordinación en Diputación de Sevilla como única organización que dispone de tal figura en su formación. Todas las medias son muy positivas, estando por encima del 4.5. El ítem mejor valorado fue el segundo, referido al control de asistencia. El año 2005 fue el mejor valorado y el 2006 el puntuado más negativamente, dándose sin excepción en todos los ítems (aunque no por ello las puntuaciones en el último año 2006 dejaron de ser positivas).

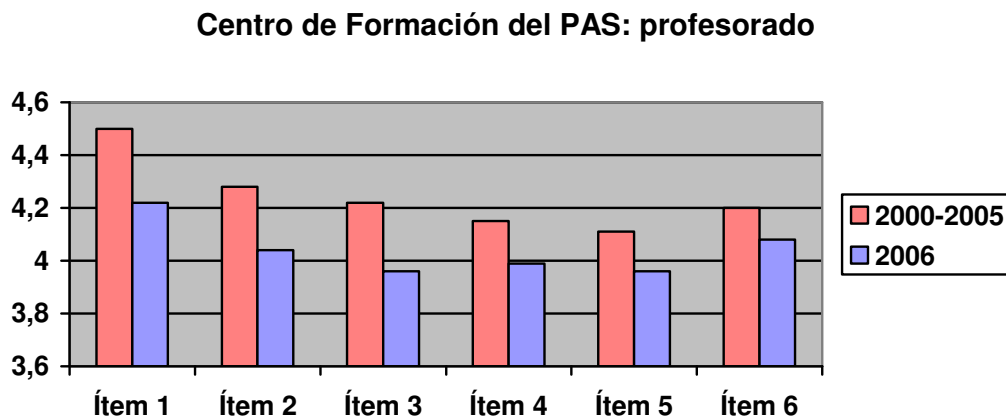


Figura 7.6. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en el Centro de Formación y Perfeccionamiento del PAS.

La figura 7.6 presenta las medias del profesorado del Centro de Formación y Perfeccionamiento del PAS desde el año 2000 al 2005 en conjunto y del 2006 por otro lado. Todas las valoraciones fueron positivas, estando las medias por encima de 4 o

cercanas a este valor. El ítem valorado más positivamente fue el referido al dominio de la materia por parte del profesor (ítem 1); el menos valorado, el referente a la motivación que el profesorado logró despertar en los participantes (ítem 5).

Diputación de Sevilla: profesorado

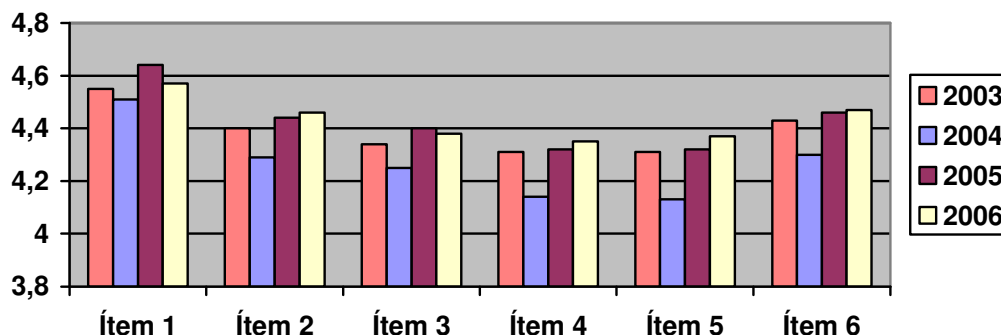


Figura 7.7. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en Diputación de Sevilla.

El profesorado de las acciones formativas de Diputación de Sevilla obtuvo también buenas puntuaciones, todas por encima del 4 de media. Los años en los que la valoración fue más positiva fueron los dos últimos (2005 y 2006); el año con valoración más baja, el 2004. En líneas generales, puede afirmarse que el ítem mejor valorado fue el 1 referido al dominio de la materia por parte del profesorado y los valorados de manera más negativa, el 4 y el 5, referidos a si el profesorado es ameno y a la motivación e interés que despiertan respectivamente.

Instituto Andaluz del Deporte: profesorado

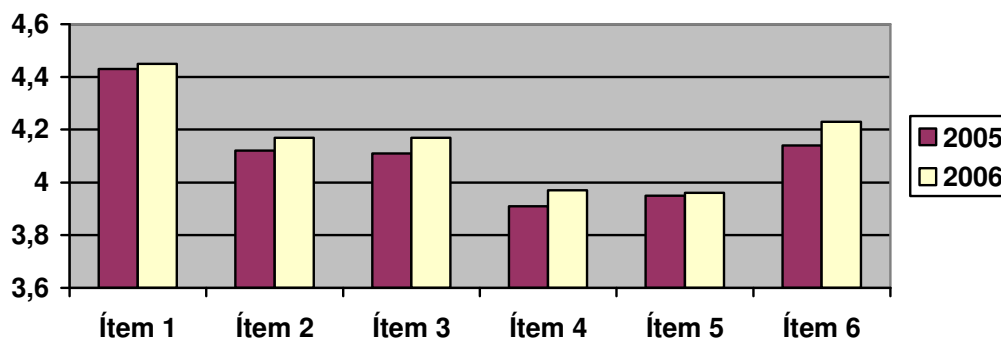


Figura 7.8. Medias obtenidas en la evaluación de la satisfacción respecto al profesorado en el Instituto Andaluz del Deporte.

Aunque los resultados obtenidos por el profesorado en el IAD no fueron tan positivos como en los centros anteriormente comentados, también podría decirse que se obtuvieron medias bastante elevadas, cercanas al 4 o por encima de este valor. Aunque no existen grandes diferencias entre los resultados del año 2005 y los del 2006, en todos los ítems se obtuvieron valoraciones algo más elevadas en este último año. Por último cabe destacar que, siguiendo con la línea de los resultados obtenidos en los centros anteriormente comentados, el ítem mejor valorado fue el referido al dominio de la materia por parte del profesorado (ítem 1) y los valorados más negativamente, los

referidos a si resultó ameno y a si motivó y despertó interés en los participantes (ítems 4 y 5 respectivamente).

4. DISCUSIÓN.

Tal como se comentó inicialmente, uno de los objetivos de la evaluación de la satisfacción de los participantes a las acciones formativas era proporcionar un posible referente de reflexión desde el que valorar en términos generales los aspectos positivos a mantener y otros posibles aspectos a mejorar. Por ello, a continuación se presentan puntos fuertes que destacan, puntos débiles y posibles mejoras para solventar dichas debilidades, aunque teniendo en cuenta que la denominación de puntos "fuertes" o "débiles" se considera algo relativo, nunca absoluto, ya que el objetivo es fomentar una mejora continua de las acciones formativas.

En referencia a los **puntos positivos** obtenidos de los resultados mostrados, destacan los recursos humanos; concretamente, el profesorado y la coordinación:

En general, los participantes en las acciones formativas hicieron valoraciones muy positivas del **profesorado** que impartió los cursos en modalidad presencial y de las personas que tutorizaron las acciones formativas en la modalidad teleformación a distancia, en especial con respecto al dominio de la materia que mostraron.

Los participantes consideraron que el papel desempeñado por los/as **coordinadores/as** de las acciones formativas fue óptimo y relevante para el buen desempeño de las mismas.

Respecto a los **puntos negativos** a mejorar, un primer aspecto que podría destacarse es la evolución de la **participación** en la evaluación de la satisfacción con las acciones formativas, caracterizada por una marcada irregularidad en la muestra. Este hecho puede estar relacionado con distintas cuestiones:

Las acciones formativas se impartieron en dos modalidades diferentes: a distancia a través de teleformación y de manera presencial. El porcentaje de participación es bastante más elevado en la última modalidad señalada ya que aún no se ha logrado dar con un método con el que las personas se animen a enviar vía web los cuestionarios cumplimentados. Resulta bastante más sencillo lograr que, en modalidad presencial, en el último cuarto de hora de la última sesión estando allí en persona, los participantes se molesten en responder al cuestionario. Teniendo en cuenta que la proporción de acciones formativas en modalidad teleformación fue aumentando a lo largo del tiempo, una posible explicación al descenso en la participación en los últimos años puede ser la influencia de esta variable.

Los inicios suelen ser duros. Generalmente, cuesta introducir en una dinámica nueva a los usuarios de un programa. Por ello, la participación no fue excesivamente elevada los primeros años en que se implementó la evaluación de la satisfacción en las tres organizaciones, aunque especialmente se notó en el IAD (también debido a que no se contabilizó la participación desde el principio del año 2005) y en el Centro de Formación del PAS. Este punto de inicio tan bajo puede ser la explicación de que el ascenso en la participación fuera tan pronunciado al principio pero que, posteriormente, este aumento se suavizara e incluso en ocasiones se convirtiera en declive.

Y en relación a los ítems del cuestionario, cabe destacar que la **duración** fue la menos valorada a lo largo del tiempo y en las distintas organizaciones en estudio. Este aspecto relaciona el tiempo de duración de la acción formativa con los temas que querían tratarse. Y también engloba otros aspectos como los horarios o la intensidad con la que se impartieron las acciones (cuántas horas cada día, si fueron días continuados o alternos, etc.).

Algunas **mejoras** se han ido introduciendo a lo largo del tiempo:

En relación al **instrumento**, como previamente se ha visto, se fueron realizando diferentes modificaciones:

1. En 2003, se amplió el número de ítems; concretamente, en referencia a la acción formativa en sí se pasó de 6 a 13 ítems y, respecto al profesorado, de 3 a 6.
2. En 2004, se logró con un único modelo evaluar acciones formativas de cualquier modalidad. Además, se pasó de que la coordinación se evaluara con un solo ítem general a que se usara para ello cuatro ítems más específicos.
3. En 2005, la información recogida con la pregunta abierta referida a otras acciones formativas consideradas necesarias se hizo más detallada, ya que se solicitó concretamente el nivel de dificultad, la modalidad, los destinatarios, la utilidad, los objetivos y los contenidos.

Se considera que con estos cambios el cuestionario mejoró porque cada vez se logró ir recogiendo una información más detallada acerca de la acción formativa en sí, sobre otras posibles acciones formativas a desarrollar y sobre el profesorado y la coordinación. Logrando un único cuestionario aplicable a cualquier modalidad de impartición, se consiguió además aumentar la comparabilidad de los resultados entre las diferentes acciones formativas.

Prueba de la calidad del dato recogido con este instrumento son los estudios de sus cualidades métricas (Camacho y Sánchez, 1997; Martínez, 1995), con los que se constató la validez del modelo de medida y que gozaba de unos niveles óptimos de fiabilidad, discriminación y precisión de medida.

También se ha notado una mejora en relación a la **difusión** de los resultados encontrados: mientras que anteriormente éstos eran de difícil acceso, actualmente cualquier persona que por algún motivo esté interesada en consultarlos puede acceder de manera fácil vía Internet.

Pero aún quedan cosas por hacer. Como **mejoras futuras**, se plantean las siguientes:

Respecto a la **poca participación** en la evaluación de la satisfacción por parte del alumnado en la modalidad de teleformación, habrían de ser estudiados los momentos de **entrega, cumplimentación, envío y recogida de los cuestionarios**, puesto que el problema podría encontrarse en una fase del proceso o en varias a la vez: en la entrega puede ocurrir que la información de la posibilidad de cumplimentar el cuestionario no llegue a todos los participantes; en la cumplimentación, por ejemplo, que los participantes no estén lo suficientemente motivados o que les parezca un

engorro; en el envío, por ejemplo, que sea costoso el método; y en la recogida que, al llegar primero al personal de INPRO en el caso de Diputación y después al personal de Formación Continua en lugar de ser recibida directamente, las probabilidades de que la información se pierda aumentan.

Una posible solución sería hacer más fácil la cumplimentación y envío del cuestionario, de tal manera que sólo tuvieran que marcar con el ratón en una opción de respuesta, en lugar de tener que escribir el número elegido, que es como aparece ahora; y que sólo tengan que pinchar un botón para que se enviara directamente, sin necesidad de adjuntar el documento a un correo electrónico como ocurre por ejemplo en Diputación de Sevilla.

La probabilidad de extraviar la información disminuiría considerablemente si se creara una base a la que llegaran todos los datos enviados, tal y como ya se hace en el Centro de Formación y Perfeccionamiento del PAS, y el personal evaluador tuviera acceso directo a ella. Además, aligeraría el proceso de codificación de datos, con lo que aumentaría la eficiencia del método de evaluación.

Respecto a las **características de la muestra**, hasta ahora los cuestionarios de satisfacción fueron de cumplimentación anónima. Debido a esto, los posibles cálculos a realizar están muy limitados y es difícil ir más allá de meros resultados descriptivos generalistas. Por ello, se espera tener un control personalizado de los datos solicitando cierta información identificativa, con lo que se conseguiría tener un seguimiento individualizado a lo largo del proceso de evaluación y se abriría la posibilidad de poner en relación diferentes variables e incluso de evaluar impacto más adelante. La organización más avanzada en este sentido es el Centro de Formación y Perfeccionamiento del PAS, donde se recogen algunos de estos datos y se ha llegado a implementar la evaluación de la transferencia.

Sobre la **duración** de las acciones formativas, algunas posibles mejoras son las siguientes:

- Tratar de adecuar los objetivos y contenidos del curso a la duración: a partir del conocimiento del número de horas del que se va a disponer, sería trabajo del profesorado crear una planificación clara y realista de los contenidos a tratar y los objetivos a alcanzar, lo que aumentaría además la probabilidad de **conseguir los objetivos previstos**.
- Cuando se detectara que el contenido que se necesita impartir excede a la duración de la que se dispone, una posible solución sería dividir en dos cursos de distintos niveles dicha información, y así tener la posibilidad de **tratar todos los temas de interés con suficiente profundidad**.
- **Mejorar el horario**: en numerosas ocasiones, las quejas acerca de la duración no se deben tanto al número de horas del curso, sino a la inadecuación de los horarios. En este sentido, dos son las propuestas de mejora más señaladas por los asistentes a las acciones formativas al responder al ítem 13 referente a posibles aspectos que mejorar en próximas acciones:
 - Crear un horario menos concentrado: parece ser que las personas se cansan en exceso cuando se dan muchas horas en pocos días. Esto puede tener

repercusiones negativas en la adquisición de conocimientos por cuestiones de fatiga y de poco tiempo de asimilación y consolidación de lo aprendido.

- Posibilidad de elección mañana/ tarde: este hecho aumentaría la satisfacción de los participantes porque facilitaría su asistencia ya que, por ejemplo, en algunas situaciones un alto volumen de trabajo urgente podría imposibilitar la asistencia de una persona en horario laboral.

En referencia al **instrumento** utilizado, además del estudio de las cualidades métricas ya realizado, se pretende analizar la invarianza de los datos obtenidos con el cuestionario de satisfacción en las diferentes organizaciones, con la finalidad de confirmar que mide el mismo concepto independientemente de la población de la que se recogieron los datos, con lo que se confirmaría la capacidad de generalización en el uso de este instrumento.

En relación a la **difusión** de los resultados obtenidos, a pesar de que en los últimos años ha mejorado, podrían proponerse algunas mejoras que quedaron por hacer:

En ocasiones, los evaluadores no pueden dar respuesta inmediata a los participantes de las acciones formativas porque la manera como está planteado el proceso de evaluación exige una alta inversión en tiempo para obtener los resultados. Esto ocurre principalmente en Diputación de Sevilla y el IAD, ya que en el Centro de Formación del PAS se consiguió automatizar bastante más el proceso. Por esto, una mejora que se propone es lograr un mayor grado de **automatización** del proceso en la línea de lo obtenido en la última organización mencionada.

En la página web del IAD, los resultados de evaluación se presentan en el informe global, en un solo archivo con lo que puede ocurrir que, si a una persona le interesa el resultado obtenido en una acción formativa concreta, le resulte muy engorroso tener que buscar en todo el documento. En las páginas web de Diputación de Sevilla y del Centro de Formación del PAS se presentan los resultados de las acciones formativas únicamente por separado, con lo que una persona interesada por ejemplo en el proceso de evaluación no encontraría la información que busca. Ante esta situación, la propuesta de mejora se concreta en aunar la presentación de los resultados en las tres organizaciones, mostrando en los tres casos los informes generales para que los interesados puedan ahondar en cuestiones de procedimiento y así se promueva la claridad y transparencia del proceso; y los informes individuales de cada acción formativa por separado para aquellas ocasiones en que el dato que interesa es concreto de una sola acción.

5. CONCLUSIONES.

La información sobre las opiniones de los participantes en la acción formativa es de mucha utilidad, pues una acción formativa que no es bien recibida por los participantes, aunque su diseño teórico sea bueno, no se puede considerar un éxito.

Al ser el nivel de evaluación de los resultados más fácilmente aplicable, se tuvo la oportunidad de presentar, a grandes rasgos, cómo se realizó en tres organizaciones distintas y los resultados descriptivos que se obtuvieron.

Tanto la metodología seguida como los resultados obtenidos en estas tres organizaciones presentaron grandes similitudes, aunque podría decirse que la satisfacción de los participantes a las acciones formativas promovidas por el IAD fue algo más baja que la obtenida en las otras dos organizaciones.

Aunque son ya muchos los años en los que se ha evaluado la satisfacción de los participantes y fueron muchas las mejoras implantadas a lo largo del tiempo, aún quedan aspectos que pulir centrados principalmente en las siguientes cuestiones: la duración de las acciones formativas, la participación, la identificación de los usuarios, el instrumento utilizado y la agilización y automatización del proceso.

La falta de operacionalización de los constructos en los que generalmente se centra la evaluación y la carencia de estudios de validez llevó a la necesidad de elaborar un instrumento de medida del constructo **satisfacción**, estudiar sus cualidades métricas y validar el modelo utilizado.

Este primer nivel de evaluación de resultados a través del estudio del grado de satisfacción de los participantes es interesante, pues de él se obtiene información que permite una primera aproximación en relación a la detección de aspectos que mejorar en las acciones formativas, pero hay que ir más allá. El curso puede haber despertado reacciones favorables pero no haber sido todo lo eficaz que debiera. Se pueden obtener buenas valoraciones y, sin embargo, haber servido para poco en cuanto a los conocimientos o habilidades adquiridas o en cuanto a la aplicación en el puesto de trabajo de estos conocimientos (Campbell, 1971). Es por ello que, en el **capítulo 8** que se presenta a continuación, se sigue avanzando en el proceso de evaluación ahondando en niveles más profundos: evaluación del conocimiento y de la transferencia.

CAPÍTULO 8. EVALUACIÓN DESPUÉS DE LAS ACCIONES FORMATIVAS (II). HACIA LA EVALUACIÓN DE LOS NIVELES SUPERIORES: CONOCIMIENTO Y TRANSFERENCIA

1. Evaluación del conocimiento.....	289
1.1. Introducción	289
1.2. Método	291
1.2.1. Muestra.....	291
1.2.2. Instrumentos	293
1.2.3. Procedimiento	295
1.3. Resultados	298
1.4. Discusión.....	300
2. Evaluación de la transferencia.....	305
2.1. Introducción	305
2.2. Método	306
2.2.1. Muestra.....	306
2.2.2. Instrumentos	306
2.2.3. Procedimiento	307
2.3. Resultados	308
2.4. Discusión.....	309
3. Conclusiones	314

CAPÍTULO 8. EVALUACIÓN DESPUÉS DE LAS ACCIONES FORMATIVAS (II). HACIA LA EVALUACIÓN DE LOS NIVELES SUPERIORES: CONOCIMIENTO Y TRANSFERENCIA.

En el capítulo anterior, se planteó la evaluación de la satisfacción de los participantes a las acciones formativas, fase de evaluación de resultados que, según Kirkpatrick (1999), es la más superficial y más sencilla de poner en marcha, por lo que su uso está bastante extendido entre las organizaciones.

En el presente capítulo 8, se sigue con la evaluación de los resultados según el modelo de Kirkpatrick, pero pasando a unas fases que requieren mayor profundización en el estudio de los casos y son más difíciles de implantar, por lo que su uso no es tan común: la evaluación de los **conocimientos** adquiridos en la acción formativa y la **transferencia** de estos conocimientos a las situaciones concretas planteadas en el puesto de trabajo.

1. EVALUACIÓN DEL CONOCIMIENTO.

1.1. Introducción.

El **aprendizaje** se puede definir como el grado en que los participantes en una acción formativa cambian actitudes, amplían conocimientos y/o mejoran habilidades como consecuencia de asistir a la misma (Kirkpatrick, 1999).

Su resultado en muchas ocasiones es independiente del nivel de evaluación anteriormente tratado; así como la evaluación de la satisfacción se centra en el fenómeno de la percepción subjetiva que los participantes tuvieron de la acción formativa, en esta fase la evaluación apunta hacia el fenómeno del aprendizaje que, como constructo, es algo más objetivo. Con frecuencia, las medidas de reacción más gratificantes se corresponden con un buen índice de aprendizaje, pero no siempre es así (Ramírez del Río, 1997).

El **objetivo** de la **evaluación de conocimientos** es proporcionar evidencias empíricas sobre la medida en que los participantes asimilaron los conocimientos, habilidades o actitudes que la acción formativa pretendía; en definitiva, conocer si los objetivos pedagógicos se cumplieron.

Para disponer de un referente de comparación válido desde el que poder evaluar “el valor añadido” de la acción formativa, aislando el efecto provocado por dicha acción del de otras posibles variables influyentes, es necesario medir qué sabían los participantes antes de implementar la acción formativa y qué supieron tras la misma (Chacón, López y Sanduvete, 2004).

En este sentido, **antes de iniciar la acción formativa** se persiguen fundamentalmente los objetivos de conocer el nivel de conocimientos del alumnado, y que el formador defina empíricamente los objetivos de la acción y los contenidos que va a impartir. **Finalizada la acción formativa**, se persiguen objetivos relacionados con la

determinación de las cotas de aprendizaje alcanzadas y la identificación de los puntos débiles de la prueba utilizada para la evaluación del aprendizaje.

En definitiva, el objetivo de la evaluación de conocimientos no es conocer el rendimiento a nivel individual de cada participante (de hecho, los instrumentos se cumplimentan de manera anónima), sino que se pretende (Chacón, Pérez, Sanduvete, Holgado, Sánchez y Lozano, 2007):

- Que, al tener que elaborar una prueba de medición de conocimientos, el profesorado tenga que reflexionar de una forma más sistemática acerca de los contenidos en los que va a incidir y profundizar, de tal manera que se conseguiría una buena planificación de los objetivos a lograr y cierta adecuación al tiempo disponible.
- Que, al estudiarse el rendimiento del grupo en su globalidad, el profesorado pueda detectar aspectos en los que tendría que incidir con mayor profundidad y/o claridad en siguientes acciones formativas ya que, considerándose aspectos importantes de los contenidos, no fueron comprendidos por los participantes.
- Que la realización de la prueba previa pueda ayudar al profesorado a determinar el nivel de conocimiento en que se encuentra su alumnado y, dependiendo de éste, adaptar el ritmo de aprendizaje. También podría servir para la homogeneización de los grupos si se pudiese aplicar con suficiente antelación.

Además de llevar a cabo la evaluación de conocimientos de cada una de las acciones formativas, también se hizo un estudio anual de los resultados en conjunto, con lo que se pretendía, además de emitir juicios de valor válidos sobre el nivel de consecución de los objetivos, la obtención de un flujo continuo de retroalimentación del sistema que permitiera poner en evidencia aquellos elementos que no resultaran del todo satisfactorios y, por tanto, poner en marcha, una vez conocidas las deficiencias, los mecanismos que permitieran su reajuste.

Según la literatura, respecto a los instrumentos utilizados para medir conocimientos, es generalizado el uso de **pruebas de rendimiento de elección múltiple** que, en el mejor de los casos, son administradas antes y después de la implementación de la acción formativa tal y como aquí se propone (Artés, 1998; Kirkpatrick, 1999).

Entre las técnicas de **análisis** que la literatura recomienda, se encuentra la distribución de frecuencias, medidas de tendencia central tales como la media y mediana, medidas de variabilidad tales como el rango, y medidas de la ganancia de conocimientos que, básicamente, se traducen en el uso de pruebas de significación para la diferencia de medias (*T* de *Student* mayoritariamente o la prueba de *Wilcoxon*, en función de si los datos cumplen o no los supuestos). Excepcionalmente, se plantea el uso de series temporales y el análisis de la covarianza para la medida del cambio (Arvey y Cole, 1989).

A continuación, se presenta la metodología utilizada y los resultados obtenidos en Diputación de Sevilla y el Centro de Formación del PAS, ya que en el IAD no llegó a realizarse evaluación de conocimientos. Concretamente, en Diputación de Sevilla se presenta la evaluación de conocimientos de los años 2005 y 2006 ya que anteriormente

sólo se había dado un estudio piloto donde se implantó esta evaluación en un reducido número de acciones formativas; y en el Centro de Formación y Perfeccionamiento del PAS, se presentan los datos desde el año 2000 al 2005, por no encontrarse disponibles los del 2006.

1.2. Método.

1.2.1. Muestra.

En la tabla 8.1 que a continuación se presenta, aparece la evolución de la muestra en las pruebas de conocimiento tanto en el Centro de Formación del PAS como en Diputación de Sevilla.

En la primera columna, se muestra la organización concreta a la que pertenecen los datos; la segunda columna presenta el año; la tercera columna, el número total de cursos que se impartieron (concretamente, todos los valores hacen referencia a cursos en global, sin diferenciar entre ediciones, a excepción del dato referente a Diputación 2006 donde sí se contabilizan las distintas ediciones); en la cuarta columna se presenta el número de acciones donde se llevó a cabo la evaluación de conocimientos; la quinta columna muestra el porcentaje de acciones formativas, respecto al total, donde se llevó a cabo la evaluación de conocimientos.

Las restantes columnas hacen referencia al número de participantes; concretamente, la sexta columna se refiere al número total de participantes; la séptima, al número de personas que participaron en una acción formativa donde se evaluó el conocimiento adquirido; la octava, al número de cuestionarios emparejados que pudieron recogerse (es decir, al número de personas cuyos resultados pudieron formar parte de la evaluación de conocimientos, al haberse recogido de ellas tanto la prueba previa al inicio de la acción formativa como la posterior a la finalización de ésta); y la novena y última presenta el porcentaje de personas cuyos resultados pudieron formar parte de la evaluación en relación al total de personas que participaron en alguna acción formativa donde se evaluó el conocimiento adquirido.

Las casillas en blanco se deben a la no disponibilidad del dato.

ORG.	AÑO	CURSOS	CURSOS CONOCIM.	% CURSOS	SUJETOS	SUJETOS CONOCIMIENTO	PRUEBAS VÁLIDAS	% SUJETOS
CENTRO DE FORM. DEL PAS	2000	36	23	63.89	1512	---	---	---
	2001	46	13	28.26	1839	---	---	---
	2002	62	14	22.58	1664	---	---	---
	2003	64	37	57.81	2450	---	---	---
	2004	71	28	39.44	3099	---	---	---
	2005	91	17	18.68	---	---	---	---
DIP.	2005	83	29	34.94	3093	1603	987	61.57
	2006	163	95	58.28	3139	1690	1164	68.88

Tabla 8.1. Participación en la evaluación de conocimientos en las distintas organizaciones a lo largo del tiempo.

El porcentaje de acciones formativas donde se implementó la evaluación de conocimientos en cada uno de los planes del Centro de Formación del PAS se mostró

oscilante a lo largo del tiempo; en Diputación, a pesar de que sólo se tienen datos de dos planes de formación, parece ser que tiende a la consolidación poco a poco.

De todos modos, en ninguno de los dos casos puede decirse que el porcentaje de acciones formativas en las que se implementó la evaluación de conocimientos fuera elevado, pues nunca se pasó del 65% del total.

La figura 8.1 muestra los porcentajes de acciones formativas donde se realizó la evaluación de los conocimientos adquiridos por los participantes.

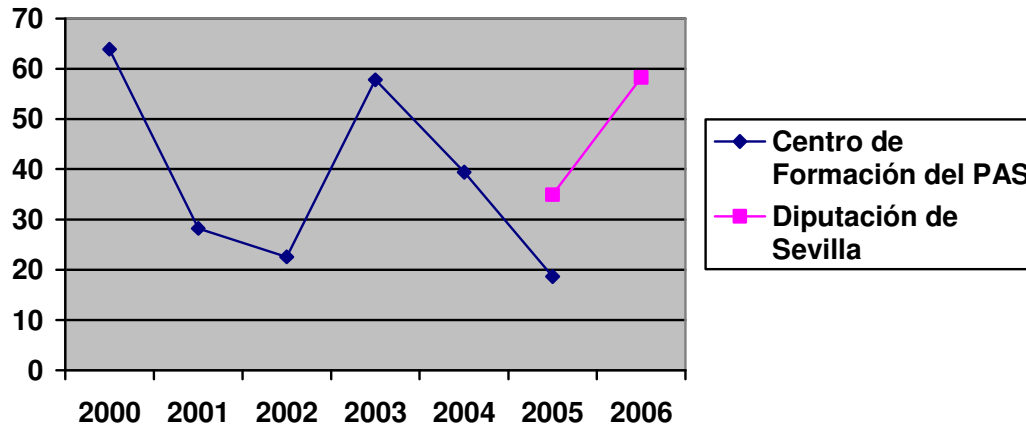


Figura 8.1. Porcentaje de acciones formativas donde se llevó a cabo la evaluación de conocimientos en las distintas organizaciones a lo largo del tiempo.

Concretamente, en lo que al Centro de Formación y Perfeccionamiento del PAS respecta, en el año 2000, del total de 36 acciones formativas que se llevaron a cabo, en 23 de ellas se evaluaron los conocimientos adquiridos, lo que supuso el 63.89% del total; en el 2001 y el 2002, este porcentaje fue disminuyendo: en el 2001, en sólo 13 de las 46 acciones formativas que se implementaron se evaluaron los conocimientos adquiridos, lo que supuso el 28.26% del total; y en 2002, se evaluaron 14 de 62 (el 22.48%). En el año 2003, se presentó un aumento de la implementación de la evaluación de los conocimientos adquiridos, puesto que se evaluaron 37 de 64 acciones formativas (el 57.81%); pero a partir del 2004, el porcentaje volvió a disminuir, dándose el menor valor en el año 2005: concretamente, en el 2004 se evaluaron 28 de las 71 acciones formativas implementadas (el 39.44%); y en el 2005, 17 de 91 (el 18.68%).

En Diputación de Sevilla, de las 83 acciones formativas que se llevaron a cabo en el plan de formación de 2005, únicamente en 29 de ellas se evaluaron conocimientos, lo que supuso un porcentaje del 34.94%.

Respecto al número de personas, de las 3093 que finalizaron alguna acción formativa, 1603 participaron en aquellas en las que se llevó a cabo la evaluación de conocimientos, lo que supuso algo más de la mitad; concretamente, el 51.83%. De estas personas, se consiguió obtener la información necesaria para realizar el análisis (es decir, la medida previa y la posterior con el número de identificación para emparejarlas) en 987 casos, lo que supuso un porcentaje del 61.57%.

En Diputación de nuevo, esta vez en el año 2006, de las 163 acciones formativas que se llevaron a cabo, en 95 de ellas se evaluaron conocimientos, lo que supuso un porcentaje del 58.28% (mayor que en el año 2005).

Respecto al número de personas, de las 3139 que finalizaron alguna acción formativa, 1690 participaron en aquellas en las que se llevó a cabo la evaluación de conocimientos, lo que supuso algo más de la mitad; concretamente, el 53.83%. Y de estas personas, se consiguió obtener tanto la medida previa como la posterior en 1164 ocasiones, lo que supuso un porcentaje del 68.88%.

1.2.2. Instrumentos.

Las **pruebas de opción múltiple** que los participantes a las acciones formativas cumplimentaron para medir los conocimientos adquiridos fueron realizadas por el profesorado de cada acción, ya que es éste quien tiene los conocimientos sustantivos acerca de la materia y quien determinó qué objetivos habrían de cumplirse con la acción formativa que se impartió.

Como ayuda para la realización de dichas pruebas, se facilitaron las siguientes directrices a seguir (Chacón, Holgado, López y Sanduvete, 2006):

- **Tener siempre presente cuáles son los objetivos** que se persiguen con la acción formativa; una vez que estén claramente explicitados, procurar “traducirlos” a preguntas concretas en la prueba de evaluación que se elabore.
- Dadas las características de las acciones formativas, **la prueba ha de ser corta**; una duración en torno a los 15 minutos puede ser adecuada. Así, con **una prueba de 10 preguntas con tres/cuatro alternativas de respuesta**, contabilizando un minuto y medio por cada pregunta, el participante utilizaría alrededor de 15 minutos para responder al cuestionario.
- **Utilizar preguntas objetivas de tres/cuatro alternativas.**
- **Hacer que cada pregunta tenga una sola respuesta correcta.** Si hay más de una, los participantes se confunden, la puntuación es más difícil de obtener y la prueba pierde validez.
- **Procurar que las preguntas y sus alternativas se entiendan fácilmente.** El objetivo de la prueba es comprobar el grado de conocimiento de los alumnos; no evaluar su nivel de comprensión lectora. Para ello, es útil procurar:
 - Evitar los enunciados en forma negativa; son más difíciles de procesar.
 - Redactar frases cortas y evitar estructuras de frases complicadas.
 - Evitar o minimizar el uso de expresiones como “raras veces”, “a menudo”, o “generalmente”.
 - Asegurar la concordancia gramatical entre el enunciado general y las alternativas.

- Evitar alternativas o enunciados jocosos.
 - Conseguir que la longitud de las distintas alternativas de respuesta sea aproximadamente la misma.
 - Evitar las reiteraciones de texto innecesario en todas las alternativas.
 - Presentar ordenadamente cifras que pudiesen aparecer en las alternativas de respuesta.
 - Evitar el uso de la última alternativa “todas las anteriores son correctas”, “ninguna de las anteriores es correcta” y las del tipo “a) y b) son correctas”.
 - Evitar dar pistas topográficas sobre la respuesta correcta variando el tamaño de los espacios en blanco o destacando una de las respuestas múltiples como única alternativa razonable.
 - No ofrecer información engañosa. Las preguntas “con trampa” provocan respuestas equivocadas.
 - Distribuir la posición de las respuestas correctas al azar para evitar que las personas evaluadas deduzcan una posible pauta de posición de respuestas correctas.
- **Procurar que las respuestas múltiples no hagan referencia a las respuestas de otras preguntas.** Si las respuestas están estrechamente relacionadas entre sí, la puntuación del alumnado puede resultar gravemente afectada si contesta incorrectamente a la primera relacionada con otras. Esta situación puede afectar a los resultados y la validez de la prueba.
- **Utilizar diferentes niveles de dificultad en la prueba.** Redactar preguntas complicadas, medianamente difíciles y fáciles. En lo posible, redactar el doble de preguntas que compondrán la prueba, para posteriormente seleccionar de entre ellas la mitad que resulte más apropiada.
- **Simplificar la mecánica de la prueba.** El objetivo de la prueba es medir los conocimientos del alumnado; no su habilidad para comprender instrucciones complejas.
- **Revisar la prueba antes de realizarla** para comprobar que resulta adecuada y que responde a los objetivos de la acción formativa. Hacer con ella un estudio piloto si fuera posible.

Además de las pruebas de evaluación elaboradas por los formadores, otros instrumentos que se utilizaron fueron los siguientes:

Los softwares **SPSS 13.0** y/o **Microsoft Excel** para la introducción de los datos en una base, el cálculo de las medias previas y posteriores a la acción formativa y la realización de la prueba de contraste.

El software **Pumuki** para convertir los datos del formato propio de los archivos de SPSS a un formato compatible con el programa Metrix.

El software **Metrix** para el análisis métrico de la prueba.

1.2.3. Procedimiento.

Los asistentes **cumplimentaron** la misma prueba de opción múltiple (elaborada previamente por el formador de la acción formativa) dos veces: antes del comienzo de la acción formativa y una vez finalizada ésta. Los datos fueron recogidos de manera anónima, pero se dio un número de identificación a cada participante porque el análisis requería el emparejamiento de los resultados previos y posteriores de cada persona concreta, para luego poder hacer un estudio grupal, que es el que realmente interesaba.

En Diputación, donde existe la figura del **coordinador**, éste fue el encargado en cada acción formativa de administrar las pruebas de conocimiento tanto el primer día como el último, explicitando la necesidad de que cada participante recordara el número de identificación otorgado el primer día porque en la prueba posterior habría de marcar el mismo. Además, se encargó de entregar ambas pruebas en la sección de formación, donde se le hizo llegar al equipo de evaluación para la codificación y análisis de datos. En el Centro de Formación y Perfeccionamiento del PAS, donde se prescinde de coordinador, estas mismas tareas fueron realizadas por el **profesorado**.

Este es el medio de **aplicación** de las pruebas y de **recogida** de la información en la modalidad presencial; en teleformación, el cuestionario se puso a disposición en la plataforma virtual como primera actividad a realizar y tras finalizar la acción formativa como última tarea. En ambos momentos, los cuestionarios fueron recogidos y guardados por los responsables de la plataforma que, posteriormente, enviaron en un archivo las respuestas obtenidas.

Una vez recogidos los datos y codificados con el software SPSS 13.0 o Microsoft Excel, los archivos resultantes fueron tratados con el programa Pumuki para convertir la información a un formato que fuera compatible con el programa Metrix.

Con este programa, se **analizaron** los datos para determinar las cualidades métricas generales de la prueba (fiabilidad, discriminación y dificultad), y el funcionamiento de cada uno de los ítems por separado, con el estudio de la discriminación y la dificultad, principalmente. Siguiendo unos criterios explicitados, se decidió si cada ítem era o no mínimamente apropiado.

Concretamente, respecto a las **cualidades métricas de la prueba**, éstos fueron los índices que se consideraron óptimos:

- Respecto a la **fiabilidad**, definida como la consistencia de la medida, se consideró aceptable un índice superior a 0.6.
- En relación a la **discriminación**, considerada como la capacidad de la prueba para diferenciar a las personas con niveles más altos de conocimiento de aquellas con niveles más bajos, se aceptó como válido un valor superior a 0.3. Un ítem

discrimina bien si las respuestas de cada una de las personas al ítem están directamente relacionadas con sus puntuaciones globales.

- La **dificultad** es entendida como la proporción de aciertos sobre el número de respuestas dadas. Su fórmula por tanto es (Pérez, 2006-07):

$$D_j = \frac{A_j}{N_j}$$

donde:

A_j = número de aciertos.

N_j = número de respuestas dadas.

Se estimó para este índice un valor óptimo de 0.625, aunque este punto de corte varió en función del número de opciones de la pregunta. Puede oscilar entre 0 (máxima dificultad, pues implicaría que no hubo ningún acierto) y 1 (máxima facilidad, que implicaría que todos los que respondieron acertaron). Cuanto más extremo sea, como más inadecuado será interpretado, ya sea por excesiva facilidad o excesiva dificultad (de ninguna de las dos maneras resultaría ser una prueba discriminante).

Y en relación al **estudio métrico de cada ítem por separado**, los pasos que se siguieron para considerarlo o no adecuado y para mejorarlo en caso de que fuera necesario, fueron los siguientes:

1º. Estudio del nivel de **discriminación**. Un índice menor de 0.15 fue considerado inapropiado, por lo que se recomendaba al profesor que elaboró la prueba que eliminara dicho ítem concreto y los resultados obtenidos con esta pregunta no eran incluidos en el estudio de comparación de medias que se realizaba posteriormente. Generalmente, esta falta de discriminación del ítem iba asociada a un índice de dificultad inapropiado, ya fuera por excesiva dificultad (cercano a 0) por lo que la inmensa mayoría de participantes no acertaba o, por el contrario, por excesiva facilidad (cercano a 1), donde hasta las personas con puntuaciones más bajas en la prueba acertaban.

2º. Con los ítems que pasaron la primera criba (con un índice de discriminación al menos cercano al adecuado), se hizo un **estudio de las distintas opciones de respuesta**. Concretamente, se profundizó en la capacidad de discriminación del ítem analizando la correlación biserial (r_{bp}) de cada una de ellas. Este índice oscila entre -1 y +1 y su fórmula es (Palmer, Jiménez y Montaña, 2000):

$$r_{bp} = \frac{m_p - m_x}{S_x} \sqrt{\frac{p_i}{q_i}}$$

Siendo:

m_p = la media de las puntuaciones obtenidas en la prueba por aquellos que dieron como respuesta la opción en estudio.

m_x = la media de las puntuaciones de todos los que realizaron la prueba.

S_x = la desviación estándar de las puntuaciones obtenidas por todos los examinados en el test.

p_i = la dificultad del ítem.

$q_i = 1 - p_i$.

Se espera correlaciones positivas para la opción correcta y negativas para la incorrecta. Estudiando exclusivamente el numerador de la fórmula, esto implicaría que quienes respondieron la respuesta correcta obtuvieron una puntuación alta en la prueba (mayor que la media obtenida por el grupo), mientras que quienes respondieron con una opción incorrecta obtuvieron una puntuación baja en la prueba (menor que la media).

3°. Una vez realizado este estudio para cada una de las opciones de los distintos ítems, se pasó a hacer **recomendaciones** al profesorado para mejorar la pregunta:

- Cuando una opción no era elegida por ningún participante, se recomendaba que se cambiara por otra por ser excesivamente fácil de discriminar: todos la habían descartado por lo que se deduce que, por alguna razón, era evidente que no se trataba de la respuesta correcta.
- Cuando una opción incorrecta obtenía una correlación biserial negativa ($-1 < r_{bp} < 0$), se recomendaba que se cambiara por otra por resultar ambigua y por provocar dudas a quienes obtuvieron altas puntuaciones en la prueba.

Después de realizar el estudio de los índices métricos, incluyendo sólo los resultados obtenidos a través de los ítems que fueron considerados adecuados, se realizó con los programas SPSS 13.0 o Microsoft Excel una comparación entre la media obtenida por el grupo antes de comenzar la acción formativa y la media obtenida tras finalizar ésta, para concluir si se dieron diferencias significativas en el nivel de conocimiento. Las pruebas que se realizaron fueron la T de Student cuando los datos cumplían los supuestos para pruebas paramétricas y la prueba de Wilcoxon en caso de necesitarse una prueba no paramétrica.

La difusión de esta información se dio por dos vías diferenciadas:

- Con los resultados del análisis métrico de la prueba, se elaboró un informe de cada acción formativa que se envió por correo electrónico al formador que había elaborado la prueba con la intención de que, en caso de que se repitiera dicha acción formativa, tuviera posibilidad de mejorar el instrumento de evaluación. Un modelo de este informe se presenta en el anexo XIII (pág. lxxvii).
- Las medias previas y posteriores y la significación de la comparación de cada acción formativa por separado se presentaron en Internet, en las correspondientes páginas web de ambas organizaciones, junto con los resultados de satisfacción, para que cualquier persona interesada pudiera consultarlo.

1.3. Resultados.

A continuación, a modo ilustrativo, en la tabla 8.2 se presenta un resumen de los índices de fiabilidad y discriminación que obtuvieron las pruebas de evaluación de conocimientos en el Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla, desde el año 2000 al 2005. En la primera columna aparecen los índices estudiados (fiabilidad y discriminación); en la segunda, los intervalos donde se encuadraron los resultados de cada prueba, colocando como punto de corte el valor mínimo aceptable: 0.6 para la fiabilidad y 0.3 para la discriminación, tal y como se aclara en la última columna; de la columna tercera a la octava, se presenta el número de pruebas que se encuadraron en cada uno de los intervalos desde el año 2000 al 2005.

	Intervalo	FRECUENCIAS						
		2000	2001	2002	2003	2004	2005	
Fiabilidad	[0-0.6]	13	8	6	25	19	11	<i>Valor mínimo aceptable > 0.6</i>
	(0.6-1]	10	5	8	12	9	6	
Discriminación	[0-0.3]	7	5	6	17	11	5	<i>Valor mínimo aceptable > 0.3</i>
	(0.3-1]	16	8	8	20	17	12	

Tabla 8.2. Índices de fiabilidad y discriminación de las pruebas de evaluación de los conocimientos utilizadas en el Centro de Formación del PAS.

Como puede apreciarse, en la mayoría de las ocasiones las pruebas no alcanzaron un nivel aceptable de fiabilidad; sin embargo, el grado de discriminación fue óptimo en la mayoría de las ocasiones, dándose esta tendencia a lo largo de prácticamente todos los años estudiados.

Y a continuación, de manera global, en la tabla 8.3 se presentan los resultados en el estudio de comparación de medias tanto en Diputación de Sevilla con en el Centro de Formación del PAS. En la primera columna, se diferencia entre ambas organizaciones; en la segunda, se presenta el año; la tercera columna muestra el número de acciones formativas que obtuvieron diferencias significativas al menos con un nivel de confianza del 95% al comparar las medias previas a la acción formativa y las obtenidas tras la finalización de ésta; la cuarta columna presenta el número de acciones formativas donde se evaluaron los conocimientos adquiridos; y la quinta y última columna presenta el porcentaje de acciones formativas donde se encontraron diferencias significativas en relación con el total de acciones evaluadas.

ORGANIZACIÓN	AÑO	SIGN.	TOTAL	%
CENTRO DE FORMACIÓN DEL PAS	2000	18	23	78.26
	2001	4	13	30.77
	2002	6	14	42.86
	2003	18	37	48.65
	2004	12	28	42.86
	2005	7	17	41.18
DIPUTACIÓN DE SEVILLA	2005	42	85	49.41
	2006	59	95	62.1

Tabla 8.3. Acciones formativas donde se encontraron diferencias significativas en el estudio de las medias obtenidas en las pruebas de conocimiento.

Antes de continuar, es interesante aclarar un par de cuestiones:

El número total de acciones formativas es muy dispar si se comparan las realizadas en el Centro de Formación del PAS y las de Diputación de Sevilla. Realmente, tales diferencias no existen de manera tan abismal; lo que ocurre es que en el Centro de Formación se evaluó cada acción formativa en global (en un estudio se incluyeron todas las ediciones de la misma acción) mientras que, en Diputación, las evaluaciones se realizaron por cada acción formativa por separado.

En la columna “sign.” donde se recoge el número de acciones formativas que presentaron diferencias significativas al 95% de nivel de confianza, se incluyen únicamente las ocasiones en que estas diferencias se encontraron a favor de la medida posterior; es decir, se contabilizaron sólo aquellas acciones donde la media obtenida tras la finalización de la acción formativa fue superior a la media obtenida antes de dicha acción, y la diferencia resultó ser estadísticamente significativa. Se excluyeron, por tanto, los casos excepcionales donde las diferencias significativas se presentaron a favor de la media tomada antes de la realización de la acción formativa.

En la figura 8.2, se presentan los porcentajes de acciones formativas que obtuvieron resultados significativos en el estudio de comparación de medias.

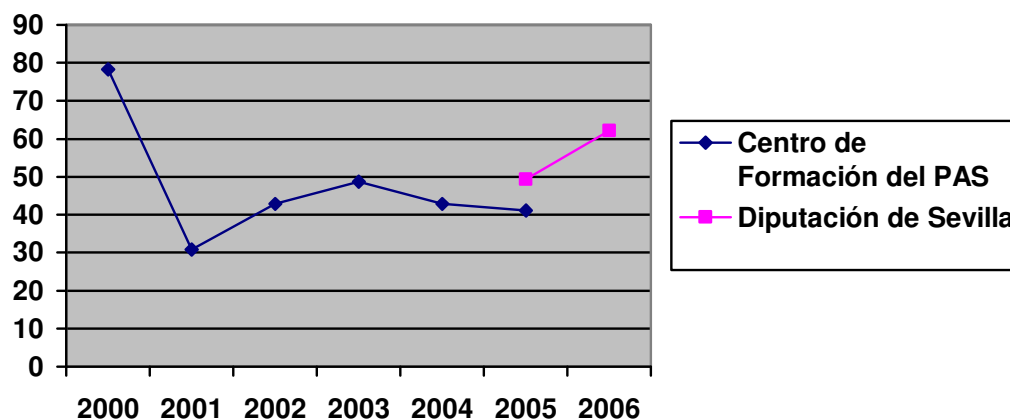


Figura 8.2. Evolución del porcentaje de acciones formativas que presentaron diferencias estadísticamente significativas en el estudio de evaluación de conocimientos.

Como puede apreciarse en la figura 8.2, en el Centro de Formación y Perfeccionamiento del PAS, tras un inicio algo inestable donde se obtuvo un porcentaje muy positivo en el año 2000 y un porcentaje muy bajo en el 2001, en los siguientes años se encontraron porcentajes bastante homogéneos; concretamente, el porcentaje de acciones formativas donde se encontraron diferencias estadísticamente significativas entre medias estuvo algo por debajo de la mitad del total desde el 2002 en adelante.

En Diputación de Sevilla, donde se obtuvieron datos de únicamente dos años, se detectó una mejora pasando de casi la mitad de acciones formativas con diferencias significativas a algo más del 60%.

1.4. Discusión.

A continuación, se presentan algunas fortalezas encontradas en el proceso de evaluación de conocimientos hasta ahora descrito, los puntos débiles descubiertos y posibles propuestas de mejora ligadas a éstos.

Como **puntos fuertes**, destacaron los siguientes:

1. Con esta evaluación, se consiguió **determinar si el alumnado adquirió los conocimientos previstos**, constructo más objetivo que la satisfacción con la acción formativa recibida. Se puede afirmar, por tanto, que con esta fase de evaluación se consiguió una información útil que aportó un grado más al nivel anterior para conocer con mayor detalle los resultados de cada acción formativa.
2. La información recogida con esta fase de evaluación se puede utilizar como prueba para entregar **certificados “con aprovechamiento”** en caso de que el alumnado llegue al nivel mínimo de conocimiento necesario, en lugar de únicamente certificados de asistencia. Esta está siendo la tendencia actual porque a quienes solicitan el currículo de un empleado o futuro empleado de su empresa no le interesa si éste acudió a alguna acción formativa, sino más bien si aprendió con ella y le sirvió para mejorar el desempeño en su puesto de trabajo.
3. Para el profesorado, la elaboración de esta prueba puede servir para realizar una **reflexión sistemática acerca de los contenidos** en los que van a incidir y profundizar, de tal manera que exija una buena planificación de los objetivos a lograr y cierta adecuación al tiempo disponible.
4. El **análisis de los índices métricos** que se elaboró de cada prueba fue **útil** para mejorarla en aquellas ocasiones en las que presentaban flaquezas infranqueables que impedían una buena medición del conocimiento de los participantes. Concretamente, se propusieron estrategias con las que mejorar los ítems que los formadores realizaron para que fueran representativos del material dado y realmente midieran lo que se pretendía. En ocasiones, se evalúa de manera excesivamente rápida y a mansalva sin cuestionarse la calidad del dato; en este caso, antes de valorar si hubo mejora en el grado de conocimiento tras la acción formativa, se estudió la capacidad de la prueba de medir dicho constructo.
5. La realización del pretest puede, además, ayudar al profesorado a **determinar el nivel de conocimiento del que parte su alumnado** y, dependiendo de éste, adaptar el ritmo de aprendizaje. También podría servir para la homogeneización de los grupos si se pudiesen aplicar con suficiente antelación.
6. La evaluación de conocimientos en Diputación de Sevilla parece que se fue **afianzando** más cada año desde el 2004 donde se realizó un estudio piloto con un número reducido de acciones formativas, tal y como reflejó la evolución en el número de participantes, de acciones formativas donde se implementó esta evaluación, y la significación de los resultados obtenidos. No se puede decir lo mismo de la evolución en el Centro de Formación y Perfeccionamiento del PAS, aspecto que se considera una debilidad que en el siguiente apartado se explicita.

Como **puntos débiles** detectados en esta fase de evaluación de conocimientos, destacan los siguientes aspectos:

1. En el Centro de Formación y Perfeccionamiento del PAS, el porcentaje de acciones formativas donde se evaluó conocimientos se presentó oscilante a lo largo del tiempo, debido a que **los formadores no tendieron a participar** en el proceso de evaluación por varios motivos:
 - Existen cursos que en los últimos años pasaron a ser seminarios de un solo día de duración, lo que supuso un primer motivo para no cumplimentar los cuestionarios ya que no fue posible dedicar el tiempo necesario para esta labor. Entre entrega, cumplimentación y recogida de estas pruebas dos veces en un día, pudo perfectamente haberse invertido más de media hora de un seminario que, por ejemplo, duró en total sólo cuatro.
 - Parte del profesorado carece de concienciación acerca de la importancia de cumplimentar los cuestionarios en pro de la calidad del plan de formación en general y de las acciones formativas concretas. Entre las causas que los formadores solieron alegar por no haber pasado a los participantes los cuestionarios se encontraron la falta de tiempo del que dispusieron para dar la materia y la escasa motivación de los asistentes por cumplimentarlas.
2. El **número de personas** que pasaron tanto por la prueba del pre-test como por la del post-test solió ser un número muy **bajo**, hecho que influyó en que en el análisis de la comparación entre una prueba y otra no se obtuvieran resultados significativos. Los resultados de aquellas personas que sólo cumplimentaron la prueba en una de las dos ocasiones, no fueron incluidos en el estudio de comparación.
3. El **proceso** en Diputación de Sevilla aún es excesivamente **lento**; cuando se acaba obteniendo el análisis métrico y se le hace llegar al profesorado los resultados, generalmente ha pasado tanto tiempo que no hay posibilidad de mejora de la prueba para la siguiente edición, porque ya está en marcha. Esto ocurre incluso cuando la evaluación se realiza a través de Internet. Habría que lograr agilizar el proceso con lo que se conseguiría optimizar la utilidad de esta evaluación:
 - Con el análisis métrico, el profesorado podría incluir las mejoras antes de la siguiente edición de la acción formativa.
 - Con el análisis de los resultados en la prueba previa, se conseguiría tener información de utilidad para la creación homogénea de los grupos en función del nivel previo de conocimiento acerca de la temática en cuestión que presentaran los participantes.

Algunos aspectos que conllevan el enlentecimiento de la salida de los resultados son los siguientes:

- Los datos han de ser codificados manualmente, incluso cuando las personas cumplimentaron las pruebas a través de internet porque, o llegan en papel

(impreso de un archivo) o llegan en un archivo de formato del que no se puede exportar directamente a los programas con los que se realiza el análisis.

- Para el análisis, los datos pasan por tres transformaciones: del formato de SPSS o Microsoft Excel al de Pumuki, y del de éste al de Metrix. Aunque el proceso de análisis se hace con programas informáticos y no manualmente, no están nada automatizados, por lo que se invierte gran cantidad de tiempo para obtener los resultados y aún más para elaborar el informe.
4. **¿Realmente las diferencias se deben a la acción formativa?** Para responder a esta pregunta, se propuso el diseño de grupo único con medida previa y posterior que podría representarse con la figura 8.3 que a continuación se presenta (Chacón, Shadish y Cook, en prensa):

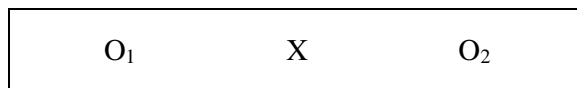


Figura 8.3. Diseño de un solo grupo con prueba previa y posterior.

“O₁” representa la medida previa; “O₂”, la posterior; y “X”, la acción formativa. Teniendo una medida previa con la que comparar la posterior, se consiguió obtener un punto de referencia a partir del cual poder concluir si realmente hubo cambios en el nivel de conocimiento del grupo antes y después de la acción formativa.

Pero aún quedan muchas amenazas a la validez que podrían estar influyendo en los resultados, como la historia (todo aquello que sucede durante la acción formativa pero que no tiene relación con ésta), la regresión estadística (comparar a partir de una medida previa extraordinariamente elevada o baja), o problemas de instrumentación y aplicación diferencial de la prueba.

Ligada a cada una de las cuatro carencias enumeradas, se presentan **propuestas de mejora** para tratar de solventarlas:

1. Para tratar de **aumentar el número de acciones formativas** donde se realice evaluación de conocimientos, y en vista de que una de las razones de la no puesta en marcha de este proceso es la desmotivación del profesorado, se hace necesario realizar un seguimiento de éste y concienciarlo de la importancia de la evaluación, de que es la figura clave para que pueda llevarse a cabo, de que está en sus manos, de que sin su colaboración es imposible. Es necesario hacer que sean conscientes de que administrar el cuestionario es una parte necesaria de la acción formativa, y no una decisión propia.

Concretamente, una posible medida a tomar al respecto sería realizar **seminarios formativos** en los que se les concienciara de la necesidad de realizar los cuestionarios para fomentar la calidad del proceso de evaluación de la formación, y en el que se les guiara acerca de cómo realizar las pruebas a partir de las directrices generales presentadas.

2. Respecto a la **muestra**, se propone lo siguiente:

- **Incrementar el número de participantes** en las pruebas de evaluación de conocimiento en ambos momentos: antes de llevar a cabo la acción formativa y tras finalizar ésta. Uno de los problemas que provocó la aparición de valores perdidos fue el hecho de que el alumnado puede faltar al 20% de la acción formativa y, aun así, lograr el diploma. Cuando no asistieron el primer o el último día, no realizaron una de las dos pruebas y la comparación no fue posible, por lo que los datos se eliminaron del análisis. Para evitar esto, se podría dar la posibilidad de realizar la prueba previa el segundo día a quienes faltaron el primero y permitir la realización de la prueba posterior en el penúltimo día del curso si el participante sabía previamente que el último día no iba a asistir.
- **Inferir valores perdidos.** Para tratar de evitar la pérdida de datos recogidos una posibilidad sería, en aquellos casos en los que se hubiera aportado la prueba previa o la posterior pero no ambas, inferir los valores que faltan. Por ejemplo, se podrían completar los huecos con las mismas respuestas dadas por el participante en la prueba que presentó; otra posibilidad sería dar de media en la prueba no presentada la media del grupo en dicha prueba. De este modo, la N aumentaría considerablemente, lo que mejoraría las condiciones del estudio de comparación de medias.
- **Identificación.** Como ya se comentó en referencia a la evaluación de la satisfacción, se espera en un breve periodo de tiempo activar un sistema de identificación de los datos, con el que se podrían relacionar los resultados de una persona en satisfacción y conocimientos, por ejemplo, con lo que las posibilidades de análisis de datos aumentarían en gran medida.

3. Para lograr la **agilización del proceso**, se presentan distintas posibilidades:

- Buscar una alternativa a la codificación manual de los datos: podría utilizarse, por un lado, la **máquina lectora** para los cuestionarios de las acciones formativas presenciales donde no se trabajara con ordenador; y por otro lado, podría crearse una **base de datos** de formato compatible con los programas con los que después se llevan a cabo los análisis, donde desembocaran todas las respuestas dadas por los participantes de las acciones formativas en las que se usa ordenador (todas las de modalidad teleformación y algunas presenciales concretas).
- **Automatizar en mayor medida el proceso de análisis de datos:** lo ideal sería que, de la misma base en la que se codifican los datos inicialmente, se pudieran sacar todos los resultados, tanto los del estudio de los índices métricos de la prueba como los del estudio de comparación de medias. Lo ideal sería que además de esto, se lograra con programación que nada más se vertieran los datos salieran automáticamente los índices que se necesitan e incluso se realizara el informe que luego se envía al profesorado.
- Utilizar como instrumentos de evaluación los **Tests Adaptativos Informatizados** (Holgado, Chacón, Ruiz, López y Barbero, en elaboración). Hasta ahora, se solicitó a cada profesor que elaborara su propia prueba de conocimientos; lo que en un futuro se pretende es que estas pruebas ya existan y que, además, sean tests adaptativos informatizados (TAI). Con este formato, a

los participantes se les presentaría en ordenador una prueba consistente en una serie de ítems a los que habría que responder. Detrás de este formato familiar, se escondería un banco de ítems de los que el propio programa iría escogiendo en función de los aciertos y errores del participante. En definitiva, con pocas preguntas se conseguiría afinar mucho respecto al nivel de conocimientos previos y adquiridos posteriormente, ya que la prueba se adaptaría al nivel que va mostrando quien responde.

Algunas ventajas se lograrían con este nuevo instrumento, entre las que destacan la agilización del proceso pues los resultados se obtendrían en el momento, y la comodidad para el profesorado, pues no tendría que crear una prueba para cada acción formativa ni modificarla cuando el estudio métrico alertara de esta necesidad.

Ya se están almacenando ítems para algunos cursos de ofimática. Aunque aún queda por hacer, se espera que la puesta en marcha de este instrumento no se demore en exceso.

4. Para controlar algunas variables que pudieran estar influyendo en la medida de los resultados, se espera incluir en el proceso de evaluación a posteriori algunos avances:

- La realización de un **diseño cuasiexperimental de grupo control no equivalente y prueba previa y posterior**. Consistiría en conformar un grupo de potenciales usuarios beneficiarios de las acciones formativas pero que no asistiera a éstas (grupo control) y solicitarles su participación en la realización de las mismas pruebas previas y posteriores. Con esta práctica se conseguiría, por ejemplo, un control de posibles efectos de la historia. Este diseño, representado en la figura 8.4, es uno de los más difundidos en la investigación educativa (Shadish, Cook y Campbell, 2002b; Chacón, Shadish y Cook, en prensa).

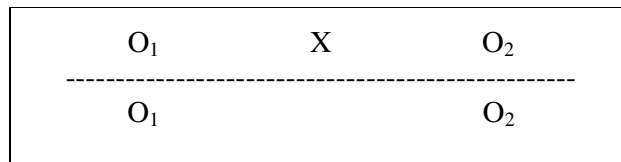


Figura 8.4. Diseño con grupo control no equivalente y prueba previa y posterior.

Al igual que para la figura 8.3, “O₁” representa la medida previa; “O₂”, la posterior; y “X”, la acción formativa. La línea discontinua que separa ambos grupos simboliza que son no equivalentes; es decir, que ambos grupos no fueron conformados aleatoriamente.

Este diseño aún mejoraría más si se pudieran tomar varias medidas previas y varias posteriores, para tener mayor certeza de la consistencia del dato.

- La realización de un **meta-análisis** para obtener un dato global que incluya el tamaño de efecto alcanzado por el plan de formación en su totalidad y que permita conocer posibles variables moduladoras que estén facilitando o

entorpeciendo el proceso de formación y la consecución de los objetivos previstos.

En **resumen**, se puede afirmar que la evaluación de conocimientos es útil porque supone dar un paso más en pro de obtener información objetiva y fidedigna a partir de la cual decidir si las acciones formativas cumplieron los objetivos que con ellas se trataron de alcanzar.

Pero interesa dar más pasos en el sentido de verificar, sobre todo, si lo que se aprendió en la acción formativa se puso en práctica en el puesto de trabajo. Si no se aprendió lo suficiente, es imposible que la puesta en práctica en la realidad laboral fuera idónea. Sin embargo, se puede dar el caso de que se produjera un buen nivel de aprendizaje, pero la transferencia al puesto de trabajo no se diera adecuadamente. Es por ello que el estudio del nivel de transferencia constituye una de las fases de evaluación clave en el proceso de evaluación.

2. EVALUACIÓN DE LA TRANSFERENCIA.

2.1. Introducción.

Hasta ahora, respecto a resultados, se ha evaluado la satisfacción de los participantes con las acciones formativas en las que han participado y el nivel de conocimientos adquiridos.

El siguiente paso a desarrollar en la evaluación de resultados es la valoración de la transferencia, que consiste en analizar el grado en que los participantes en las acciones formativas transfirieron los conocimientos, habilidades o actitudes adquiridas a sus respectivos puestos de trabajo, y en qué medida esto sirvió para mejorar su desempeño en el puesto de trabajo (Tharenou, 1991; Broad y Newstrom, 1992; Villar, De Vicente, Gallego, León, Molina, Moral y Pérez, 1996; Lara, Díaz y Chacón, 2002; Chacón y López, 2003).

Esta evaluación deriva de la necesidad de que las acciones formativas se ajusten a las demandas específicas de los puestos de trabajo ya que, en principio, tiene poca utilidad una acción formativa que provoque un alto grado de satisfacción en los participantes y donde éstos presenten un alto nivel de adquisición de conocimientos, pero tras la cual se olviden los contenidos o no se pongan en práctica en el lugar de trabajo (Chacón, S., Holgado, F. P., López, J. M., Sanduvete, S., 2006).

Los **objetivos** en esta fase de evaluación de la transferencia son los siguientes:

- Valorar el grado de **utilidad** de las acciones de formación para la mejora en el desempeño de las tareas en el puesto de trabajo por parte de los participantes.
- Evaluar la **eficacia**; es decir, determinar si se cumplieron los objetivos que perseguía la acción formativa y si se dio un **cambio de conducta** en el puesto de trabajo a partir de la formación, que conllevara un aumento de la calidad de las

tareas realizadas.

- Evaluar la **eficiencia**; es decir, si se dieron **mejoras a nivel organizacional** que, en el caso de instituciones públicas, se traducen en un aumento de la calidad de los servicios que se prestan a la sociedad (Arvey y Cole, 1989).
- Proponer **mejoras en las acciones formativas** para subsanar las deficiencias detectadas.

En la literatura, respecto a los **instrumentos** a emplear para la recogida de información en este nivel de evaluación se recomienda la encuesta, la entrevista y el registro de conducta principalmente (Chacón, Sanduvete, Holgado y Sánchez, 2006).

En cuanto a las **técnicas de análisis de datos**, se recomiendan las mismas expuestas para los dos primeros niveles (frecuencias, medidas de tendencia central, medidas de variabilidad, etc.), junto con el uso de medidas de asociación (correlación, principalmente) con objeto de medir la posible asociación entre variables relacionadas con la formación (Basarab y Root, 1992).

En Diputación de Sevilla y en el IAD, la evaluación de la transferencia no se ha puesto aún en marcha, aunque el diseño ya está elaborado; en el Centro de Formación y Perfeccionamiento del PAS, es un aspecto que se evalúa desde el año 2006. A continuación, se pasa a describir método y resultados de la evaluación de la transferencia acaecida en esta última organización durante el año 2006 (Unidad de Evaluación del Centro de Formación y Perfeccionamiento del PAS de la USE, 2006a). Hay que destacar que fue un estudio piloto ya que, al ser la primera vez que se desarrolló este tipo de evaluación, sirvió como prueba, para ver su aceptación entre los participantes en las acciones formativas y para estudiar posibles mejoras que realizar para próximas ocasiones.

2.2. Método.

2.2.1. Muestra.

La **población potencial** para este estudio fueron todas las personas que participaron en alguna acción formativa desde enero hasta noviembre de 2006; concretamente, la cifra ascendió a 1800. La **muestra** fue de 1089, lo que supuso la participación del 60.5% respecto al total de posibles participantes.

2.2.2. Instrumentos.

El instrumento principal de evaluación fue un **cuestionario** de formato semicerrado. En el anexo XIV (pág. lxxxii) puede verse el utilizado en el Centro de Formación del PAS (instrumento 1) y el que será utilizado en Diputación una vez se implemente la evaluación de la transferencia (instrumento 2), además de otros tantos posibles modelos que también responden a las principales dimensiones para evaluar la transferencia recogidas a nivel nacional e internacional según un estudio bibliométrico elaborado acerca del tema (Chacón, Holgado, López y Sanduvete, 2006).

El cuestionario que será utilizado para la evaluación de la transferencia en Diputación de Sevilla una vez se ponga en marcha ésta (anexo XIV, instrumento 2, pág. lxxxiv) consta, al inicio, de datos de identificación que se cumplimentan automáticamente (vienen dados por el programa, por lo que no han de ser escritos por el participante); a continuación, se presentan seis ítems que se responden con una escala de valoración de cinco puntos, referidos a la aplicación de lo aprendido en la acción formativa al puesto de trabajo; posteriormente, un ítem de opción múltiple (en el que se puede marcar más de una opción) referido a las razones por las que la persona hizo el curso; y un último apartado de formato abierto para posibilitar la anotación de observaciones.

El cuestionario con el que se evaluó la transferencia en el Centro de Formación y Perfeccionamiento del PAS (anexo XIV, instrumento 1, pág. lxxxiii), presenta tres cuestiones cerradas que se responden con una escala de valoración de cinco puntos, acerca de la aplicación en el trabajo de lo aprendido en la acción formativa, la mejora en los resultados de este trabajo y la disposición de los recursos necesarios para poder aplicar lo aprendido; además, presenta una última pregunta abierta para cualquier otro comentario que se considere oportuno.

Se necesitó además acceso a **Internet**, puesto que por esta vía se difundió y recogió la información. A modo de ejemplo en el instrumento 2 del anexo XIV (pág. lxxxiv), en los dos primeros párrafos, se presenta textualmente el **correo electrónico** que se va a enviar cuando se inicie esta evaluación en Diputación de Sevilla. Sencillamente, se solicitará al participante de la acción formativa que cumplimente el cuestionario que se enviará adjunto, se le agradecerá su participación y se le facilitará el enlace a través del cual acceder al cuestionario y enviar las respuestas.

Como vía alternativa, se utilizó también el **teléfono** para la difusión y recogida de la información de aquellas personas que no tenían acceso a Internet.

Para la recogida de los datos derivados del cuestionario, se utilizó una **base** donde se almacenaron automáticamente las respuestas de los participantes a los diferentes ítems del cuestionario, así como otros aspectos esenciales para la identificación y descripción de la persona; concretamente, el código de la edición del curso, la fecha en que se llevó a cabo, el nombre y los apellidos, el DNI, la fecha de nacimiento, el puesto de trabajo concreto, la subunidad, la unidad, el centro destino, la situación administrativa, el turno, el grupo, el nivel y la titulación.

Por último, el análisis de la información recogida se llevó a cabo con el software **SPSS (v. 13.0)**.

2.2.3. Procedimiento.

El primer paso fue el **envío del aviso de la disponibilidad del cuestionario** cuatro meses después de finalizada la acción formativa a evaluar, transcurrido por tanto un tiempo prudencial una vez concluida ésta (Ortí, 1994; Bou, Marro y Pineda, 1999; Fernández Salinero, 1999; Salanova y Grau, 1999; Julià, 2001), para así dar tiempo suficiente como para que los participantes tuvieran la posibilidad de consolidar y poner en práctica lo aprendido, teniendo por más seguro que el uso se mantendría a lo largo

del tiempo y no se debía a un primer momento de ilusión rápidamente precedido (Chacón, Holgado, López y Sanduvete, 2006).

Se dejaron tres semanas para la **recepción de los cuestionarios**. En el correo enviado, se facilitó a los participantes un enlace en el que podían encontrar el cuestionario de evaluación. Marcando el botón “enviar”, los datos se almacenaban directamente en la base de datos creada para tal efecto.

Para aquellos casos en los que no se conocía el correo electrónico del participante o éste no tenía acceso fácil a Internet, se realizó el contacto por **teléfono** y se le solicitó la respuesta a cada uno de los ítems del cuestionario.

Una vez cerrado el plazo de recogida de información, se procedió al **análisis de los datos** obtenidos. Las medias de las tres preguntas cerradas y sus figuras correspondientes fueron calculadas automáticamente. Otros estudios adicionales fueron realizados con el programa SPSS 13.0.

Finalmente, se realizó un **informe** para cada acción formativa donde se mostraron las conclusiones obtenidas a partir del contraste de los datos cualitativos con los cuantitativos.

Todos los informes están disponibles en la página web junto con los resultados de satisfacción y conocimiento, con lo que toda persona interesada puede de manera fácil acceder a la información.

Por último, se mantuvo abierta una **vía de comunicación** con la unidad de formación y de evaluación por si a alguien se le planteaba alguna duda o comentario.

2.3. Resultados.

A continuación, en la tabla 8.4, se presentan los resultados obtenidos en las preguntas de formato cerrado del cuestionario de transferencia.

Ítem	N	Media
1. ¿Aplica en su puesto de trabajo lo que aprendió en el curso?	1086	3.42
2. ¿Cree que los resultados de su trabajo han mejorado por haber participado en el curso?	1083	3.30
3. ¿Dispone en su entorno de trabajo de los recursos y condiciones necesarios para poder aplicar la formación?	1080	3.47

Tabla 8.4. Resultados globales de las preguntas de formato cerrado del cuestionario de transferencia en el Centro de Formación del PAS (año 2006).

Estas mismas medias se presentan gráficamente en la figura 8.5.

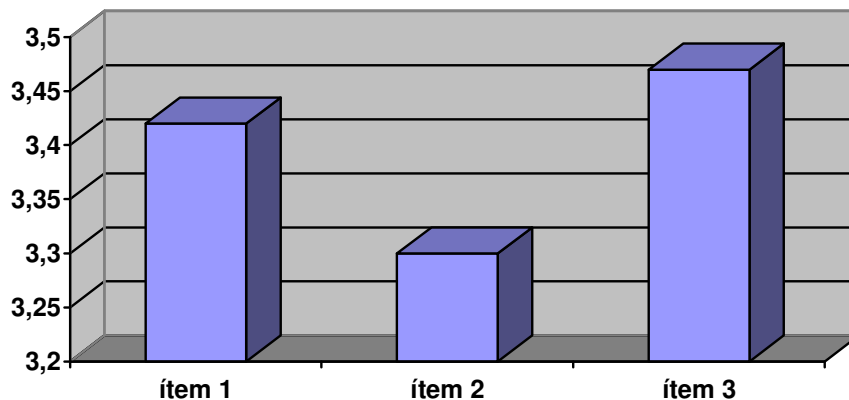


Figura 8.5. Medias obtenidas en la evaluación de la transferencia en el Centro de Formación del PAS en el año 2006.

Tras estos resultados se puede observar que, en los tres aspectos estudiados (aplicación en el puesto de trabajo de lo aprendido, mejoras del desempeño en el puesto de trabajo tras la acción formativa y disponibilidad de los recursos y condiciones necesarios para aplicar lo aprendido en el puesto de trabajo), las puntuaciones obtenidas fueron mediocres, estando entre el 3 y el 3.5.

De las tres cuestiones, el ítem mejor valorado fue el 3 referido a la disponibilidad de los recursos y condiciones necesarios para aplicar lo aprendido en el puesto de trabajo, mientras que el peor valorado fue el ítem 2 referido a las mejoras en el desempeño del trabajo debido a la acción formativa.

2.4. Discusión.

A continuación, se presentan los puntos fuertes detectados en el proceso de evaluación de la transferencia de los conocimientos adquiridos en las acciones formativas a los puestos de trabajo, las debilidades y posibles mejoras para fortalecer dichos puntos débiles en posteriores evaluaciones.

Los **puntos fuertes** más destacables fueron los siguientes:

1. **Alta participación de usuarios/as.** Fue una grata sorpresa el constatar que más de la mitad de todos los posibles usuarios participaron en la evaluación de la transferencia, sobre todo teniendo en cuenta que era la primera vez que esta evaluación se ponía en marcha; que no era un proceso totalmente anónimo, método al que estaban acostumbrados hasta ahora los participantes; y que el envío y recogida de la información se hicieron a distancia, medio en el que solemos encontrar un número elevado de abstenciones.

Esta buena aceptación puede deberse a la metodología seguida a la hora de instaurar esta nueva dinámica al intentar, por un lado, que fuese un proceso dinámico, participativo y lo menos disruptivo posible para los participantes con la menor intrusión posible en la dinámica habitual de trabajo; y al tratar, por otro lado, de que el participante fuera conocedor de qué es la evaluación de la transferencia y

de los beneficios que puede reportarle, con lo que se animaría a invertir un tiempo mínimo en cumplimentar una encuesta accesible, rápida y fácil.

2. **Puntuaciones medias en transferencia.** En una escala del 1 al 5, los tres ítems de formato cerrado obtuvieron un valor medio entre el 3 y el 3.5. Aunque podría mejorarse (en satisfacción suelen obtenerse medias más elevadas), no puede decirse que sea un mal resultado, teniendo en cuenta que la evaluación de este constructo no ha hecho más que ponerse en marcha. Con las posibles mejoras detectadas relacionadas con el fomento de la transferencia del conocimiento adquirido a los puestos de trabajo, se espera encontrar mejoras en las valoraciones obtenidas en próximos años.
3. **Identificación de los participantes.** Los datos que se habían recogido hasta el momento en satisfacción y conocimientos eran totalmente anónimos. Con el nuevo procedimiento de recogida de información acerca de la transferencia de lo aprendido al puesto de trabajo, se logró obtener datos personales y profesionales de los participantes. Con esta información adicional, se abre la posibilidad de realizar análisis en mayor profundidad, como la correlación entre variables para detectar factores influyentes en el éxito de las acciones formativas, como podrían ser la formación previa o los años de experiencia.

Esta identificación resulta crucial ya que con ella se puede comenzar a pensar en posibles criterios de asignación a las distintas acciones formativas, en la evaluación del impacto y de la rentabilidad, etc.

Esto ha supuesto afinar más en una evaluación de la formación que, hasta ahora, había tenido tintes genéricos, por lo que se perdía mucha información y capacidad de análisis.

Además de fortalezas, también se detectaron **puntos débiles** a solventar, como los que a continuación se enumeran:

1. **Problemas para la puesta en marcha de la evaluación de la transferencia.** Durante muchos años, la evaluación de la transferencia estaba ya planificada pero no se llegaba a implementar. De hecho, en Diputación de Sevilla y en el IAD aún no se ha llevado a cabo. Las mayores barreras fueron **trabas institucionales**, muchas de las cuales no estaban directamente relacionadas con la formación.
 - En ocasiones, los estudios en transferencia, además de por problemas contextuales, suelen ser abortados por los mismos **responsables de los organismos** por provocar más problemas que soluciones, como pueden ser reivindicaciones sindicales, el rediseño de puestos de trabajo, la petición de recursos no previstos en los que no se puede o no se quiere invertir, etc.
 - Además, forzar la situación, incitar hacia la evaluación de la transferencia cuando la organización y su personal aún no estaban preparados, podría haber llevado a los participantes a ver la actuación de los evaluadores como una intrusión. Esta **reactividad negativa** en los usuarios podría haber mermado futuros esfuerzos para la puesta en marcha de la evaluación de la transferencia.

2. **Diseño con debilidades.** El diseño que hasta ahora se ha utilizado es tan simple como la recogida de una sola medida al final de la acción formativa pasado algún tiempo. Esto presenta algunos problemas, como el hecho de que no hay punto de comparación ni previo (para determinar si hubo cambios desde antes de la acción formativa hasta cuatro meses después de finalizada ésta), ni posterior (para estudiar si las mejoras se mantuvieron en el tiempo o si por el contrario fueron sólo cambios perecederos).
3. **Análisis de datos a nivel superficial.** El análisis de datos hasta ahora realizado en la evaluación de la transferencia fue meramente a nivel descriptivo, con el conteo del número de respuestas a los ítems de formato cerrado, el cálculo de medidas de tendencia central y no mucho más. Para mejorar en este punto, habría que plantearse el desarrollo de otros análisis de interés.
4. **Mejora del instrumento de evaluación.** Aunque el cuestionario utilizado para evaluar la transferencia se elaboró recopilando información acerca de los métodos de evaluación de este constructo utilizados a nivel nacional e internacional recogidos en la bibliografía; y si bien es cierto que se contó para su construcción con la colaboración de expertos en formación, este instrumento presenta alguna que otra debilidad, como el hecho de que, hasta ahora, nunca se estudiara su validez.

Para tratar de fortalecer los puntos débiles encontrados, se presentan a continuación las siguientes **propuestas de mejora** (Campos, Ruiz, Sánchez y Chacón, 2007), que se espera puedan ser implementadas en un breve periodo de tiempo:

1. **Puesta en marcha** del proceso de evaluación de la transferencia en Diputación de Sevilla y **consolidación** en el Centro de Formación y Perfeccionamiento del PAS. Una vez se logró dar el primer paso hacia la evaluación de la transferencia con el estudio piloto desarrollado en el Centro de Formación y Perfeccionamiento del PAS, las siguientes directrices a seguir habrían de ser consolidar la evaluación de este constructo en esta organización y tratar de implementarla por primera vez en las otras organizaciones.

Aunque es cierto que, entre los participantes del Centro de Formación del PAS, la evaluación de esta fase obtuvo una buena acogida, se podrían pensar estrategias para aumentar aún más la participación. Como aún no se conoce qué respuesta habrá en Diputación de Sevilla, estas **recomendaciones** también podrían venir bien a la hora de poner en marcha la evaluación en esta organización:

Una opción es tratar de que el **cuestionario** sea lo más **rápido y sencillo** posible de cumplimentar. Por esta razón, se pretende que simplemente marcando con el ratón en un vínculo se acceda al instrumento en sí; que los datos de identificación del curso se muestren automáticamente para que el participante no tenga que cumplimentarlos; y que, para responder a los ítems con escala de valoración sólo haya que marcar con el ratón, de tal manera que sólo se tenga que utilizar el teclado y escribir en caso de que se quiera aportar alguna observación en el último ítem. El cuestionario puesto en uso en el Centro de Formación y Perfeccionamiento del PAS, además, sólo presenta 4 ítems (3 cerrados y uno abierto), con lo que su cumplimentación es tremendamente rápida; el cuestionario propuesto para Diputación de Sevilla, contando los ítems de formato abierto y

cerrado, tiene 8 (el doble). Si en Diputación la acogida de la evaluación y la participación no son buenas, habría que tener en cuenta esta diferencia como posible factor influyente.

Otra posibilidad podría ser dar **feedback** lo más **inmediatamente** posible, para que los participantes vean que realmente su opinión se ha tenido en consideración y, en la próxima ocasión que les surja, vuelvan a aportar la información requerida.

Una última alternativa sería, en la medida de lo posible, fomentar el **canal telefónico** para la recogida de información ya que, aunque para el evaluador sea un proceso más costoso, es posible que al participante le sea más cómodo responder a las preguntas verbalmente, sobre todo para aquellos con dificultades para acceder a Internet.

2. **Variación del diseño:** Como ya se comentó, el diseño hasta ahora utilizado en la evaluación de la transferencia es bastante pobre y conlleva algunas debilidades. Por ello, se propone modificar el número de momentos de medida y pasar, de una evaluación sólo posterior, a un estudio de seguimiento a lo largo del tiempo con tres momentos de recogida de datos: al inicio de la acción formativa, tres meses después de la finalización de ésta y seis meses después. El nuevo diseño, por tanto, se representa a partir de la siguiente figura:



Figura 8.6. Diseño con medida previa y múltiples medidas posteriores.

Cada “O” simboliza una observación distinta en diferentes momentos temporales, mientras que la “X” representa una intervención (en este caso, la ejecución de una acción formativa).

Con este diseño, se logra tener un punto de referencia previo a la realización de la acción formativa (representado por O₁) y dos posteriores diferenciados en tres meses el uno del otro (representados por O₂ y O₃), con lo que se puede tener mayor certeza de que hubo cambios en el desempeño del puesto de trabajo por la acción formativa y que dicho cambio se dio de manera consistente a lo largo del tiempo.

Pero este diseño aún presenta posibles amenazas a la validez por lo que se propone, como mejora en un futuro algo más lejano, la implementación de un diseño cuasiexperimental de grupo control no equivalente con medida previa y posterior (Gaines y Robinson, 1998). Como ya se explicó para la evaluación de conocimientos, se trataría de comparar a un determinado grupo del personal que haya participado en algún programa de formación (grupo experimental) con otro grupo del personal de similares características (con mismos jefes, áreas organizativas, funciones, nivel de experiencia, edad o educación) que no hayan participado en el mismo (grupo control).

A pesar de que este tipo de diseño nos permite obtener ciertas garantías sobre la transferencia efectiva de la formación al puesto de trabajo (por ejemplo, reduce el efecto de la historia), es necesario considerar por un lado que, además de

la formación, existen otras muchas variables que pueden estar afectando al rendimiento en el trabajo; y por otro lado, que en la realización de estos diseños también se pueden dar distintas circunstancias que amenazan o ponen en duda la conclusión sobre la eficaz transferencia de la formación al puesto de trabajo, o del aumento de la calidad institucional entendida desde la mejora en la gestión. Entre estas amenazas destacan la instrumentación, la regresión estadística y la historia focal diferencial (Shadish, Cook y Campbell, 2002b).

3. Con la recogida de la información adicional acerca de características personales y profesionales y con la identificación de los participantes, se abrió la posibilidad de llevar a cabo estudios de correlación entre variables, por ejemplo, para conocer factores que aumentan las probabilidades de éxito en la transferencia.

Y si esta identificación se aplicara también a las demás fases de evaluación (detección de necesidades, evaluación durante la acción formativa, satisfacción y conocimientos), las posibilidades de análisis aumentarían aún más, además de que se dispondría de una visión holística del proceso formativo de cada empleado individualmente y como grupo.

Además, si se llegara a implementar el primer diseño propuesto donde se recogen varias medidas a lo largo del tiempo (una antes de la acción formativa, otra a los tres meses después de finalizada y la última a los seis meses), se podrían realizar **análisis** de tendencia.

4. Para mejorar el cuestionario utilizado para recoger la información, se propone un **estudio de validez de contenido**. La necesidad de este estudio viene dada por el hecho de que no se encontró en la bibliografía un instrumento validado para evaluar transferencia. Esto llevó a la elaboración de un instrumento propio partiendo de los encontrados en la literatura, pero se desconoce si es **exhaustivo** (si mide todos los aspectos fundamentales referidos a la transferencia), **relevante** (si no se añadieron ítems que no eran de interés) y **representativo** (si todos los ítems son útiles para medir el constructo). Es para responder a estas cuestiones, por lo que se propone realizar la validez de contenido del instrumento elaborado.

El **objetivo** de este estudio de validez de contenido sería por tanto validar el cuestionario utilizado para medir la transferencia de los conocimientos adquiridos en las acciones formativas al puesto de trabajo.

El **método** a seguir para conseguir este objetivo sería el siguiente:

- La **muestra** sería el mayor número de expertos posible de distintas organizaciones, considerándose como tales a aquellas personas que, teniendo a su cargo a otros trabajadores y/o dedicándose directamente a la organización de acciones formativas, están especialmente motivados por el tema de la formación y consideran que ésta es un medio eficaz para conseguir la mejora del desempeño en los puestos de trabajo de los profesionales a su cargo.
- El **procedimiento** consistiría en elaborar un cuestionario que recogiera todos los posibles ítems de evaluación de la transferencia obtenidos tras una revisión sistemática de la bibliografía disponible referida al tema y solicitar a los expertos

que valoraran de 0 a 3 cada uno de los ítems en relación a tres aspectos relevantes: representatividad, utilidad y viabilidad. Una vez recogida esta información, se procedería al análisis de los datos a través del índice de congruencia de Osterlind (1998).

- Los **instrumentos** utilizados serían el cuestionario elaborado para medir la transferencia y el programa Microsoft Access para realizar el análisis de datos.

Como **resultado**, se interpretarían los índices de congruencia obtenidos por cada ítem en cada uno de los tres aspectos valorados, considerando como óptimos aquellos que obtuvieran un valor superior a 0.5 (Osterlind, 1998) en al menos dos de las tres dimensiones estudiadas. Comparando los datos hallados con el cuestionario de partida, se incluirían aquellos ítems considerados óptimos y se excluirían los que fueron desechados.

3. CONCLUSIONES.

Con la evaluación de los conocimientos adquiridos y con la transferencia de estos conocimientos al puesto de trabajo se ha dado **un paso adelante** en la evaluación de la formación continua: generalmente, según la literatura, sólo se miden los resultados con la satisfacción de los participantes a las acciones formativas; añadiendo estas fases de evaluación, se han conseguido obtener otros datos realmente útiles para llegar a concluir de manera más fidedigna si la acción formativa fue eficaz (si provocó el cambio de conducta esperado en la persona) y eficiente (si este cambio de conducta se tradujo a beneficios a nivel organizacional).

Puede decirse que la evaluación de conocimientos tiene algo más de rodaje que la evaluación de la transferencia que apenas se acaba de iniciar pero, en ambas fases, aún **queda mucho que mejorar** y pulir: el grado de implementación en las distintas organizaciones, el nivel de participación, los instrumentos utilizados, los análisis de datos, etc., son pruebas de esta necesidad.

A pesar de esto, las ventajas obtenidas tras iniciar la evaluación en estas fases son incuestionables. Tras la obtención de resultados en satisfacción, conocimiento y transferencia, se abre la posibilidad de **rediseñar todo el proceso en general**, dependiendo de los resultados y conclusiones extraídos de las evaluaciones cada año. Probablemente, siempre habrá algún aspecto que mejorar pero, lejos de ser un indicador negativo, este hecho responde a la filosofía de la evaluación formativa, de mejora constante a lo largo del proceso. Realmente, no existe el diseño de evaluación perfecto, entre otras cuestiones, porque la realidad es cambiante y la evaluación ha de ir reciclándose y adaptándose a estos cambios.

En este capítulo, como en otros anteriores, se trató de hacer aportaciones que sirvieran para superar las limitaciones que, tras un estudio de la literatura, se concluyó que generalmente se dan en evaluación de la formación.

De este modo, se demostró que realmente la **evaluación sumativa y formativa** no están reñidas ni son contrarias, sino que se complementan: la evaluación de

conocimientos y de transferencia, al igual que la de la satisfacción, forman parte de la fase llevada a cabo una vez finalizada la acción formativa; sin embargo, se han enfocado de tal manera que sus resultados, además de responder al grado de eficacia y eficiencia del programa, aporta directrices a partir de las cuales mejorar el proceso de intervención y mejorar la evaluación en sí misma.

Además, contra la no **operativización del constructo**, aquí se trató de definir con claridad qué se entendía por evaluación de resultados y, dentro de este amplio constructo, qué papel jugaban la obtención de conocimientos y la transferencia de éstos al puesto de trabajo. Esta estructura es compartida por varios modelos de evaluación de la formación, como por ejemplo el de Kirkpatrick (1977), pero además de esta definición teórica se hizo el esfuerzo por proponer una definición empírica clara y concisa, que se consigue sobre todo al aportar instrumentos concretos para la evaluación de cada fase. Con estos instrumentos se concretó qué aspectos se incluían en cada constructo y se contabilizó empíricamente el peso de cada uno.

Por último, para aportar pruebas de la calidad del dato recogido, se realizaron diversos **estudios métricos** con los que se evaluó si los instrumentos utilizados sirven realmente para lo que se les creó y miden los constructos que se espera que midan. Además, se presentan propuestas concretas para avanzar en este sentido.

Así, respecto a la evaluación de conocimientos, se explicita el proceso por el cual las pruebas elaboradas por cada formador pasan por un análisis general y de cada ítem en particular antes de responder con los datos obtenidos por este examen si hubo diferencias significativas en el nivel de conocimientos antes y después de la acción formativa. Además, se espera seguir trabajando en la implementación de los tests adaptativos computerizados, lo que conllevaría la homogeneización del instrumento en distintas acciones formativas de misma temática y donde cada ítem del banco creado tendría un estudio de su calidad métrica por separado y la calibración de su dificultad.

Y en relación a la evaluación de la transferencia, se propuso un instrumento que, según la literatura y unos expertos en la materia, recogía las dimensiones fundamentales del constructo. Pero no conforme con esto, se propone además un estudio de validez de contenido con el que determinar con aún más certeza si cabe la calidad del dato que con tal instrumento se recoja.

Como reto para el futuro, una vez consolidadas las fases de evaluación de conocimientos y transferencia, se espera poder pasar a la **evaluación del impacto**, consistente en analizar la contribución de la formación para mejorar la calidad del servicio prestado por la organización (Gaines y Robinson, 1998). Pero no tiene sentido pasar rápidamente a esta fase de evaluación sin estar seguros de que las fases previas fueron aceptadas por los participantes, los interesados y la organización en su conjunto, que no es vista como un elemento disruptivo en el funcionamiento natural de la empresa y que las evaluaciones ya implementadas funcionan correctamente y aportan resultados de calidad.

CAPÍTULO 9. CONCLUSIONES Y DESARROLLO FUTURO.

Una de las características que diferenció al ser humano del resto de animales fue el desarrollo de la capacidad de transmitir de manera explícita y concisa su conocimiento a futuras generaciones a través de la escritura, de tal modo que este hecho marcó el paso de la prehistoria a la historia. A partir de ahí, también con ayuda de otros factores, la ciencia avanzó en pocos cientos de años más de lo que lo había hecho en miles anteriormente.

La formación continua es una expresión de esta transmisión de conocimientos, concretamente a personas empleadas adultas. Por sí misma confirma la importancia del aprendizaje a lo largo de toda la vida, y no exclusivamente en la niñez y adolescencia.

En ocasiones, este tipo de formación se ha **desvirtuado**:

En el momento en que se conoció el hecho de que se daba la posibilidad de recibir ayudas del Fondo Común Europeo, algunos vieron la ocasión de obtener beneficios económicos, llegando en ocasiones al extremo de casos de **corrupción** donde se obtuvo el dinero y, sin embargo, finalmente no se realizó ningún tipo de acción formativa. Como control de estas situaciones, se solicita desde las organizaciones encargadas de conceder las ayudas, una justificación exhaustiva de los planes de formación realizados.

Un servicio que puede llegar a ser una ocasión para aumentar las competencias personales, en algunas circunstancias sólo se aprovecha exclusivamente para fines más superficiales como la **promoción interna** (obtener puntos para las siguientes oposiciones como sea), de tal modo que la formación se convierte en un medio para lograr un fin, en lugar de un fin por sí mismo. Por esta razón, en ocasiones un profesor puede encontrarse con un alumnado sin implicación, que no participa activamente y que trata de obtener el título con el menor esfuerzo posible. Tanto el nivel de conocimiento obtenido como el grado de transferencia en estos casos suele ser muy bajo.

Sin embargo, los **beneficios** aportados superan con creces este uso inapropiado de la formación.

A nivel **personal para el empleado**, dota de **instrumentos** a los profesionales por lo que, en lugar de fomentar las ayudas puntuales de manera asistencialista, se busca un avance más profundo a partir del aumento de la eficacia en el trabajo, de un mayor número de habilidades entre las que elegir y de la disminución de la probabilidad de ser manipulados o negativamente influidos por factores externos.

Cuando la recompensa no sólo es extrínseca, cuando realmente la persona está interesada en aprender para mejorar en el desempeño de su puesto de trabajo, considera que además redunda en su desarrollo personal y que es un derecho al que no está dispuesta a renunciar, la implicación se hace máxima y la utilidad de la acción formativa aumenta.

A nivel **personal para el usuario**, se busca su **satisfacción** con el servicio prestado, por lo que indirectamente son beneficiarios de la formación que los trabajadores recibieron; por ejemplo, cuando se forma a los trabajadores de una

residencia de personas mayores, se espera que las destrezas en sus puestos de trabajo aumenten (que disminuya el estrés, que hagan sus tareas mejor y con menor esfuerzo), pero que también aumente la satisfacción de las personas mayores con la residencia y sus empleados (que se les trate correctamente, que se vean atendidas en sus necesidades, que se les permita tomar decisiones acerca de sus propias vidas, etc.).

Este hecho se hace especialmente relevante en aquellas ocasiones en que la vida de otras personas depende de la eficacia de los profesionales. Es el caso, por ejemplo, de la formación dirigida a bomberos, cuya función en situaciones de crisis consiste en rescatar a personas; o el caso de personas que trabajan con inmigrantes, gente que en ocasiones lo han perdido todo, que tienen que empezar una nueva vida y que tiene, como única vía para poder permanecer en el país de manera legal y con un empleo digno, el trabajo en conjunto con el profesional.

A nivel **organizacional**, la formación se concibe como una **inversión** que aportará beneficios a medio y/o largo plazo. Cada vez más empresas están optando por el reciclaje y la formación continua de sus empleados porque, en este mundo constantemente cambiante, o te renuevas o mueres. Para ser competente y competitivo al día de hoy, es necesaria la mejora continua en todo momento.

Este trabajo se realizó fundamentalmente con la intención de aportar propuestas concretas para tratar de mejorar las acciones dirigidas a formar a los trabajadores. Por ello, lo primero que se hizo fue buscar limitaciones que podrían estar influyendo negativamente en la calidad de la formación a partir del estudio bibliográfico sobre modelos teóricos, la elaboración de la escala de calidad y la aplicación de dicha escala a estudios de formación continua. El meta-análisis aportó una prueba más de la influencia de algunas de las variables incluidas. A continuación, se explicita más detalladamente cada uno de los pasos seguidos en busca de tales limitaciones:

Tras el estudio bibliográfico, se constató que existe una gran cantidad de **modelos de evaluación** de la formación, algunos con muchos aspectos coincidentes entre sí, pero muy pocos prácticos y aplicables, puesto que no concretan lo teórico con propuestas empíricas. Por eso, en este trabajo se ha tratado de especificar las distintas fases de evaluación y de aportar una descripción clara del procedimiento seguido y de los instrumentos utilizados: con la intención de que fuera posible su uso en otras organizaciones.

Para determinar qué factores influyen en la calidad de un programa, a través de un proceso deductivo-inductivo se obtuvo una versión integradora de una **escala** para medir la calidad metodológica de estudios primarios.

Este instrumento aportó una posible **definición empírica del concepto calidad**, que hoy día es muy utilizado, pero en pocas ocasiones definido con detalle, sino que más bien se suele dar una definición conceptual general. Además, este concepto en muchas ocasiones está excesivamente dirigido a los resultados, a la evaluación sumativa, sin considerar el proceso; sin embargo, la definición aportada en este trabajo tiene también en cuenta aspectos relevantes relacionados con la evaluación formativa.

Hay que tener en consideración que este instrumento no pretende ser algo definitivo e incuestionable porque, por ejemplo, dependiendo de la temática concreta de

los programas en estudio, podrían variar algunas categorías. Así, en referencia al tamaño de la muestra, puede que el número ideal de participantes para educar en valores sea mayor que el número ideal para aprender a realizar cuidados paliativos.

Además, su aplicación en mayor número de ocasiones posibilitará la realización de modificaciones como adaptaciones a casos distintos a los hasta ahora encontrados; y posibilitará la contrastación con los resultados de otras escalas.

La escala se aplicó concretamente a los **estudios de formación continua**. De este modo, se detectaron los aspectos en los que más se flaqueaba según la literatura y se plantearon posibles propuestas para reforzarlos.

Finalmente, con un **estudio meta-analítico** se trató de constatar qué variables de las detectadas realmente ejercían influencia en la eficacia de la formación. Sin embargo, los resultados obtenidos se tomaron con cierta cautela porque fue un número muy bajo de estudios el que cumplió los criterios de inclusión.

Para solventar las limitaciones encontradas en la mayoría de los estudios de formación continua en la práctica de evaluación en las distintas organizaciones, en la tabla 9.1 se explicitan las actuaciones ya realizadas y las acciones planificadas para el futuro.

La primera columna presenta la limitación a solventar. A continuación, cada fase de evaluación muestra dos columnas: en la de la izquierda, aparecen las mejoras realizadas para cubrir las limitaciones encontradas; en la columna de la derecha, las mejoras que tratarán de implantarse en el futuro.

LIMITACIONES	EVALUACIÓN DE RESULTADOS									
	DET. NECESIDADES		EVAL. DURANTE		EVAL. SATISFACCIÓN		EVAL. CONOCIMIENTOS		EVAL. TRANSFERENCIA	
	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar
Sin modelo teórico explicitado	Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfec. del PAS, 1999 Chacón, Lara y Pérez, 2002c, adaptado de Witkin y Altschuld, 1995	---	Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfec. del PAS, 1999	---	Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfec. del PAS, 1999; Kirkpatrick, 1999	---	Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfec. del PAS, 1999; Kirkpatrick, 1999	---	Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfec. del PAS, 1999; Kirkpatrick, 1999	Futura implantación de evaluación del impacto
Evaluación sumativa	Evaluación antes de intervención	---	Evaluación durante intervención	Consolidar	Detección de mejoras	---	Detección de mejoras	---	Detección de mejoras	---
Sin operativización de variables	Dada por el contenido del instrumento	---	Los ítems recogidos marcan qué se espera de coordinación y profesorado	---	Constructo con 3 factores	---	Dada por el profesor de cada acción formativa	TAIs	Dada por el contenido del instrumento	---
Instrumentos sin estudios de fiabilidad y validez	---	---	---	Validez de contenido de cuestionario y checklist	Validez del modelo de medida; fiabilidad, discriminación y precisión de medida	Invarianza	Estudio de fiabilidad, discriminación, dificultad	Homogeneidad con TAIs	---	Validez de contenido

Tabla 9.1. Limitaciones encontradas, mejoras ejecutadas y otras mejoras planificadas para el futuro.

LIMITACIONES	EVALUACIÓN DE RESULTADOS									
	DET. NECESIDADES		EVAL. DURANTE		EVAL. SATISFACCIÓN		EVAL. CONOCIMIENTOS		EVAL. TRANSFERENCIA	
	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar	Mejoras realizadas	Mejoras por realizar
Poca concreción de cómo se llevó a cabo la intervención	Descripción detallada	---	Descripción detallada	---	Descripción detallada	---	Descripción detallada	---	Descripción detallada	---
Poca concreción de las características de la muestra	Descripción lo más detallada posible	Participantes identificados	Descripción lo más detallada posible	Participantes identificados	Descripción lo más detallada posible	Participantes identificados	Descripción lo más detallada posible	Participantes identificados	Descripción lo más detallada posible	Participantes identificados
Sin grupo de comparación	---	---	---	---	Comparación entre años	---	Comparación entre años	Diseño cuasiexperimental de grupo control no equivalente y prueba previa y posterior	---	Comparación entre años
Sin azar	Posibilidad de participación de toda la población	---	Posibilidad de participación de toda la población	---	Posibilidad de participación de toda la población	---	Posibilidad de participación de toda la población	---	Posibilidad de participación de toda la población	---
Mortalidad global elevada	No procede	---	No procede	---	No procede	---	---	Inferir valores perdidos	---	
Medida en un solo momento y sin seguimiento	Medida en cualquier momento	Relacionar con otros momentos de medida	Medida durante	Relacionar con otros momentos de medida	Medida después	Relacionar con otros momentos de medida	Medidas previa y posterior	Relacionar con otros momentos de medida	Medida después	Una medida previa y dos posteriores
Análisis de datos a nivel descriptivo	Estudio de métodos de escalamiento (priorización)	Triangulación y correlación		Triangulación y correlación		Triangulación y correlación	Comparación entre medias	Tamaño de efecto y correlación		Análisis de tendencia, triangulación y correlación

Tabla 9.1 (continuación). Limitaciones encontradas, mejoras ejecutadas y otras mejoras planificadas para el futuro.

Sin modelo teórico explicitado

Se encontró que en pocas ocasiones las evaluaciones de la formación halladas en la literatura presentaban un modelo teórico que guiara y justificara las actuaciones realizadas. La propuesta de evaluación aquí presentada se basa en tres modelos:

1. El modelo de evaluación no lineal (Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999), plantea un bucle evaluativo que comienza con la detección de necesidades, sigue con la evaluación durante la acción formativa y continúa con la evaluación de resultados. No podría decirse que la evaluación acabara aquí porque, con toda la información recopilada, además de intervenir en el momento para mejorar, se propondrían mejoras para futuros planes de formación y se volvería a empezar con la evaluación de necesidades, pero incluyendo estas mejoras en las fases correspondientes.
2. El modelo de Chacón, Lara y Pérez (2002c), adaptado de Witkin y Altschuld, (1995), para la etapa concreta de detección de necesidades, que diferencia tres fases distintas: pre-evaluación, evaluación principal y post-evaluación.
3. El modelo de Kirkpatrick (1999) para la evaluación de resultados, donde se diferencian cuatros niveles. De menor a mayor profundidad, son la evaluación de la satisfacción, de los conocimientos adquiridos, de la transferencia al puesto de trabajo y del impacto. Generalmente, la evaluación de la satisfacción es la más utilizada; en el caso de la evaluación en este trabajo presentada, si bien es cierto que ésta es la fase más consolidada, también pudieron ponerse en marcha las dos siguientes de evaluación de conocimientos y de transferencia.

En el futuro, se intentará implementar la evaluación del impacto (cuarto y último nivel de evaluación de resultados), no sin antes haber consolidado la evaluación de conocimientos y transferencia.

Evaluación sumativa

Se constató que el tipo de evaluación más comúnmente realizado era la evaluación sumativa o de resultados. En este trabajo, siguiendo como previamente se comentó el modelo de evaluación no lineal (Anguera y Chacón, en prensa, b; Alvira, 1991; Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla, 1999), se llevó a cabo una evaluación formativa o de proceso, con recogida de datos antes de la intervención (detección de necesidades), durante dicha intervención y tras ésta. Además, incluso se trató de que la evaluación de resultados adquiriera tintes de evaluación formativa puesto que se dirigió, por encima de responder al nivel de eficacia del programa, a determinar qué factores concretos podrían mejorarse y cómo.

Como mejora futura, se espera consolidar la fase de evaluación durante la acción formativa a través del fomento de la participación y de un análisis más profundo de los datos cualitativos obtenidos, además de la relación de estos resultados con los encontrados en las otras fases de evaluación.

Sin operativización de variables

Se constató que otra limitación comúnmente encontrada en los programas de formación continua era la falta de operativización de las variables en estudio. Esta limitación trató de superarse principalmente con la aportación de instrumentos concretos en los que se explicitaron los aspectos considerados relevantes para cada constructo.

Así por ejemplo, en detección de necesidades, con el instrumento utilizado se puede vislumbrar que el constructo “necesidad” está compuesto por los conceptos clave “objetivos”, “contenidos” o “destinatarios”; “satisfacción” sería un constructo conformado por tres factores: “objetivos y contenidos” “metodología y ambiente” y “utilidad y valoración global”. Y el constructo “transferencia” estaría constituido por el nivel de aplicación de lo aprendido, el grado de facilitación del entorno para dicha aplicación y las mejoras en el trabajo. Lo mismo ocurre con los instrumentos para evaluar durante la intervención: uno de ellos coincide básicamente con los elementos tenidos en cuenta en la evaluación de la satisfacción; el otro fue diseñado a partir de los manuales del formador y del coordinador.

En la evaluación de conocimientos, como los instrumentos son elaborados por los profesores que imparten la acción formativa, los constructos son definidos por ellos, aunque más adelante modifiquen dichos instrumentos en función de sus cualidades métricas.

Una mejora para el futuro en este sentido es el uso de los TAI (Tests Adaptativos Computerizados), con lo que se conseguiría obtener un constructo común para cada temática de formación.

Instrumentos sin estudios de fiabilidad y validez

Otra limitación extendida en las intervenciones de formación continua es que los instrumentos de evaluación no presentan comprobación de su fiabilidad y validez.

Para solventar esta carencia, instrumentos concretos se presentaron como medio para recoger datos en las distintas fases de evaluación y, en la mayoría de ellos, se realizó algún estudio de fiabilidad y validez:

- El cuestionario de satisfacción obtuvo índices óptimos en fiabilidad, precisión y discriminación. Además, se demostró empíricamente la validez del modelo utilizado.
- Las pruebas de medición del conocimiento adquirido son sometidas a un estudio métrico de fiabilidad, discriminación y dificultad y se proponen mejoras concretas para próximas ediciones.
- El cuestionario de transferencia se elaboró a partir de una revisión bibliográfica y la consulta a expertos en la materia.

Como mejoras para el futuro, se espera profundizar en estudios de fiabilidad y validez en diferentes instrumentos:

- Respecto al cuestionario de satisfacción, se espera realizar un estudio de la invarianza para sopesar la posibilidad de generalizar el uso de este instrumento a otras organizaciones.
- Se tiene pensado realizar estudios de validez de contenido a los instrumentos utilizados para la evaluación durante la implementación del programa (el cuestionario y el listado de verificación) y al cuestionario de transferencia.
- En relación a la evaluación de conocimientos, el uso de los TAIs permitirá la homogeneización de los contenidos sobre los que versarán las pruebas de una misma temática y, además, cada banco de ítems habrá de pasar por estudios de fiabilidad, discriminación y dificultad para su calibración y comprobación de su adecuación.

Poca concreción de cómo se llevó a cabo la intervención

Un punto débil generalmente encontrado es la poca concreción de cómo se llevó a cabo la intervención. Para solventar esta carencia, en este trabajo se explicitó con detalle la muestra, los instrumentos utilizados y el procedimiento seguido para llevar a cabo no sólo las distintas fases de evaluación, sino también los estudios bibliográficos y meta-analítico y para elaborar la escala de calidad de estudios primarios.

Además, en busca del mayor grado de transparencia posible, en muchas ocasiones estos datos también se dejan disponibles a los participantes de las acciones formativas de las distintas organizaciones, para que todos tengan acceso a conocer la metodología utilizada, si fuera de interés.

Como mejora, se espera tener la posibilidad de dejar disponible esta información en todas las ocasiones (de todas las evaluaciones y en todas las organizaciones).

Poca concreción de las características de la muestra

Generalmente, las características de la muestra no son explicitadas en los programas de formación continua. En las evaluaciones presentadas en este trabajo, hasta ahora, no pudo obtenerse información muy concreta acerca de la muestra porque se utilizaron para la recogida de la información instrumentos anónimos en los que, además, no se incluyeron ítems referidos a características demográficas.

El único instrumento que recogió algo de este tipo de información fue el cuestionario de necesidades, donde el participante concretaba el lugar en que trabajaba y su puesto de trabajo además de que, de manera opcional, se le daba la oportunidad de aportar su nombre.

Como propuesta de mejora para el futuro, se espera conseguir la identificación de cada participante en todas las fases de evaluación. No interesa realmente quién es la persona en cuestión; los dos grandes beneficios serían, por un lado, obtener información acerca de las características demográficas del grupo para poder poner en relación estas variables con los resultados obtenidos; y, por otro lado, tener la posibilidad de comparar los resultados obtenidos por cada persona en las distintas fases, para así tener la

posibilidad de predecir la eficacia de una acción formativa en función del perfil de sus participantes.

Sin grupo de comparación

Otra deficiencia fuertemente extendida es la no existencia de grupo de comparación en los estudios de formación continua. Para solventar esta cuestión, lo que hasta ahora se ha hecho ha sido comparar a lo largo del tiempo; por ejemplo, en la evaluación de conocimientos se compara una medida antes de la intervención y otra después de ésta para determinar si hubo un aumento en los aciertos; y tanto en la evaluación de satisfacción como de conocimientos, los resultados anuales se compararon con años anteriores.

Dos mejoras se piensan implementar en este sentido:

- El próximo año, realizar la misma comparación general con los datos de evaluación de transferencia. Hasta ahora no pudo hacerse porque era el primer año en el que se obtenían estos datos.
- Implantar un diseño cuasiexperimental de grupo control no equivalente y prueba previa y posterior. Este grupo de comparación estará compuesto por aquellas personas que quedaron en “lista de espera” para ser incluidos en la próxima edición de la acción formativa.

Sin azar

En muy pocas ocasiones la muestra es elegida al azar de entre todos los potenciales usuarios y los participantes son asignados aleatoriamente a los distintos grupos. En las evaluaciones aquí presentadas, la muestra no es elegida al azar, sino que es escogida por la comisión de formación siguiendo unos criterios previamente explicitados y no existe asignación de los distintos grupos porque hasta ahora no hay grupos diferentes.

Seguir unos criterios claros a la hora de conformar una muestra, aunque a nivel estadístico no sea más positivo que la aleatorización, sí lo puede ser desde el punto de vista de la equidad, puesto que supuestamente participarán en la formación quienes más lo necesiten.

Un aspecto positivo que va en esta línea es que, en lo que a la evaluación respecta, pueden participar voluntariamente todas las personas; no es necesario tomar una muestra.

Mortalidad global elevada

Otra carencia encontrada en bastante número de ocasiones fue una mortalidad global elevada. Para calcular este dato, al menos habrían de compararse dos medidas. Por ello, y mientras no se pongan en relación los datos obtenidos en las distintas fases de evaluación, este dato sólo puede estudiarse en la evaluación de conocimientos.

Efectivamente, con las evaluaciones realizadas pudo constatarse que la mortalidad global era bastante elevada, puesto que en muchas ocasiones sólo se recogía un solo dato de los dos necesarios para llevar a cabo la comparación.

Como mejoras de futuro, además de tratar de evitar la pérdida de datos se inferirán, a la hora de realizar la comparación, los valores que no se recogieron.

Medida en un solo momento y sin seguimiento

Otra característica frecuente es la medición en un solo momento y la no medición pasado algo de tiempo tras la finalización de la acción formativa.

En la evaluación aquí propuesta, en cambio, se toman medidas antes de la intervención, durante ésta y una vez finalizada. Además, sólo en la evaluación de conocimientos se realizan dos medidas. Por último, la evaluación de la transferencia podría considerarse una medición de seguimiento, puesto que se realiza unos meses después de finalizada la evaluación.

Como mejora, cuando sea posible la identificación de los usuarios, se tratará de relacionar los resultados obtenidos en las distintas fases de evaluación para poder realizar un estudio más pormenorizado de todo el proceso de evaluación a lo largo del tiempo.

Otra posibilidad sería obtener en una acción formativa concreta y con el mismo instrumento, una medida antes de la acción formativa, otra justo tras finalizar ésta y una última pasados unos meses. De este modo, podría realizarse un análisis de tendencia.

Análisis de datos a nivel descriptivo

Generalmente, los análisis de datos realizados en formación continua son descriptivos. En la evaluación planteada en este trabajo, se trató de ir más allá:

- En la fase de detección de necesidades, además del estudio de la frecuencia, se realizó en la post-evaluación la priorización del listado de acciones formativas con el método más conveniente según un estudio comparativo realizado.
- En la fase de evaluación de conocimientos, se realiza la comparación entre dos medidas (previa y posterior) en cada acción formativa y se estudia la significación de las diferencias.

Algunas mejoras se pretenden realizar en el futuro:

- Triangular los datos obtenidos a través de técnicas cuantitativas y cualitativas. Por ejemplo, en la mayoría de los cuestionarios utilizados, se recoge información de ambos tipos y, aunque en cierto modo ya se ponen en relación, sería interesante sistematizar y perfeccionar un procedimiento concreto para analizar los datos en conjunto.

Además, en detección de necesidades está prevista la consolidación del uso de entrevistas o grupos de discusión en la evaluación principal y de mesas técnicas

en la post-evaluación, con lo que se obtendrán más datos cualitativos, hasta ahora menos valorados que los cuantitativos.

- Como ya se ha comentado, si se consigue identificar a los usuarios, se podrían poner en relación datos personales con los resultados obtenidos en la evaluación de necesidades, durante la acción formativa, y en la evaluación de la satisfacción, conocimiento y transferencia de cada persona, lo que podría servir para detectar variables que favorecen o dificultan el éxito de las acciones formativas y la evolución a lo largo del tiempo mediante un análisis de tendencia.
- En la evaluación de conocimientos, además de las comparaciones por parejas de cada acción formativa, se pretende realizar un meta-análisis para determinar la eficacia del plan de formación en conjunto y detectar posibles variables moderadoras que influyan en los resultados.

Nuevas tecnologías

Un aspecto especialmente relevante que no podía dejar de ser comentado en este capítulo es el desarrollo de las nuevas tecnologías aplicadas a la evaluación de la formación continua. Cada año que va pasando, la importancia de estos instrumentos va aumentando. Los beneficios que hasta ahora han aportado y las que aportarán en el futuro apenas eran imaginables unos años atrás.

Las nuevas tecnologías han abierto fronteras de tal manera que, en la actualidad, se ha hecho más fácil la participación de gran cantidad de personas, sin que la dispersión geográfica sea un problema.

Es importante señalar en este sentido que, con el uso de Internet, se hace imprescindible utilizar un método que asegure la privacidad de los datos de quienes participaron.

Con el uso de las nuevas tecnologías para la evaluación, han surgido nuevos modos de comunicación como, por ejemplo, el uso de foros y chats en la fase de detección de necesidades o los SMS (mensajes por móvil).

Los tiempos se han hecho más cortos porque los procesos de envío y de recepción de información se han agilizado mucho en comparación con aquellos tiempos donde las únicas opciones eran el envío por correo ordinario o en persona.

Con ayuda de los programas informáticos, se logró la recogida de datos y la obtención de los análisis de una forma más rápida y con menor probabilidad de errores.

Se posibilitó la difusión de la información (sobre la calendarización, el procedimiento a seguir y los resultados) a cualquier persona que estuviera interesada, con lo que se ganó en transparencia.

Aunque las mejoras hasta ahora conseguidas han sido de gran ayuda para aumentar la calidad de la evaluación de la formación continua, se trabaja actualmente por conseguir la automatización total del proceso. La intención es que, utilizando la programación, en el mismo momento en que la persona participante cumple el

instrumento, reciba feedback. Se trata de lograr que el análisis de resultados sea inmediato y que, a través de Internet, la persona interesada pueda consultar los resultados hasta ese momento obtenidos en la evaluación.

Como comentario final, cabe destacar que el trabajo nunca acaba. En esta tesis podrían haberse desarrollado muchos puntos que se han tocado superficialmente y que son interesantes y útiles para mejorar la calidad de la evaluación de la formación, pero es imposible abarcarlo todo y fue necesario delimitar unas actuaciones concretas. Con los puntos abiertos pero no estudiados en profundidad en esta tesis (estudio de validez de convergencia de la escala para medir la calidad metodológica de estudios primarios, estudio de la invarianza del cuestionario de satisfacción, desarrollo de tests adaptativos computerizados para el estudio de conocimientos, estudio de validez de contenido en el cuestionario de transferencia, etc.), se espera tener la posibilidad de seguir avanzando y continuar respondiendo a los interrogantes planteados.

Y cuanto más se estudia y se avanza, más aspectos en los que profundizar se encuentran. Por lo que más que como un punto y final, esta tesis es considerada como un punto y aparte donde quedan muchas cosas en el tintero por explorar. Esto es lo interesante de la ciencia: siempre quedarán más incógnitas.

REFERENCIAS BIBLIOGRÁFICAS

- Airasian, P. W. (1988). Societal experimentation. En G. F. Madaus, M. S. Scriven y D. L. Stufflebeam (Eds.), *Evaluation models. Viewpoints on educational and human services evaluation (7th edition)* (pp. 163-175). Boston: Kluwer-Nijhoff Publishing.
- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D. R., Gotzsche, P. C. & Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663-694.
- Alvira, F. (1991). *Metodología de la evaluación de programas*. Madrid: Centro de Investigaciones Sociológicas.
- Amat, O. (1992). Evaluación económica de la inversión en formación. *Revista de Profesiones y Empresas*, 2, 70-74.
- Anguera, M. T. y Chacón, S. (en prensa, a). Aproximación conceptual en evaluación de programas. En M. T. Anguera, S. Chacón, y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 1-32). Registro general de la propiedad intelectual: número de asiento registral 02/2005/2101 (efectos desde el 21/12/2004). Madrid: Síntesis.
- Anguera, M. T. y Chacón, S. (en prensa, b). Bases metodológicas en evaluación de programas. En M. T. Anguera, S. Chacón, y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 33-92). Registro general de la propiedad intelectual: número de asiento registral 02/2005/2101 (efectos desde el 21/12/2004). Madrid: Síntesis.
- Anguera, M. T., Chacón, S., Holgado, F. P. y Pérez, J. A. (en prensa). Instrumentos en evaluación de programas. En M. T. Anguera, S. Chacón, y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 173-224). Registro general de la propiedad intelectual: número de asiento registral 02/2005/2101 (efectos desde el 21/12/2004). Madrid: Síntesis.
- Arce, C. (1994). *Técnicas de construcción de escalas psicológicas*. Madrid: Síntesis.
- Arranz, P., Ulla, S. M., Ramos, J. L., del Rincón, C. & López-Fando, C. (2005). Evaluation of a counseling training program for nursing staff. *Patient Education and Counselling*, 56, 233-239.
- Artés, J. (1998). *Cómo crear pruebas objetivas*. Barcelona: Epise.
- Arvey, R. D. & Cole, D. A. (1989). Evaluating change due to training. En I. L. Goldstein and associates (Eds.), *Training and development in organizations* (pp. 89-118). San Francisco: Jossey-Bass.

- Ato, M. y Vallejo, G. (2007). Modelos de regresión I: regresión simple y múltiple. En M. Ato y G. Vallejo, *Diseños experimentales en psicología* (pp. 59-100). Madrid: Pirámide.
- Barbero, M. I. (1993). *Métodos de elaboración de escalas*. Madrid: UNED.
- Basarab, D. J. & Root, D. K. (1992). *The training evaluation process*. Boston: KluwerAcademia Publishers.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D. & Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. *JAMA*, 28(276), 637-639.
- Bosch, F., Guardiola, E. y grupo de trabajo Foundation Workshop 2002 (2003). Lista de comprobación (checklist) abreviada para la evaluación de artículos de investigación biomédica básica. *Medicina Clínica*, 121(6), 228-230.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D. R. & de Vet, H. C. W. (2003). *Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative*. Retrieved January 30, 2006, from <http://www.consort-statement.org/stardstatement.htm>.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D. R., de Vet, H. C. W. & Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clinical Chemistry*, 49(1), 7-18.
- Bou, S. Marro, F. y Pineda, P. (1999). ¿Cómo se evalúa la formación en las organizaciones? (I y II). *Revista Capital Humano*, 127, 68-78.
- Bray, J. H. & Howard, G. S. (1980). Methodological considerations in the evaluation of a teacher-training program. *Journal of Educational Psychology*, 72(1), 62-70.
- Bria, L., Doltra, M., Moreno, E., Pedrals, J., Juan, J. y Boldú, J. (2004). *Los libros de los filósofos*. Barcelona: Ariel.
- Broad, M. L. y Newstrom, J. W. (1992). *Cómo aplicar el aprendizaje al puesto de trabajo*. Madrid: CERASA.
- Brown, K. G. & Gerhardt, M. W. (2002). Formative evaluation: an integrative practice model and case study. *Personnel Psychology*, 55, 951-983.
- Brown, S. A. (1991). Measurement of quality of primary studies for meta-analysis. *Nursing Research*, 40(6), 352-355.
- Buckley, R. y Caple, J. (1991). *La formación. Teoría y práctica*. Madrid: Díaz de Santos.

- Camacho, C., López, A. M. y Arias, M. A. (curso 2006-07). Regresión lineal simple. Apuntes no publicados de la asignatura *Análisis de datos II* de la licenciatura de Psicología, Universidad de Sevilla.
- Camacho, C. y Sánchez, E. F. (1997). *Psicometría*. Sevilla: Kronos.
- Campbell, D. T. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Campbell, J. P. (1971). Personnel training and development. *Annual Review of Psychology*, 22, 256-602.
- Campbell, M. K., Elbourne, D. R. & Altman, D. G. (2004). CONSORT statement: extension to cluster randomized trials. *BMJ*, 328, 702-708.
- Campos J. R., Ruiz, J. A., Sánchez, M. y Chacón, S. (2007, febrero). Evaluación de la transferencia en formación continua “vía Web”. Un estudio de caso. Comunicación presentada en el *X Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Cascio, W. F. (1989). Using utility analysis to assess training outcomes. En I. L. Goldstein and associates (Eds.), *Training and development in organizations* (pp. 63-88). San Francisco: Jossey-Bass.
- Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla (1999). *II Plan de formación para el Personal de Administración y Servicios de la Universidad de Sevilla*. Sevilla: Centro de Formación y Perfeccionamiento del PAS.
- Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla (2001). *Guía del formador*. Sevilla: Centro de Formación y Perfeccionamiento del PAS.
- Chacón, S. (2004, November). Assessing the quality of randomized control trials vs. quasi-randomized control trials. Conference invited in *Method C2 Workshop. Nordic Campbell Center*. Copenhagen.
- Chacón, S., García, L., Alarcón, D. & Sanduvete, S. (2003, October). Quality of programs interventions for elderly people. A comparative study between USA and Europe (European Union Countries). Paper presented at the *XI Cochrane Colloquium*. Barcelona.
- Chacón, S., Gómez, N., Sánchez, M. y Sanduvete, S. (2006). *Formación en el Instituto Andaluz del Deporte: evaluación y diseño (4 vols.)*. Málaga: Instituto Andaluz del Deporte, Consejería de Turismo, Comercio y Deporte.
- Chacón, S., Holgado, F. P., López, J. M. y Sanduvete, S. (2006). *Evaluación de la formación continua: fundamentos teóricos y herramientas metodológicas*. Sevilla: Universidad de Sevilla.

- Chacón, S., Lara, A. y Pérez, J. A. (2001). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2000*. Sevilla: Diputación de Sevilla.
- Chacón, S., Lara, A. y Pérez, J. A. (2002a). *Evaluación de necesidades formativas para la programación del Plan Agrupado de Formación Continua de 2003*. Registro de entrada número 81, fecha 18/12/2002. Sevilla: Diputación de Sevilla.
- Chacón, S., Lara, A. y Pérez, J. A. (2002b). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2001*. Registro de entrada número 35, fecha 14/05/2002. Sevilla: Diputación de Sevilla.
- Chacón, S., Lara, A. y Pérez, J. A. (2002c). Needs assessment. In Fernández-Ballesteros, R. (Ed.), *Encyclopedia of psychological assessment, vol. 2* (pp. 615-619). Londres: SAGE Publications.
- Chacón, S. y López, J. M. (2003, octubre). *Ciclo de conferencias. Innovaciones metodológicas en evaluación de programas de formación continua y ocupacional*. Sevilla.
- Chacón, S., López, J. M., Campos, J. R., y Sanduvete, S. (2003). Herramientas metodológicas para la evaluación de programas de formación continua y ocupacional. Comunicación presentada en el *II Congreso de la Sociedad Española de Evaluación*. Jerez.
- Chacón, S., López, J. M. y Pérez, J. A. (2003). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2002*. Registro de entrada número 149, fecha 05/08/2003. Sevilla: Diputación de Sevilla.
- Chacón, S., López, J. M. y Sanduvete, S. (2004). *Evaluación de acciones formativas en Diputación de Sevilla. Una guía práctica*. Sevilla: Diputación de Sevilla.
- Chacón, S., López, J. M., Sanduvete, S. y Holgado, F. P. (2003). *Evaluación de necesidades formativas para la programación del Plan Agrupado de Formación Continua de 2004*. Registro de entrada número 161, fecha 16/09/2003. Sevilla: Diputación de Sevilla.
- Chacón S., Pérez, J. A., Holgado, F. P. y Lara, A. (2001a). Evaluación de la calidad universitaria: validez de contenido. *Psicothema, 13*, 294-301.
- Chacón, S., Pérez, J. A., Holgado, P. y Lara, A. (2001b). *Memoria del periodo 1998-2001 de la Unidad de Evaluación del Centro de Formación y Perfeccionamiento del PAS de la Universidad de Sevilla*. Informe técnico no publicado, fecha de junio de 2002. Sevilla: Centro de Formación y Perfeccionamiento del PAS.
- Chacón, S., Pérez, J. A. y Lara, A. (2000a). *Evaluación de necesidades formativas en los Municipios adheridos al Plan de Formación Continua*. Registro de entrada número 100, fecha 05/07/2000. Sevilla: Diputación de Sevilla.

- Chacón, S., Pérez, J. A. y Lara, A. (2000b). *Memoria comparativa de los Planes de Formación Continua 1995-98 y memoria de cada edición formativa*. Registro de entrada número 99, fecha 05/07/2000. Sevilla: Diputación de Sevilla.
- Chacón, S., Pérez, J. A., Lara, A. y Holgado, F. P. (2001). *Evaluación de necesidades formativas de la Diputación y Municipios adheridos. Propuesta de diseño del plan de formación del bienio 2001-2002*. Registro de entrada número 77, fecha 22/05/2001. Sevilla: Diputación de Sevilla.
- Chacón, S., Pérez, J.A., Sanduvete, S., Holgado, F. P., Sánchez, M. y Lozano, J. A. (2007). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2006*. Registro de entrada número 7, fecha 19/10/2007. Sevilla: Diputación de Sevilla.
- Chacón, S., Sánchez, M., Holgado, F. P. y Sanduvete, S. (2007). *Formación en el Instituto Andaluz del Deporte: evaluación y diseño (4 vols.)*. Málaga: Instituto Andaluz del Deporte, Consejería de Turismo, Comercio y Deporte.
- Chacón, S., Sánchez-Meca, J., Alarcón, D. & Marín, F. (2002, September). Evaluative results generalization, design quality and meta-analysis. Main characteristics of intervention designs of published papers in the European context. Paper presented at the *I Annual Campbell Collaboration Methods Group Meeting*. Baltimore.
- Chacón, S., Sánchez-Meca, J. y Sanduvete, S. (2007, febrero). Midiendo la calidad de los estudios primarios en meta-análisis. Comunicación presentada en el *X Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Chacón, S., Sánchez-Meca, J., Sanduvete, S. & Alarcón, D. (2006, February). A scale to measure the quality of primary studies for meta-analysis. A content validity study. Paper presented at the *VI Annual Campbell Collaboration Colloquium*. Los Angeles.
- Chacón, S., Sánchez-Meca, J., Sanduvete, S. y Alarcón, D. (en elaboración). Validation of a scale to measure the quality of primary studies for meta-analysis. *Evaluation and program planning*.
- Chacón, S., Sanduvete, S. y Alarcón, D. (2005, septiembre). Validez de contenido de una escala para valorar la calidad de los estudios primarios en meta-análisis. Comunicación presentada en el *IX Congreso de Metodología de las Ciencias Sociales y de la Salud*. Granada.
- Chacón, S., Sanduvete, S., y Holgado, F. P. (2004). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2003 (I y II)*. Registro de entrada número 2, fecha 18/01/05. Sevilla: Diputación de Sevilla.

- Chacón, S., Sanduvete, S. y Holgado, F. P. (2005a). *Evaluación de necesidades formativas para la programación del Plan Agrupado de Formación Continua de 2005*. Registro de entrada fecha 18/01/05 y número 2. Sevilla: Diputación de Sevilla.
- Chacón, S., Sanduvete, S. y Holgado, F. P. (2005b). *Evaluación de necesidades formativas para la programación del Plan Agrupado de Formación Continua de 2006*. Registro de entrada número 41, fecha 17/11/2005. Sevilla: Diputación de Sevilla.
- Chacón, S., Sanduvete, S., y Holgado, F. P. (2005c). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2004 (3 tomos)*. Registro de entrada número 42, fecha 28/11/2005. Sevilla: Diputación de Sevilla.
- Chacón, S., Sanduvete, S. y Holgado, F. P. (en elaboración). Herramientas metodológicas y análisis de datos en detección de necesidades de formación continua. *Psicothema*.
- Chacón, S., Sanduvete, S., Holgado, F. P., Rodríguez, J. L. y Gómez, N. (2007, febrero). Evaluación de necesidades. Técnicas de escalamiento. Comunicación presentada en el *X Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Chacón, S., Sanduvete, S., Holgado, F. P. y Sánchez, M. (2006). *Evaluación general correspondiente al Plan Agrupado de Formación Continua de la Diputación de Sevilla de 2005*. Registro de entrada número 07, fecha 05/02/2007. Sevilla: Diputación de Sevilla.
- Chacón, S., Sanduvete, S., Holgado, F. P. y Sánchez, M. (2007). *Evaluación de necesidades formativas para la programación del Plan Agrupado de Formación Continua de 2007*. Registro de entrada número 6, fecha 05/02/2007. Sevilla: Diputación de Sevilla.
- Chacón, S., Sanduvete, S., Sánchez, M. & Sánchez-Meca, J. (2007a, May). Enhancing utility of systematic reviews from methodological quality of primary studies. A case study in training programs. Paper presented at the *VI Annual Campbell Collaboration Colloquium*. London.
- Chacón, S., Sanduvete, S., Sánchez, M. & Sánchez-Meca, J. (2007b, May). Measuring methodological quality of primary studies for meta-analysis. Preliminary results from a proposed scale. Paper presented at the *VI Annual Campbell Collaboration Colloquium*. London.
- Chacón, S. y Shadish, W. R. (en prensa). Validez en evaluación de programas. En M. T. Anguera, S. Chacón, y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 93-132). Registro general de la propiedad intelectual: número de asiento registral 02/2005/2101 (efectos desde el 21/12/2004). Madrid: Síntesis.

- Chacón, S., Shadish, W. R. y Cook, T. D. (en prensa). Diseños evaluativos de intervención media. En M. T. Anguera, S. Chacón, y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 279-311). Registro general de la propiedad intelectual: número de asiento registral 02/2005/2101 (efectos desde el 21/12/2004). Madrid: Síntesis.
- Chelimsky, E. (1997). The coming transformations in evaluation. En E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century. A handbook* (pp. 1-26). Londres: Sage Publications.
- Connolly, S. (1983). Participant evaluation: finding out how well training worked. *Training and Development Journal, October*, 92-96.
- Coogle, C. L., Parham, I. A., Welleford, E. A., & Netting, F. E. (2002). Evaluation of a distance learning course in geriatric interdisciplinary teaming. *Educational Gerontology, 28*, 791-804.
- Cook, T. D. (1997). Lessons learned in evaluation over the past 25 years. En E. Chelimsky y W. R. Shadish (Eds.), *Evaluation for the 21st century. A handbook* (pp. 30-52). Londres: Sage Publications.
- Cook, T. D., Campbell, D. T. & Peracchio, L. (1990). Quasi-experimentation. En M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 699-777). New York: Randon House.
- Cronbach, L. J. (1982). *Designing evaluation of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L. J. (1988). Course improvement through evaluation. En G. F. Madaus, M. S. Scriven & D. L. Stufflebeam (Eds.), *Evaluation models. Viewpoints on educational and human services evaluation* (pp. 101-115). Boston: Kluwer-Nijhoff Publishing.
- De Diego, R., Fraile, A. y Boada, J. (2004). Necesidades formativas de los técnicos deportivos. *Psicothema, 16*(1), 84-87.
- Des Jarlais, D. C., Lyles, C., Crepaz, N. & the TREND group (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health, 94*(3), 361-366.
- Di Paolo, A. J. & Patterson, A. C. (1983). Selecting a training program for new trainers. *Training and Development Journal, January*, 96-101.
- Education Group for Guidelines on Evaluation (1999). Guidelines for evaluating papers on educational interventions. *BMJ, 318*, 1265-7.
- Emerson, J. D., Burdick, E., Hoaglin, D. C., Mosteller, F. & Chalmers, T. C. (1990). An empirical study of the possible relation or treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clinical Trials, 11*, 339-352.

- Fernández-Ballesteros, R. (1995). El ciclo de intervención social y evaluación. En R. Fernández-Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 50-74). Madrid: Síntesis.
- Fernández-Salineró, C. (1999). El diseño de un plan de formación como estrategia de desarrollo empresarial: estructura, instrumentos y técnicas. *Revista Complutense de Educación*, 10(1), 181-242.
- Follete, W. C. & Beitz, K. (2003). Adding a more rigorous scientific agenda to the empirically supported treatment movement. *Behavior Modification*, July, 369-386.
- Fondo Social Europeo (2007). *Iniciativa empresarial y formación continua. Programa operativo: objetivo 1*. Consulta realizada el 15/11/2007 en <http://www.mtas.es/uafsecgi/download/iniemp-1.pdf>.
- Gaines, D. y Robinson, J. C. (1998). *Seguimiento de resultados operativos (número 3)*. Barcelona: Epise.
- García, Y. (1994). Análisis de la evaluación de cursos de formación. *Revista de psicología del Trabajo y de las Organizaciones*, 10(27), 19-37.
- Goldstein, I. L. (1989). Critical training issues. Past, present and future. En I. L. Goldstein and associates (Eds.), *Training and development in organizations* (pp. 1-21). San Francisco: Jossey-Bass.
- Grafinger, D. (1998). *Evaluar los resultados del aprendizaje*. Barcelona: Epise.
- Greenland, S. (1994). Quality scores are useless and potentially misleading. *American Journal of Epidemiology*, 140(3), 300-301.
- Hedges L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Hernández, S., Pozo, C., y Alonso, E. (2004). La aproximación multimétodo en evaluación de necesidades. *Apuntes de Psicología*, 22(3), 293-308.
- Holgado, F. P. (2002). *Evaluación institucional: alternativas metodológicas en la delimitación y mejora de la calidad universitaria*. Sevilla: tesis doctoral no publicada.
- Holgado, F. P., Chacón, S., Barbero, M. I. & Sanduvete, S. (2006). Training satisfaction rating scale: development of a measurement model using polychoric correlations. *European Journal of Psychological Assessment*, 22(4), 268-279.
- Holgado, F. P., Chacón, S., Losada, J. L., Pérez, J. A. y López, J. M. (2003). Validación de un instrumento para medir la reacción de los participantes a las acciones formativas. Comunicación presentada en el IX Congreso de Metodología de las Ciencias Sociales y de la Salud. Valencia.

- Holgado, F. P., Chacón, S., Ruiz, J. A., López, J. M. y Barbero, M. I. (en elaboración). *TAI en evaluación de la formación continua*.
- Instituto de Formación y Estudios Sociales (I.F.E.S.) (2002). *Manual del formador*. Madrid: Grayda.
- Instituto Nacional de Administración Pública (2001). *III Acuerdo de la Formación Continua en la Administración Pública* (BOE 098 de 24/04/2001).
- Irvine, A. B., Ary, D. V. & Bourgeois, M. S. (2003). An interactive multimedia program to train professional caregivers. *The Journal of Applied Gerontology*, 22(2), 269-288.
- Julià, E. L. (2001). El seguimiento de la formación: cómo hacer que los resultados de la formación perduren. *Revista Capital Humano*, 145, 52-55.
- Jüni, P., Altman, D. G. & Egger, M. (2001). Assessing the quality of randomized controlled trials. In M. Egger, G. D. Smith & D. G. Altman (Ed.), *Systematic reviews in health care*. London: BMJ.
- Jüni, P., Witschi, A., Bloch, M. D. & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*, 282(11), 1054-1060.
- Kaufman, R. (1988). *Planning educational systems: a results-based approach*. Lancaster, PA: Technomic.
- Khan, H. (2002). Effectiveness of a strategic management development program. *Applied H. R. M. Research*, 7(2), 49-52.
- Kirkpatrick, D. L. (1977). Evaluating training programs: evidence vs. proof. *Training and Development Journal*, November, 9-12.
- Kirkpatrick, D. L. (1999). *Evaluación de acciones formativas*. Barcelona: Epise.
- Komras, H. (1985). Evaluating your training programs. *Training and Development Journal*, September, 87-88.
- LaLonde, R. J. (1986). Evaluating the econometric evaluation of training programs with experimental data. *American Economic Review*, 76, 604-620.
- Lara, A., Díaz, M. y Chacón, S. (2002, febrero). La gestión y evaluación de la transferencia aplicada a la formación en seguridad. Comunicación presentada en el 2nd *International Conference on Occupational Risk Prevention*. Madrid.
- Latham, G. P. (1988). Human resource training and development. *Annual Review of Psychology*, 39, 545-582.
- Letón, E. y Pedromingo, A. (2001). *Introducción al análisis de datos en meta-análisis*. Madrid: Díaz de Santos.

- Lintzeris, N., Ritter, A. & Dunlop, A. (2002). Training primary health care professionals to provide buprenorphine and LAAM treatment. *Substance Abuse*, 23(4), 245-254.
- López, J., Trigo, M. E. y Arias, M. A. (1999). La lógica del diseño experimental y del análisis estadístico de los datos. En *Diseños experimentales. Planificación y análisis*. Sevilla: Kronos.
- MacPherson, H., White, A., Commings, M., Jobst, K., Rose, K. & Niemzow, R. (2002). Standards for reporting interventions in controlled trial of acupuncture: the STRICTA recommendations. *Acupuncture in Medicine*, 20(1), 22-25.
- Martínez, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- McGuire, J., Bates, G. W., Dretzke, B. J., McGivern, E., Rembold, K. L., Seabold, D. R., Turpin, B. A. M. & Levin, J. R. (1985). Methodological quality as a component of meta-analysis. *Educational Psychologist*, 20, 1-5.
- Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P. & Walsh, M. D. (1995). Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, 62-73.
- Moher, D., Jadad, A. R. & Tugwell, P. (1996). Assessing the quality of randomized controlled trials: current issues and future directions. *International Journal of Technology Assessment in Health Care*, 12(2), 195-208.
- Moher, D., Jones, A. & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*, 285(15), 1992-1995.
- Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P. & Klassen, T. P. (1998). Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet*, 22(352), 609-614.
- Moher, D., Schulz, K. F. & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *The Lancet*, 14(357), 1191-1194.
- Moreski, R. T., Eliades, M. J., Bhimani, M. A., Bunney, E. B. & VanRooyen, M. J. (2001). Preparing international relief workers for health care in the field: an evaluation of organizational practices. *Prehospital and Disaster Medicine*, 16(4), 257-262.
- Munoz, M. A & Munoz, M. D. (1999). *The role of occupational training and evaluation*. U.S. Department of Education, Office of Educational Research and Improvement (OERI), National Library of Education (NLE) & Educational Resources Information Center (ERIC).

- Newes-Adeyi, G., Helitzer, D. L., Roter, D. & Caulfield, L. E. (2004). Improving client-provider communication: evaluation of a training program for women, infants and children (WIC) professionals in New York State. *Patient Education and Counselling*, 55, 210-217.
- Newstrom, J. W. (1978). Catch 22: the problems of incomplete evaluation of training. *Training and Development Journal*, 32, 22-24.
- Olivares, J., Rosa, A. I. y Sánchez-Meca, J. (2000). Meta-análisis de la eficacia de las habilidades de afrontamiento en problemas clínicos y de salud en España. *Anuario de Psicología*, 31(1), 43-61.
- O'Rourke, K. & Detsky, A. S. (1989). Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *Journal of Clinical Epidemiology*, 42(10), 1021-1024.
- Ortí, C. (1994). Evaluación de la formación: un caso práctico. *Revista de Psicología del Trabajo y de las Organizaciones*, 28(10), 63-69.
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157-159.
- Osterlind, S. J. (1998). *Constructing tests items*. Boston: Kluwer Academic Publishers.
- Palmer, A., Jiménez, R. y Montaña, J. J. (2000). Tutorial sobre coeficientes de correlación con una o dos variables categóricas. *Psicología.com (revista electrónica de Psicología)*, 4(2). Consulta realizada el 11/11/2007 en <http://www.bibliopsiquis.com/psicologiacom/vol4num2/2830/#bis>.
- Panter, K. L. (1994). Greenhouse pesticide application safety training. *Hort Technology*, July-September, 299-301.
- Parker, T. C. (1973). Evaluation: the forgotten finale of training. *Personnel*, November-December, 61-63.
- Pérez, J. A. (curso 2006-07). Evaluación del instrumento de medida: análisis de ítems. Apuntes no publicados de la asignatura *Psicometría* de la licenciatura de Psicología, Universidad de Sevilla.
- Pérez, J. A., Barbero, M. I., Holgado, F. P., Chacón, S., Ruiz, J. A., Campos, J. R., López, J. M. y Suárez, J. C. (2003). Acreditación en ofimática mediante tests computerizados en evaluación de programas de formación continua. Comunicación presentada en el *VIII Congreso de Metodología de las Ciencias del Comportamiento*. Madrid.
- Phillips, J. J. (1990). *Handbook of training evaluation and measurement methods*. Londres: Kogan Page.
- Phillips, P. P. & Phillips, J. J. (2001). Symposium on the evaluation of training. *Internacional Journal of Training and Development*, 5(4), 240-247.

- Pineda, P. (2000). Evaluación del impacto la formación en las organizaciones. *Educar*, 27, 119-133.
- Quinet, K., Nunn, S. & Kincaid, N. L. (2003). Training police: a case study of differential impacts of problem-oriented policing training. *Police Practice and Research*, 4(3), 263-283.
- Ramírez del Río, A. (1997). *Valoración de la formación. Cómo rentabilizar los costes de formación*. Madrid: Griker y Asociados.
- Ritter, L. A. & Sue, V. M. (2007). Managing online survey data. In Using online surveys in evaluation, (special issue), *New Direction for Evaluation*, 115, autumn (fall) 51-55.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641.
- Salanova, M. y Grau, R. (1999). Análisis de necesidades formativas y evaluación de la formación en contextos de cambio tecnológico. *Revista de Psicología General y Aplicada*, 52(2-3), 329-350.
- Sánchez-Meca, J. (1997). Methodological issues in the meta-evaluation of correctional treatment. In S. Redondo, V. Garrido, J. Pérez & R. Barberet (Ed.), *Advances in psychology and law. International contributions*. Berlin & New York: Walter de Gruyter.
- Sánchez-Meca, J. (2000, mayo). Generalización de resultados evaluativos mediante meta-análisis. Conferencia invitada en el ciclo *Validez y Generalización de Resultados en Evaluación de Programas: Avances Metodológicos en Diseño y Análisis*. Sevilla.
- Sánchez-Meca, J. y Ato, M. (1989). Meta-análisis: una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Arnau y H. Carpintero (Eds.), *Historia teoría y método*. Vol. 1 del *Tratado de psicología general* (J. Mayor y J. L. Pinillos, Eds.) (pp. 617-669). Madrid: Alhambra.
- Sánchez-Meca, J. Rosa, A. I. y Olivares, J. (1999). Las técnicas cognitivo-conductuales en problemas clínicos y de salud: meta-análisis de la literatura española. *Psicothema*, 11(3), 641-654.
- Sanduvete, S. (2004). Calidad de vida en las personas mayores. *Apuntes de Psicología*, 22(2), 277-288.
- Sanduvete, S. (2006). *Avances metodológicos para la mejora de las intervenciones dirigidas a las personas mayores*. Trabajo de investigación. Director: Salvador Chacón Moscoso.
- Sanduvete, S., Chacón, S., Holgado, F. P., Gómez, N. y Sánchez, M. (2006, julio). Causal analysis in training program evaluation. Póster presentado en el

simposium sobre *Análisis del Efecto Causal en Diseños Experimentales y Cuasi-experimentales*. Schloss Dornburg (Germany).

- Sanduvete, S., Chacón, S. y Sánchez, M. (2007, febrero). Diseño y evaluación de programas de formación continua: estado de la cuestión y propuestas de mejora. Comunicación presentada en el *X Congreso de Metodología de las Ciencias Sociales y de la Salud*. Barcelona.
- Scriven, M. (1967). The methodology of evaluation. En R. W. Tyler, R. M. Gagne y M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Sección de Formación Continua de Diputación de Sevilla (2007). *Manual de formación*. Consulta realizada el 21/11/2007 en http://195.57.5.45/formcont2001/manual_formacion_continua.pdf.
- Shadish, W. R., Chacón, S. & Sánchez-Meca, J. (2005). Evidence-based decision making: enhancing systematic reviews of program evaluation results in Europe. *Evaluation*, 11(1), 95-109.
- Shadish, W. R., Cook, T. D. y Campbell D. T. (2002a). Experiments and generalized causal inference. En W. R. Shadish, T. D. Cook and D. T. Campbell (Eds.), *Experimental and quasi-experimental designs for generalized causal inference* (pp. 1-32). Boston & New York: Houghton Mifflin Company.
- Shadish, W. R., Cook, T. D. y Campbell D. T. (2002b). Practical problems 1: Ethics, participant recruitment, and random assignment. En W. R. Shadish, T. D. Cook and D. T. Campbell (Eds.), *Experimental and quasi-experimental designs for generalized causal inference* (pp. 279-313). Boston & New York: Houghton Mifflin Company.
- Shadish, W. R., Cook, T. D. y Campbell D. T. (2002c). Quasi-experimental designs that use both control groups and pretests. En W. R. Shadish, T. D. Cook and D. T. Campbell (Eds.), *Experimental and quasi-experimental designs for generalized causal inference* (pp. 135-169). Boston & New York: Houghton Mifflin Company.
- Shadish, W. R., Cook, T. D. & Levinton, L. (1991). *Foundations of program evaluation. Theories and practice*. New York: Sage Publications.
- Shadish, W. R. & Myers, D. (2004). *Research design. Policy brief for C2*. Consulta realizada el 10/11/2007 en: <http://www.campbellcollaboration.org/MG/ResDesPolicyBrie.pdf>.
- Shadish, W. R. & Ragsdale, K. (1996). Random versus non random assignment in controlled experiments: do you get the same answer? *Journal of Consulting and Clinical Psychology*, 64(4), 1209-1205.

- Siegel, S. (1991a). El caso de dos muestras relacionadas. En S. Siegel, (Ed.), *Estadística no paramétrica aplicada a las ciencias de la conducta* (pp. 84-119). México D. F.: Trillas.
- Siegel, S. (1991b). Las medidas de correlación y sus pruebas de significación. En S. Siegel, (Ed.), *Estadística no paramétrica aplicada a las ciencias de la conducta* (pp. 226-278). México D. F.: Trillas.
- Sims, R. R. (1993). Evaluating public sector training programs. *Public Personnel Management*, 22(4), 591-615.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2), 5-8.
- Stufflebeam, D. L. (1971). Stufflebeam's improvement orient evaluation. En D. L. Stufflebeam y A. J. Shrinkfield (Eds.), *Systematic evaluation* (pp. 151-270). Boston: Kluwer Academic.
- Stufflebeam, D. L. & Shrinkfield, A. J. (1987). *Evaluación sistemática. Guía teórica y práctica*. Madrid: Paidós-MEC.
- Stufflebeam, D. L. & Webster, W. J. (1988). An analysis of alternative approaches to evaluation. En G. F. Manaus, M. S. Scriven y D. L. Stufflebeam (Eds.), *Evaluation models. Viewpoints on educational and human services evaluation* (pp. 23-45). Boston: Kluwer-Nijhoff Publishing.
- Suchman, E. (1967). *Evaluative research. Principles and practice in public service and social action programs*. New York: Sage Publications.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. & Sing, F. (2000). Study quality. In A. J. Sutton, K. R. Abrams, D. R. Jones, T. A. Sheldon & F. Sing (Eds.), *Methods for meta-analysis in medical research*. Chichester, New York: Wiley Series in Probability and Statistics.
- Tharenou, P. (1991). Managers' training. Needs and preferred training strategies. *Journal of Management Development*, 5, 46-59.
- Thayer, P. W. (1991). A historical perspective on training. En I. L. Goldstein and Associates (Eds.), *Training and development in organizations* (pp. 457-468). San Francisco: Jossey-Bass.
- Trevisan, M. S. (2004). Practical training in evaluation: a review of the literature. *American Journal of Evaluation*, 25(2), 255-272.
- Tritchler, D. (1999). Modelling study quality in meta-analysis. *Statistics in Medicine*, 18, 2135-2145.
- Tyler, R. (1942). General statements on evaluation. *Journal of Educational Research*, 35, 492-501.

- Unidad de Evaluación del Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla (2000). *Resultados del cuestionario sobre servicios de formación universitaria*. Consulta realizada el 17/10/2006 en <http://www.forpas.us.es/jornadas2000>.
- Unidad de Evaluación del Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla (2006a). *Informe de evaluación de la transferencia en FORPAS*. Sevilla: Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla.
- Unidad de Evaluación del Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla (2006b). *Memoria 2000-2005*. Sevilla: Centro de Formación y Perfeccionamiento del Personal de Administración y Servicios de la Universidad de Sevilla.
- Ventosa, P. (1998). *Desde la evaluación de la formación al rendimiento de la inversión*. Barcelona: Epise.
- Villar, L. M., De Vicente, P., Gallego, M. J., León, M. J., Molina, E., Moral, C. y Pérez, P. (1996). *Evaluación de programas de formación permanente para la función directiva*. Sevilla: Universidad de Sevilla.
- Weiss, C. H. (1972). The politicization of evaluation research. En C. H. Weiss (Ed.), *Evaluation action programs. Readings in social actions and educations* (pp. 327-338). Boston: Allyn & Bacon.
- Weiss, C. H. (1973). Where politics and evaluation research meet. *Evaluation, 1*, 37-45.
- Weisz, J. R., Hawley, K. M., Pilkonis, P. A., Woody, S. R. & Follette, W. C. (2000). Stressing the (other) three Rs in the search for empirically supported treatments: review procedures, research quality, relevance to practice and the public interest. *Clinical Psychology of Scientific Practice, 7*, 243-258.
- Wholey, J. S. (1983). *Evaluation and effective public management*. Boston: Little Brown.
- Wilkinson, S. M., Leliopoulou, C., Gambles, M. & Roberts, A. (2003). Can intensive three-day programs improve nurses' communication skills in cancer care? *Psycho-oncology, 12*, 747-759.
- Witkin, B. R. & Altschuld, J. W. (1995). *Planning and conducting needs assessments. A practical guide*. London: Sage.
- Wolf, F. M. (1986). *Meta-analysis: quantitative methods for research synthesis*. In J. L. Sullivan & R. G. Niemi (Eds.), *Serie Quantitative applications in the social sciences*. California: Sage Publications.

ÍNDICE DE AUTORES

A

Abrams, K. R., 79
Airasian, P. W., 39
Alarcón, D., 80, 82, 83,
114, 135, 139, xxi
Alliger, G., 59
Alonso, E., 42
Altman, D. G., 78, 79,
80
Altschuld, J. W., 19, 42,
211, 214, 219, 224,
233, 320, 322
Alvira, F., 19, 29, 41,
42, 66, 71, 72, 206,
320, 322
Amat, O., 68
Anguera, M. T., 19, 25,
29, 30, 33, 41, 42, 43,
66, 71, 72, 107, 172,
206, 268, 320, 322
Arce, C., 19, 214, 229,
230, 233, 236, 254
Arias, M. A., 190
Arranz, P., 182
Artés, J., 290
Arvei, R. D., 290, 306
Ary, D. V., 182
Ato, M., 195

B

Barbero, M. I., 19, 59,
60, 213, 214, 229, 233,
236, 240, 247, 249,
274, 303
Basarab, D. J., 59, 306
Bates, G. W., 80
Begg, C., 79
Beitz, K., 77, 139
Bennett, W., 59
Bhimani, M. A., 77, 139
Bloch, M. D., 80
Boada, J., 211
Boldú, J., 30
Bordonaro, T., 51, 68,
71
Bosch, F., 80
Bossuyt, P. M., 79, 81
Bou, S., 307

Bourgeois, M. S., 182

Bray, J. H., 182
Bria, L., 30
Broad, M. L., 60, 305
Brown, K. G., 31, 32,
50, 51, 68, 71, 72
Brown, S. A., 80
Bruno, D. E., 79, 81
Buckley, R., 17, 26
Bunney, E. B., 77, 139
Burdick, E., 80

C

Camacho, C., 190, 283
Campbell, D. T., 28, 35,
36, 38, 46, 54, 64, 65,
72, 73, 84, 85, 91, 92,
103, 105, 118, 119,
166, 169, 170, 172,
174, 304, 313
Campbell, J. P., 286
Campbell, M. K., 80
Campos, J. R., 41, 59,
60, 311
Caple, J., 17, 26
Carey, L., 50, 51, 67, 71,
72
Carmichael, L., 56
Cascio, W. F., 41, 55,
63, 69, 72, 207
Caulfield, L. E., 182
Centro de Formación y
Perfeccionamiento del
PAS de la
Universidad de
Sevilla, 19, 28, 41, 42,
66, 71, 72, 205, 206,
263, 271, 306, 320,
322
Chacón, S., 19, 25, 26,
28, 29, 30, 33, 36, 41,
42, 43, 49, 59, 60, 64,
66, 71, 72, 78, 80, 82,
83, 84, 94, 96, 98, 102,
107, 114, 135, 139,
172, 175, 196, 206,
211, 212, 213, 214,
219, 220, 224, 233,

261, 262, 268, 271,
274, 289, 290, 293,
302, 303, 304, 305,
306, 308, 311, 320,
322, xxi, lxxxii

Chalmers, T. C., 80
Chelimsky, E., 30
Cho, M., 79
Cohen, J., 195
Cole, D. A., 59, 290,
306
Comings, M., 80, 81
Connolly, S., 53, 68, 72
Coogler, C. L., 182
Cook, D. J., 79
Cook, T. D., 28, 30, 35,
36, 41, 46, 64, 73, 91,
92, 98, 103, 105, 118,
119, 166, 169, 170,
172, 174, 175, 302,
304, 313
Crepaz, N., 79, 80
Cronbach, L. J., 33, 35,
36, 37, 64, 72, 104

D

Davidoff, F., 80
de Diego, R., 211
de Vet, H. C. W., 79, 80,
81
de Vicente, P., 305
del Rincón, C., 182
Des Jarlais, D. C., 79,
80
Detsky, A. S., 79
Di Paolo, A. J., 27
Díaz, M., 305
Dick, W., 50, 51, 67, 71,
72
Doltra, M., 30
Dretzke, B. J., 80
Dunlop, A., 182
Dunn-Rankin, D., 214,
229, 230, 233, 236,
237, 238, 240, 247,
249, 250, 251, 252,
253

- E**
Eastwood, S., 79
Education Group for
Guidelines on
Evaluation, 80, 81
Egger, M., 78, 79, 80
Elbourne, D. R., 80
Eliades, M. J., 2, 77,
139
Emerson, J. D., 80
- F**
Fernández-Ballesteros,
R., 25, 29
Fernández-Salinero, C.,
307
Follete, W. C., 77, 139
Fondo Social Europeo,
17
Foundation Workshop
2002, 80
Fraile, A., 211
- G**
Gaines, D., 61, 312, 315
Gallego, M. J., 305
Gambles, M., 182
García, L., 82, 83, xxi
García, Y., 59
Gatsonis, C. A., 79, 80,
81
Gerhardt, M. W., 31, 32,
50, 51, 68, 71, 72
Glasziou, P. P., 79, 80,
81
Goldstein, I. L., 41, 54,
55, 69, 71, 72
Gómez, N., 94, 96, 139,
213, 271
Gossman, J. R., 49, 67
Gotsche, P. C., 80
Grafinger, D., 59
Grau, R., 220, 307
Greenland, S., 80
Guardiola, E., 80
- H**
Hawley, K. M., 79
Hedges, L. V., 181, 185
Helitzer, D. L., 182
Hernández, S., 42
Hoaglin, D. C., 80
Holgado, F. P., 26, 28,
37, 49, 59, 60, 71, 84,
94, 96, 107, 139, 212,
213, 268, 271, 274,
290, 293, 303, 305,
306, 308, lxxxii
Horton, R., 79, 212
Howard, G. S., 182
- I**
IFES, 26
INAP, 221
Irvine, A. B., 182
Irwig, L. M., 79, 80, 81
- J**
Jadad, A. R., 79, 80
Jiménez, R., 296
Jobst, K., 80, 81
Jones, A., 79
Jones, D. R., 79
Juan, J., 30
Julià, E. L., 307
Jüni, P., 78, 79, 80
- K**
Kaufman, R., 42, 211
Khan, H., 182
Kincaid, N. L., 182
Kirkpatrick, D. L., 19,
55, 58, 59, 60, 61, 62,
63, 70, 71, 77, 212,
271, 289, 290, 315,
320, 322
Klassen, T. P., 79
Komras, H., 53, 69, 72,
73
- L**
LaLonde, R. J., 182
Lang, T., 80
Lara, A., 19, 28, 42, 84,
211, 213, 214, 219,
220, 224, 233, 271,
305, 320, 322
Latham, G. P., 54, 69,
72, 73
Leliopoulou, C., 182
León, M. J., 305
Lepage, L., 79
Letón, E., 129, 181
Levene, 163
Levin, J. R., 41, 80
Levinton, L., 41
Lijmer, J. G., 79, 80, 81
Lintzeris, N., 182
Lipsey, M. W., 183
López, A. M., 190
López, J., 167
López, J. M., 26, 41, 49,
59, 60, 71, 213, 261,
262, 271, 274, 289,
293, 303, 305, 306,
308, lxxxii
López-Fando, C., 182
Losada, J. L., 59, 274
Lozano, J. A., 271, 290
Lyles, C., 79, 80
- M**
MacPherson, H., 80
Marín, F., 82
Marro, F., 307
Martínez, R., 283
McAlpine, L., 51, 68,
71
McGivern, E., 80
McGuire, J., 80
Moher, D., 79, 80, 81
Moher, M., 79
Molina, E., 305
Montaño, J. J., 296
Moral, C., 305
Moreno, E., 30
Moresky, R. T., 77, 139
Mosteller, F., 80
Munoz, M. A., 63
Munoz, M. D., 63
Myers, D., 77, 98, 104
- N**
Netting, F. E., 182
Newes-Adeyi, G., 182
Newstrom, J. W., 60, 62,
305
Nichol, G., 79
Niemzow, R., 80, 81
Nunn, S., 182

- O**
Olivares, J., 80, 91, 92, 93, 94, 181, 183, 184, 197
Olkin, I., 79, 181, 185
O'Rourke, K., 79
Ortega y Gasset, J., 30
Ortí, C., 307
Orwin, R. G., 195
Osterlind, S. J., 85, 91, 314
- P**
Palmer, A., 296
Panter, K. L., 182
Parham, I. A., 182
Parker, T. C., 49, 67, 71
Patterson, A. C., 27
Pedrals, J., 30
Pedromingo, A., 129, 181
Penman, M., 79
Peracchio, L., 35, 36
Pérez, J. A., 19, 28, 42, 59, 60, 84, 107, 211, 213, 214, 219, 220, 224, 233, 268, 271, 274, 290, 296, 320, 322
Pérez, P., 305
Pham, B., 79
Phillips, J. J., 58, 59, 60, 61, 62, 63, 70, 71, 212
Phillips, P. P., 59, 60, 61, 62, 63, 70, 71, 212
Pilkonis, P. A., 79
Pineda, P., 63, 307
Pitkin, R., 79
Ponsoda, V., 247
Pozo, C., 42
- Q**
Quinet, K., 182
- R**
Ragsdale, K., 77, 103, 197
Ramírez del Río, A., 289
Ramos, J. L., 182
Reitsma, J. B., 79, 81
Rembold, K. L., 80
Rennie, D. R., 79, 80, 81
Ritter, A., 182
Ritter, L. A., 85
Roberts, A., 182
Robinson, J. C., 61, 312, 315
Rodríguez, J. L., 213
Root, D. K., 59, 306
Rosa, A. I., 80, 91, 92, 93, 94, 181, 183, 184, 197
Rose, K., 80, 81, 194
Rosenthal, R., 194
Roter, D., 182
Ruiz, J. A., 59, 60, 303, 311
- S**
Salanova, M., 220, 307
Sánchez, E. F., 283
Sánchez, M., 94, 96, 98, 114, 139, 196, 197, 199, 213, 271, 290, 306, 311
Sánchez-Meca, J., 78, 79, 80, 81, 82, 91, 92, 93, 94, 96, 98, 102, 114, 116, 135, 139, 181, 183, 184, 195, 196, xxi
Sanduvete, S., 26, 41, 49, 59, 71, 78, 80, 82, 83, 94, 96, 98, 102, 114, 134, 135, 139, 166, 196, 212, 213, 261, 262, 271, 274, 289, 290, 293, 305, 306, 308, xxi, lxxxi
Scheffé, H., 163, 164
Schulz, K. F., 79, 80
Scriven, M., 29, 31, 38, 65
Seabold, D. R., 80
Sección de Formación Continua de
Diputación de Sevilla, 263
Shadish, W. R., 28, 35, 36, 41, 46, 64, 73, 77, 91, 92, 98, 103, 104, 105, 118, 119, 139, 166, 169, 170, 172, 174, 175, 197, 302, 304, 313
Sheldon, T. A., 79
Shotland, A., 59
Shrinkfield, A. J., 29, 34, 38
Siegel, S., 251, 252
Simel, D., 79
Sims, R. R., 55, 56, 70, 72
Sing, F., 79
Spearman, C., 251
Stake, R. E., 39, 65, 71, 72
Stanley, J., 38, 65, 72
"Student" (Sealy, W.), 290, 297
Stroup, D. F., 79
Stufflebeam, D. L., 29, 34, 38, 40, 66, 71, 72
Suárez, J. C., 59, 60
Suchman, E., 38, 65, 72
Sue, V. M., 85
Sutton, A. J., 79
Swierczek, F. W., 56
- T**
Tannenbaum, S., 59
Tharenou, P., 305
Thayer, P. W., 59
Thurstone, L. L., 213, 214, 229, 233, 236, 239, 253
Trave, H., 59
TREND group, 79, 80
Trevisan, M. S., 77, 139
Trigo, M. E., 167
Tritchler, D., 79
Tugwell, P., 79, 80
Turpin, B. A. M., 80
Tyler, R., 37, 64
- U**

Ulla, S. M., 182
Unidad de Evaluación
del Servicio de
Formación y
Perfeccionamiento del
PAS de la USE, 271,
306

V

Vallejo, G., 190
VanRooyen, M. J., 77,
139
Ventosa, P., 59
Villar, L. M., 305

W

Walsh, M. D., 79
Webster, W. J., 38
Weiss, C. H., 29, 41, 66
Weisz, J. R., 79
Welleford, E. A., 182
Weston, C., 51, 68, 71
White, A., 80, 81
Wholey, J. S., 29
Wilcoxon, F., 252, 290,
297
Wilkinson, S. M., 182
Witkin, B. R., 19, 42,
211, 214, 219, 224,
233, 320, 322
Witschi, A., 80
Wolf, F. M., 183
Woody, S. R., 79

ÍNDICE DE TÉRMINOS

- A**
Acreditación, 37, 64, 72
Actitud, 49, 50, 58, 59, 60, 70, 289, 305
Adecuación, 33, 78, 195, 290, 300, 324, lxxix
Agilización del proceso, 303, 304
Aleatorización, 46, 83, 104, 129, 147, 148, 169, 325, v, xxv
Allegado, 31
Ambiente, 27, 33, 52, 262, 274, 277, 323, lxxvii, lxxv
Amplitud de intervalo, 242
Análisis de datos, 30, 38, 39, 69, 70, 81, 85, 116, 165, 168, 219, 223, 229, 232, 233, 236, 237, 256, 295, 303, 306, 311, 314, 321, 326
Análisis de la bibliografía/bibliográfico, 45, 71, 83, 116, 306, xxiii, xxxi
Análisis/estudio métrico, 296, 297, 301, 304, 315, 323, lxxviii, lxxix
Año de publicación, 81, 86, 124, 133, x, xxi, xxv
APA, v, x, xiv, xxi
Apertura, 254
Aplicabilidad, 131, 141, 152, 160, 174, vi, lxxxvi, lxxxviii
Aprendizaje, 27, 50, 52, 56, 58, 59, 61, 62, 67, 70, 166, 261, 277, 289, 290, 300, 305, 317, lxxi
Área formativa, 93, 101, 113, 123, 133, 145, 159, 183, 188, 192, 193, 213, 214, 228, 231, 234, 237, xxxv, xli, xliv, xlvii
Autoeficacia, 33
Autoevaluación, 33, 60
Autoinforme estandarizado, 92, 100, 120, 143, 151, 187, 189, 191, xxxiii, xl, xliii, xlvii
Autoinforme sin estandarizar, 108
Automatizar, 285, 303
- B**
Balanceo, 91, 104, 118, 167, xxxiii
Base de datos, 86, 124, 133, 227, 232, 256, 275, 303, 308, x, xxi
Beneficio, 67, 70, 71, 134, 174, 207, 264, 310, 314, 317, 318, 324, 327, xcii
Burocratización, 31
- C**
Calendarización, 45, 53, 57, 68, 70, 265, 327
Calidad de medida, 172
Campo de formación, 146, 161, 188, 192, 193
Campo de intervención, 86, 89, 93, 101, 114, 123, 133, 159, 183, x, xxi, xxxv, xli, xlvii
Canal/vía de comunicación, 28, 30, 40, 54, 66, 212, 221, 226, 237, 255, 256, 308, 327
Capacitación, 50, 67
Características extrínsecas, 81, 84, 86, 89, 90, 117, x, xxix
Características metodológicas, 96, 98, 102, 111, 135, 139, 141, 186, 188, 191, 192, xi, xxix, xxxiii, xxxix, xlii, xlv
Características sustantivas, 112, 144, 167, 177, 183, 188, 189, 191, 192, 193, x, xxix, xxxiv, xl, xlvii
Certeza, 83, 173, 199, 218, 262, 264, 304, 312, 315
Certificados “con aprovechamiento”, 300
Chat, 255, 327
Checklist (listado de verificación), 263, 264, 265, 267, 320
Clima organizacional, 56
Codificación, 82, 95, 97, 102, 111, 116, 128, 129, 136, 140, 182, 183, 194, 200, 225, 232, 254, 256, 284, 295, 303
Coeficiente de correlación de rango, 251
Coherencia interna, 44, 45, 265
Compromiso, 52, 68
Consolidación, 207, 262, 285, 292, 311, 326
Continuo, 28, 41, 63, 66
Continuo subjetivo/psicológico, 239, 245, 290
Control de covariante, 77
Coordinación, 263, 264, 265, 268, 273, 274, 276, 278, 280, 282, 283, 320

- Correo electrónico, 85, 223, 225, 229, 231, 275, 284, 297, 307, 308, lxxx, lxxxiv
- Coste, 25, 32, 45, 48, 50, 53, 55, 60, 61, 63, 68, 70, 207, 235
- Credibilidad, 30, 32, 33, 77, 167
- Criterio de exclusión, 95, v
- Criterio de inclusión, 47, 91, 94, 116, 117, 124, 125, 126, 127, 133, 140, 170
- Criterio de selección, 91, 96, 99, 103, 118, 141, 183, 186, 188, 191, 192, 194, 197, xxxiii, xxxix, xlii, xlv
- D**
- Datos relevantes, 200
- Datos secundarios, 43, 169, 177
- Definición empírica, 175, 315, 318
- Definición teórica, 177, 315
- Democracia, 254
- Dependencia, 184, 199
- Destinatarios, 56, 93, 101, 114, 123, 146, 188, 192, 213, 214, 220, 221, 222, 224, 225, 235, 254, 274, 275, 283, 323, xlv, xlvii, lxxv
- Desviación típica, 86, 125, 133, 144, 156, 163, 168, 184, 185, 188, 189, 239, 242, 247, 248, x, xxi, xxxiv
- Diferencia estadísticamente significativa, 93, 101, 111, 122, 144, 155, 176, 187, 189, 195, 299, xxxiv, xl, xliii, xlvi
- Diferencia media tipificada, 183, 186
- Dificultad, 295, 296, 297, 315, 320, 323, 324, lxxix
- Difusión, 29, 31, 53, 57, 70, 213, 214, 215, 216, 218, 232, 255, 256, 276, 283, 285, 297, 307, 327
- Discriminación, 296, 298, 320, 323, 324, lxxix
- Diseño con grupo control no equivalente y prueba previa y posterior, 46, 47, 170, 304
- Diseño con medida previa y múltiples medidas posteriores, 312
- Diseño cuasiexperimental, 170, 304, 312, 321, 325
- Diseño de cohortes básico, 47, 170
- Diseño de discontinuidad en la regresión, 47, 48, 129, 170, 177
- Diseño de replicaciones cambiadas, 174
- Diseño de replicaciones de tratamiento múltiple intercambiado, 174
- Diseño de series temporales, 164
- Diseño experimental, 38, 39, 65, 77, 96, 164, 196, 198
- Dispersión discriminante, 239
- Distribución de frecuencias, 239, 290
- Distribución del instrumento, 85, 116
- Distribución discriminante, 239
- Dominio, 80, 81, 85, 89, 116
- Dunn-Rankin (método de), 214, 229, 230, 233, 236, 237, 238, 240, 247, 249, 250, 251, 252, 253
- E**
- Ecuación de regresión, 247, 248, 249
- Efectividad, 38, 48, 65, 127
- Efecto halo, 56
- Eficiencia, 48, 57, 199, 206, 234, 235, 284, 306, 315, xciv
- Emparejamiento previo, 46, 170
- Empowerment, 33, 174
- Entrevistas, 43, 45, 56, 61, 67, 68, 70, 102, 108, 226, 256, 268, 326
- Envío, 116, 223, 225, 231, 256, 275, 283, 284, 297, 307, 309, 327
- Equiprobabilidad, 166
- Equivalencia, 48, 171, 246, 247
- Errores de medida, 56
- Estabilidad del dato, 104, 164, 172, 177
- Estadístico de Levene, 163
- Estandarización, 77, 83, 96, 107, 136, 165, 174, 175, 176, 177
- Estímulo, 43, 102, 230, 238, 239, 240, 243, 245, 249, 251, 253, 254, xcii
- Estructuración de los resultados, 225
- Estudios teóricos, 149, 154, 157, 159, 160, 162, 194

- Ética, 29, 39, 46, 78,
145, 159, 166, 169,
Evaluabilidad, 42, 48,
83, 206
Evaluación externa, 32,
33
Evaluación formativa,
32, 41, 50, 51, 71, 72,
172, 211, 261, 314,
318, 322
Evaluación interna, 32,
33
Evaluación mixta, 33
Evaluación participativa,
33
Evaluación sumativa,
31, 32, 71, 72, 173,
261, 314, 318, 320,
322
Evidencia empírica, 86,
126, 133, 274, x, xxi
Exactitud, 29, 253
Exhaustividad, 255
Extrapolar, 34, 37, 54,
257
- F**
Factibilidad, 29, 253
Feedback, 28, 54, 56, 57,
70, 227, 232, 256, 312,
328
Fiabilidad, 55, 69, 77,
78, 83, 102, 144, 155,
200, 232, 254, 274,
283, 295, 298, 320,
323, 324, lxxix,
Financiación, 31, 56
Foro, 255, 327
- G**
Grados de libertad, 189,
191
Grupo de comparación
activo, 103
Grupo de comparación
inactivo, 103
- H**
Habilidades, 17, 28, 49,
50, 58, 59, 60, 61, 70,
226, 235, 286, 289,
305, 317, xxxv, lxxxv,
lxxxviii, xcii, xciii
Heterogeneidad, 19, 181,
189, 190, 199, 200,
205
Homocedasticidad, 168
- I**
Identificación de los
participantes, 310,
313
Indicadores, 41, 45, 66,
78, 94, lxxi
Índice de congruencia
de Osterlind, 314
Índice de impacto de la
revista, 86, xxi
Índice de variabilidad,
93, 101, 111, 122, 132,
144, 155, 156, 161,
183, 188, 189, xxxiv,
xl, xliv, xlvi
Índice global
cuantitativo, 78
Inferencia de valores
perdidos, 96, 121, 153,
171, 177
Instrucción, 27, 52, 68,
294
Instrumento de medida,
54, 286
Instrumento
normalizado/objetivo,
108
Internet, 84, 95, 116,
140, 160, 222, 223,
225, 226, 227, 229,
230, 231, 232, 237,
254, 255, 256, 276,
283, 297, 301, 307,
308, 312, 327, 328
Interpolación lineal, 171,
172
Intervalo de confianza,
144, 155
Intervalos
aparentemente iguales
(método de los), 20,
236, 240, 245, 246,
247, 248, 249, 251,
252, 253
Intervalos sucesivos
(método de), 20, 213,
214, 229, 233, 236,
240, 241, 242, 243,
245, 247, 248, 249,
250, 251, 252, 253
Investigadores, 86, 124,
133, 196, 200, x, xxi
Irrelevancias, 166
- J**
Juicio de valor, 40
Justificación teórica, 44
- L**
Limitaciones, 51, 53,
174, 199, 250, 253,
314, 318, 319, 320,
321
Límite inferior, 189,
243, 244, 246, 250,
251, 253
Límite superior, 189,
243,
Línea base, 61, 86, 126,
133, 173, v, xi, xxi
Lugar de trabajo, 216,
222, 305
- M**
Mantenimiento
constante, 1
Máquina lectora, 303
Materiales, 25, 44, 45,
50, 52, 56, 57, 59, 82,
262, 265, 274, lxvii,
lxxi, lxxii, lxxv
Media de la serie, 171
Media de puntos
adyacentes, 171
Mediana, 171, 172, 243,
244, 246, 250, 290,
294
Mediana de puntos
adyacentes, 171
Medida posterior, 83,
94, 96, 182, 190, 194,
299

- Medida previa, 61, 172, 173, 183, 255, 292, 293, 302, 304, 312, 321
- Medida retrospectiva, 173, 177
- Meta libre, 38, 65
- Modalidad, 129, 196, 213, 214, 216, 221, 222, 262, 268, 273, 275, 282, 283, 295, 303, lxxv
- Modelo teórico, 175, 211, 219, 224, 232, 274, 320, 322
- Monitorización, 61, 172, 261, 266, lxxxvii, xc
- Mortalidad diferencial, 48, 77, 92, 99, 105, 119, 130, 142, 149, 160, 171, 187, xxxiii, xxxix, xlii, xlv
- Mortalidad global, 91, 92, 96, 99, 105, 119, 130, 142, 148, 149, 161, 164, 167, 171, 177, 187, 191, 192, 200, 321, 325, 326, xxxiii, xxxix, xlii, xlv
- Motivación, 27, 52, 198, 212, 268, 281, 301, 302, lxvii, lxxxiv, lxxxvi, lxxxix
- Muestras independientes, 173
- Multidisciplinar, 34, 40, 64, 78
- N**
- Necesidad comparada, 43
- Necesidad demandada o expresada, 43, 217, 223
- Necesidad experimentada, sentida o percibida, 43
- Necesidad normativa, 43
- Nivel de confianza, 163, 189, 191, 252, 298, 299
- Nivel socioeconómico, 86, 125, 133, x, xxi
- Nuevas tecnologías, 159, 160, 221, 222, 223, 232, 237, 241, 255, 256, 327, xcii
- Número de grupos, 93, 94, 101, 112, 119, 122, 130, 132, 144, 156, 183, 188, 192, 193, xxxiv, xl, xlv, xlvii
- O**
- Objetividad, 30, 65, 107, 172, 211, 232, lxxv
- Observacional, 40, 98, 99, 102, 104, 118, 129, 141, 146, 148, 150, 162, 163, 164, 165, 169, 172, 178, 268, xxxvii
- Observaciones múltiples, 177
- Operativización, 18, 71, 72, 211, 266, 315, 320, 322, 323
- Ordenación por rangos (método de), 20, 236, 240, 247, 249, 250, 251, 252, 253
- Orientación teórica, 83, 86, 125, 133, x, xxi, xxv
- Origen cultural, 86, 125, 133, x, xxi
- P**
- Parcialidad, 30
- Parsimonioso, 211, 236, 250, 253, 256
- Participación de los implicados, 41
- Participativo, 212, 224, 225, 231, 309
- Perspectiva lineal, 71
- Pertinencia, 43
- Piloto, 52, 54, 135, 233, 235, 291, 294, 300, 306, 311, lxxxvii, xc
- Planificación, 17, 25, 27, 33, 35, 38, 42, 45, 55, 56, 64, 65, 69, 72, 135, 211, 222, 262, 284, 290, 300
- Plataforma de teleformación, 268
- Política, 31, 41, 59, 66, 207, 229, 254, 257, vi
- Postest, 50, 58, 60, 61, 104, 129, 171, 189, xxvi, lxxix
- Potencia estadística, 199
- Precisión, 80, 83, 175, 199, 274, 283, 320, 323
- Presencial, 222, 262, 268, 273, 274, 275, 282, 295, 303, lxxv
- Pretest, 50, 58, 60, 61, 104, 129, 171, 300, xxvi
- Primacía, 56, 71
- Proceso mixto, 254
- Programa de formación, 25, 28, 49, 50, 53, 55, 56, 58, 59, 60, 61, 63, 70, 139, 214, 312
- Programa de intervención, 18, 25, 36, 82, 83
- Proponente, 214, 216
- Proporción, 161, 163, 218, 228, 241, 242, 282, 296
- Propósito, 25, 32, 39, 53, 68
- Prototipo, 52
- Prueba de homogeneidad intragrupo, 189
- Prueba de opción múltiple, 295
- Prueba de Scheffé, 164
- Prueba paramétrica, 163
- Prueba situacional, 60

- Pseudoevaluación, 30
Punto de corte, 47, 104, 119, 130, 170, 296, 298
Puntuaciones típicas, 242
- R**
Reacción, 50, 58, 60, 62, 67, 71, 289
Reactividad, 32, 33, 310
Realismo, 254
Recencia, 56
Recogida/recepción de
Recursos humanos, 25, 44, 45, 54, 56, 57, 69, 72, 73, 82, 282
Recursos materiales, 25
Registro de observación, 268
Registro fisiológico, 92, 120, xxxiii, xl, xliii, xlvi
Regresión estadística, 164, 173, 302, 313
Relaciones lineales, 47, 170
Replicabilidad, 199, v
Representatividad, 18, 37, 84, 86, 87, 88, 89, 90, 91, 93, 94, 116, 117, 133, 200, 218, 314, ix, x, xxix
Retorno de la inversión, 62, 70
Revisión bibliográfica, 80, 323, xxxv
Revisión sistemática, 313
Rho, 251, 252
Rigor, 58, 199, 205, 255
Rol, 51, 68
- S**
Sectarismo, 30
Selección de la muestra, 85, 91, 96, 99, 103, 104, 116, 118, 129, 141, 148, 164, 167, 183, 186, 188, 194, 196, xxxiii, xxxix, xlii, xlv
Seminario formativo, 302
Sesgo, 19, 38, 77, 86, 122, 127, 132, 133, 135, 136, 184, 194, 195, 197, 199, 200, 236, xii, xxi
Sesgo de publicación, 122, 132, 194, 195, 197, 200
Significación, 94, 132, 163, 183, 189, 191, 195, 199, 237, 290, 297, 300, 326
Simulaciones, 28, 44
Sistematización, 25
Software, 80, 82, 84, 95, 140, 182, 185, 237, 294, 295, 307, v
Subdominio, 80, 81, 84, 89, 116
Subjetivismo, 31, 232
Suficiencia, 44, 174
- T**
Tamaño de efecto, 82, 83, 86, 88, 92, 96, 100, 110, 122, 131, 143, 154, 161, 165, 168, 177, 182, 183, 186, 187, 189, 190, 192, 193, 194, 195, 196, 197, 198, 304, 321, v, xii, xxi, xxxiv, xl, xliii, xlvi
Tamaño de la muestra, 78, 86, 88, 110, 112, 119, 126, 130, 133, 148, 183, 185, 199, 319, xi, xxi, xxv
Técnica de control de variables extrañas, 39, 99, 104, 118, 129, 139, 141, 167, 169, 186, 188, 192, 196, xi, xlii
Técnica de grupo, 43
Teléfono, 256, 307, 308
Teleformación, 268, 273, 275, 282, 283, 295, 303
Temporalización, 25, 27, 44, 45
Tendencia a valores centrales, 56
Tendencia lineal en el punto, 172
Test adaptativo informatizado (TAI), 303, 320, 323, 324
Test de rendimiento, 64, 72
Tipo de estudio, 19, 99, 102, 118, 129, 134, 141, 146, 162, 163, 183, 186, 191, 192, 196
Tipo de publicación, 86, 89, 93, 101, 114, 123, 133, 146, 160, 183, 188, x, xxi, xxv, xxxv, xli, xliv, xlvii
Toma de decisiones basada en referentes empíricos, 40, 41, 42, 57, 60, 66, 68, 70, 72, 139, 211, 212, 213, 225
Transparencia, 30, 211, 231, 233, 235, 254, 256, 285, 324, 327
- U**
Unidades, 34, 36, 64, 86, 87, 91, 93, 101, 113, 118, 123, 133, 145, 147, 158, 160, 183, 188, 220, 226, x, xi, xxi, xxxiii, xxxv, xli, xliv, xlvii
- V**
Validez de conclusión estadística, 36
Validez de constructo, 32, 36, 37
Validez de contenido, 79, 81, 83, 84, 85, 90,

91, 92, 93, 94, 114,
116, 117, 124, 125,
126, 127, 129, 131,
132, 133, 135, 268,
313, 315, 320, 324,
328, vii, ix, xiii
Validez externa, 34, 35,
36, 37, 78, vi
Validez interna, 34, 35,
36, 37, 78, 104
Valor escalar del
estímulo, 239
Valor perdido, 86, 89,
92, 96, 100, 110, 121,
129, 141, 143, 146,
149, 152, 153, 154,
157, 160, 161, 162,
168, 171, 172, 176,
177, 183, 186, 187,
189, 194, 251, 252,
303, 325, v, xii, xxi,
xxxiv, xl, xli, xliii xliv,
xlvi, xlvii
Valor subjetivo, 239
Variable dependiente no
equivalente, 175
Variable dependiente
normalizada, 86, 88,
92, 100, 107, 120, 143,
151, 165, 175, 177,
183, 187, 191, 193,
198, 200, xi, xxi, xxvi,
xxxiii, xl, xliii, xlvi
Variable moderadora,
19, 80, 178, 181, 182,
186, 190, 192, 198,
199, 200, 327
Viabilidad, 52, 68, 84,
85, 86, 87, 88, 89 90,
91, 93, 116, 117, 133,
135, 314, ix, x, xxix

ANEXOS

ANEXO I

**Ítems que contienen las escalas de medición de la calidad del
diseño más frecuentemente referenciadas**

		CONSORT	TREND	STRICTA	STARD	EDUC.
VISIÓN GENERAL	Capacidad para ser publicado					■
	Aporta nuevos y valiosos resultados					■
	Apto para los lectores en general					■
	Legible					■
TÍTULO Y RESUMEN	Asignación de los participantes	■	■			
	Identificación tipo de diagnóstico	■			■	
	Estructura recomendada APA		■			
INTRODUCCIÓN	Información población diana		■			
	Objetivos del estudio	■	■			■
	Antecedentes: teoría		■		■	■
	Justificación de la intervención			■		
MÉTODO: PARTICIPANTES	Necesidades			■		
	Descrito con suficiente detalle		■	■		■
	Criterios de inclusión	■	■		■	■
	Contexto	■	■		■	■
	Lugar donde el dato fue recogido	■	■		■	
TAMAÑO DE LA MUESTRA	Criterios de exclusión				■	
	Cómo fue determinado	■	■			
	Número participantes cada grupo	■	■			
INSTRUMENTOS	Inscripción: número de participantes analizados		■			
	Descritos con suficiente detalle					■
	Método usado para recoger el dato		■			■
INTERVENCIÓN	Software		■			
	Detalles de los diferentes grupos	■	■		■	
	Número de sesiones			■		
OTRAS INTERVENCIONES	Frecuencia de los tratamientos			■		
	Duración			■		
MÉTODO DE ASIGNACIÓN	Condiciones de la experiencia			■		
	Método aleatorio	■	■			
	Tipo		■			
	Método de implementación	■	■			
	Unidad (individual, en grupo, etc.)		■			
CONTROL VVEE	Cómo minimizar los errores debidos a la no aleatorización		■			
	Ciego: sí/no	■	■			
	Las cuestiones son respondidas	■	■			■
	Evaluación parte de objetivos					■
	Fluctuación de los participantes	■	■			
	Intervalos de confianza				■	
	Variabilidad entre grupos/subgrupos	■	■		■	
	Replicabilidad si se hizo				■	
	Momentos de medida	■	■			
	Periodo de seguimiento	■	■			
	Línea base en cada grupo	■	■			
	Tamaño de efecto	■	■			■
	Otros datos aparte del objetivo	■	■			
RESULTADOS	Eventos adversos	■	■			
	Muerte experimental				■	
	Técnicas para inferir los valores perdidos		■		■	
	Tipo de análisis		■			

	Medidas claramente definidas			
	Calidad de la medida: validez			
	Interpretación resultados/ errores			
	Aplicabilidad			
	Explicaciones alternativas			
	Éxito			
	Implicaciones políticas			
DISCUSIÓN	Generalizabilidad/ validez externa			
	Interpretación general			
	Bien estructurada			
	Puntos fuertes y débiles			
	Mejoras para el futuro			

ANEXO II

Estudio de validez de contenido de ítems para medir la calidad metodológica de las intervenciones

Anexos

***Escala para medir la calidad de los estudios primarios
Cuestionario de validez de contenido (versión castellana)***

Solicitamos tu colaboración completando este cuestionario con el fin de estudiar la validez de contenido de esta escala que pretende medir la calidad de los estudios primarios.

Concretamente, te pedimos que puntúes de -1 (nivel mínimo), 0 (nivel intermedio), a +1 (nivel máximo) cada uno de los ítems propuestos en función de tres aspectos:

1. *Representatividad*: en qué medida cada ítem específico representa el dominio donde se le asignó.
2. *Utilidad*: en qué grado el ítem específico es útil para medir la calidad del estudio respecto al dominio al que se le asignó.
3. *Viabilidad*: en qué medida la información que recoge el ítem procede y está disponible.

Para cada ítem se da la posibilidad de que recojas comentarios. Además, también damos la posibilidad de que incluyas algún otro que consideres importante y que tendría que ser tenido en cuenta.

Innovaciones metodológicas en la evaluación de la formación continua

		Ítem	Representatividad	Utilidad	Viabilidad	Comentarios
CARACTERÍSTICAS EXTRÍNECAS		1- Tipo de publicación (1.Revista, 2.Libro, 3.Tesis, 4.Congreso, 5. Otros)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		2- Año de publicación	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		3- Índice de impacto (en revistas)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		4- Base de datos (especificar)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		5- Entrenamiento de los investigadores (1. Especificado; 2. No hay datos suficientes)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		6- Estructura del artículo recomendada por la APA	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		- <i>Especificar otra posibilidad</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
CARACTERÍSTICAS SUSTANTIVAS	M U E S T R A	7- Edad (rango) referido (sí/ no)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		8- Edad (media)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		9- Edad (desviación típica)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		10- Origen cultural (1. Sólo uno; 2. Más de uno; 3. No hay datos suficientes)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		11- Nivel socioeconómico (1. Bajo; 2. Medio; 3. Alto).	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		- <i>Especificar otra posibilidad</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	C O N T E X T O	12- Contexto de intervención (1.Urbano; 2. Rural; 3. Mixto)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		13- Campo de intervención (1.Sanitario; 2.Educacional; 3.Social; 4.Clínico; 5.Organizacional; 6.Otros)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		14- País	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		- <i>Especificar otra posibilidad</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	T R A T A M I E N T O	15- Orientación teórica (1.Especificada; 2.Inferida; 3.No hay datos suficientes)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		16- Evidencia empírica previa (1. Especificada; 2. No hay datos suficientes)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		17- Periodo de tratamiento	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		18- Intensidad del tratamiento (p. e. número de dosis)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
19- Unidades (1. En grupo; 2. Individual)		(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)		
20- Los puntos fuertes y débiles son discutidos (sí/ no)		(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)		
- <i>Especificar otra posibilidad</i>		(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)		
	21- Criterios de inclusión y exclusión de las unidades de la	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)		

CARACTERÍSTICAS METODOLÓGICAS	muestra explicitados (sí/no)				
	22- Asignación aleatoria de las unidades a los grupos (1.No y sin control de variables extrañas; 2.No pero con control de variables extrañas; 3.Sí)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	23- Tipo de metodología o diseño: 1.Experimental (aleatorio); 2.Cuasi-experimental (dos grupos sin asignación aleatoria o con grupo control no equivalente con pre-test y post-test; 3.Pre-Experimental (sólo un grupo; una medida); 4.Otros (cuestionarios, observación)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	24- Tamaño de la muestra (1. $n \leq 5$; 2. $5 < n < 10$; 3. $n \geq 10$)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	25- Cálculo estadístico del tamaño de la muestra (sí/no) (magnitud del error)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	26- Mortalidad experimental (1. $\leq 30\%$; 2. $> 30\%$)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	27- Sin mortalidad (no/sí)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	28- Mortalidad experimental entre grupos (1. Homogéneo; 2. No homogéneo)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	29- Exclusiones tras la asignación aleatoria (sí/no y número) (por ejemplo, no codificados)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	30- Periodo de línea base (1. < 5 meses; 2. 6-11 meses; 3. > 12 meses)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	31- Periodo de seguimiento (1. < 5 meses; 2. 6-11 meses; 3. > 12 meses)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	32- Momentos de medida (sí/no y número; 1. post intervención; 2. pre y post intervención)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	33- Las medidas del pre-test aparecen en el pos-test (1.ninguna; 2.algunas; 3.todas)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	34- Variables dependientes normalizadas: 1.no hay (autoinformes y medidas a posteriori); 2.Cuestionarios o autoinformes estandarizados; 3.Al menos una es objetiva (medidas psicofisiológicas)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	35- Homogeneidad de la intervención/estudio (1.las personas no reciben el tratamiento en las mismas condiciones contextuales; 2.las personas reciben el tratamiento en las mismas condiciones contextuales)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	36- Técnicas de control: 1.Ciego (de beneficiarios); 2.Ciego (de implementadores); 3.Doble ciego (ambos); 4.Otros (especificar)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

Innovaciones metodológicas en la evaluación de la formación continua

	37- Definición del constructo del resultado (1.Replicable por el lector en su propio contexto; 2.Definición vaga; 3.No definición)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	38- Métodos estadísticos para tratar los valores perdidos (Sí/No)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	39- Especificación de los intervalos de confianza en los análisis estadísticos (Sí/No)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	40- Tamaño de efecto y valor	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	41- Otros datos además de los objetivos marcados (1.Efectos positivos; 2.Efectos negativos; 3.Ambos; 4.Ninguno)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	42- Interpretación de los resultados (1.Todos; 2.Algunos de ellos; 3.Ninguno)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	43- Interpretación de los sesgos en los resultados (1.Todos; 2.Algunos de ellos; 3.Ninguno)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

Anexos

***Scale to measure the quality of primary interventions
Content validity questionnaire (English version)***

We ask for your collaboration to complete this questionnaire in order to assess the content validity of a scale to measure the quality of primary studies.

Specifically we ask you for assessing from -1 (minimum level); 0 (medium level) to +1 (maximum level) each of the following items with respect to:

1. Representativeness: How much the specific item represents the quality subdomain where it is assigned.
2. Utility: How much the specific item is useful to assess the quality of the study with respect to the quality domain where it is assigned.
3. Feasibility: How feasible is that item to code.

Also, in each item we ask you for comments, and to suggest any other item that you think it is important to take into account.

Thanks in advance for your time and collaboration.

		Item	Representative	Useful	Feasible	Comments
EXTRINSIC CHARACTERISTICS		1- Type of publication (1. Journal, 2. Book, 3. Thesis, 4. Congress, 5. Other ones)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		2- Year of publication	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		3- Citation Impact factor for the journal in which an article appeared	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		4- Is the raw data from the study available?	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		5- Training of treatment implementers (1. specified; 2. No data enough)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		6- APA format	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		- Specify another possibility				
SUBSTANTIVE CHARACTERISTICS	S A M P L E	7- The study reports participant age (range) referred: Y/N	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		8- Age (mean)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		9- Age (standard deviation)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		10- Cultural origin (1. Only one; 2. More than one; 3. No data enough)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

SUBSTANTIVE CHARACTERISTICS	S E T T I N G	11- Socioeconomic level (1. Low; 2. Medium; 3. High). - <i>Specify another possibility</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		12- Implementation context (1.Urban; 2. Rural; 3.Mixed)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		13- Intervention Field (1.Inpatient clinical; 2.Educational; 3.Social; 4.Outpatient clinical; 5.Organizational; 6.Others)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		14- The authors report the country in which study was conducted - <i>Specify another possibility</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		15- Theoretical orientation (1.Specified; 2.Inferred; 3.No data enough)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	T R E A T M E N T	16- Previous Empirical Evidence (1. Specified; 2. No data enough)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		17- Period of treatment (quantitative time)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		18- Degree of Treatment Intensity (i.e. number of dosages)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		19- Units (1. In group; 2. Individual)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		20- Strengths and weakness of treatment are discussed (Y/N) - <i>Specify another possibility</i>	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	METHODOLOGICAL CHARACTERISTICS	21- Inclusion and exclusion criteria for units; provided (Y/N)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
		22- Units random assignment (1.None and without control of extraneous variables;	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

METHODOLOGICAL CHARACTERISTICS	2. None but with control of extraneous variables; 3. Yes)				
	23- Methodology or Design: 1. Experimental; randomized; 2. Quasi-experimental (two groups without randomized assignment) non-equivalent control groups with pre-test and post-test; 3. Pre-Experimental (only one group, one measure)/ others (questionnaires/observational/naturalistic)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	24- Sample size (1. $n \leq 15$; 2. $15 < n < 30$; 3. $n \geq 30$)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	25- The authors say they did a power analysis to calculate sample size (Y/N)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	26- Attrition (1. $\leq 30\%$; 2. $> 30\%$)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	27- No attrition occurred (N/Y)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	28- Attrition between groups (1. Homogeneous; 2. Non- homogeneous)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	29- Exclusions after randomization (N/Y) (specify number)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	30- How long were units studied before treatment implementation (1. < 5 months; 2. 6-11 months; 3. > 12 months)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	31- Follow-up period (1. < 5 months; 2. 5-11 months; 3. > 12 months)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
	32- Occasions of measurement on each variable (specify number; 1. Post intervention only; 2. Pre and post intervention)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

METHODOLOGICAL
CHARACTERSTICS

- Specify another possibility				
33- Measures in pre-test appear in post-test (1.None; 2.Some; 3.All of them)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
Item	Representative	Useful	Feasible	Comments
34- Standardized dependent variables: 1.Without (self-reports and post hoc records); 2. Standardized questionnaires or standardized self-reports.	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
35- Intervention context homogeneity (1. Subjects do not receive the treatment in the same contextual conditions; 2. Subjects receive treatment in the same contextual conditions)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
36- Control Techniques: 1.Blind (beneficiaries); 2.Blind (implementers); 3.Both; 4.Other ones (necessary to specify)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
37- Construct Definition of Outcome (1.Replicable by reader in own setting; 2.Vague definition; 3.No definition)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
38- Statistical methods for inputing missing data (Y/N; Specify)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
39- Specification of confidence intervals in statistic analysis (Y/N)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
40- Effect size value	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
41- Effectiveness of treatment (1.Positive effects; 2. Negative effects; 3.Both; 4.None)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
42- Interpretation of results (1.All; 2.Some of them; 3.None)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	
43- Discussion of bias and limitations (1.All; 2.Some of them; 3.None)	(-1) (0) (1)	(-1) (0) (1)	(-1) (0) (1)	

ANEXO III

Procedencia de los ítems de la escala de medición de la calidad de estudios primarios

ÍTEM	PROCEDENCIA
1. Tipo de publicación	Sánchez-Meca, 1997
2. Año de publicación	Sánchez-Meca, 1997
3. Índice de impacto de la revista	<i>Nuevo</i>
4. Base de datos en que se encontró	<i>Nuevo</i>
5. Entrenamiento de los investigadores	Sánchez-Meca, 1997
6. Estructura recomendada por la APA	TREND
7. Rango de edad	<i>Nuevo</i>
8. Media de edad	Sánchez-Meca, 1997
9. Desviación típica de la edad	Chacón, García, Alarcón y Sanduvete, 2003
10. Origen cultural	<i>Nuevo</i>
11. Nivel socioeconómico	<i>Nuevo</i>
12. Contexto de intervención	CONSORT, TREND, STARD, ED.
13. Campo de intervención	Chacón, García, Alarcón y Sanduvete, 2003
14. País	Sánchez-Meca, 1997
15. Orientación teórica	TREND, STARD, EDUC.
16. Evidencia empírica previa	<i>Nuevo</i>
17. Periodo de tratamiento	STRICTA
18. Grado de intensidad del tratamiento	STRICTA
19. Unidades (en grupo o individual)	TREND
20. Los puntos fuertes y débiles son discutidos	EDUC.
21. Criterios de inclusión y exclusión de las unidades de la muestra especificados	CONSORT, TREND, STARD
22. Asignación aleatoria de las unidades a los grupos	TREND
23. Tipo de metodología/ diseño	STRICTA
24. Tamaño de la muestra	CONSORT, TREND
25. Estadístico para calcular el tamaño de la muestra	CONSORT, TREND
26. Mortalidad experimental	STARD
27. Sin mortalidad	<i>Nuevo</i>
28. Mortalidad experimental entre grupos	<i>Nuevo</i>
29. Exclusiones posteriores a la asignación aleatoria	<i>Nuevo</i>
30. Periodo de línea base	CONSORT, TREND
31. Periodo de seguimiento	CONSORT, TREND
32. Momentos de medida	CONSORT, TREND
33. Las medidas del pre-test aparecen en el pos-test	Chacón, García, Alarcón y Sanduvete, 2003
34. Variables dependientes normalizadas	Chacón, García, Alarcón y Sanduvete, 2003
35. Homogeneidad de la intervención	Chacón, García, Alarcón y Sanduvete, 2003
36. Técnicas de control	CONSORT
37. Definición del constructo	<i>Nuevo</i>
38. Métodos estadísticos para inferir los valores perdidos	TREND, STARD
39. Especificación de los intervalos de confianza en los análisis estadísticos	STARD
40. Tamaño de efecto y valor	CONSORT, TREND, EDUC.
41. Otros datos aparte de los objetivos marcados	CONSORT, TREND
42. Interpretación de los resultados	CONSORT, TREND
43. Interpretación de los sesgos de los resultados	CONSORT, TREND

ANEXO IV

Resultados obtenidos en el primer análisis bibliográfico exploratorio para determinar las características que solían presentar los estudios primarios en los ámbitos psicológico, social y educacional

CATEGORÍA	PORCENTAJE n = 1899
1. Tipo de publicación: a. Revista b. Libro c. Tesis d. Congreso e. Otros	87.5% 3.9% 5.5% 0.8% 2.3%
2. Año de publicación: a. 1967-1989 b. 1990-1999 c. 2000-2005	2.69% 16.85% 80.46%
3. Orientación teórica: a. Especificada b. Inferida c. No hay datos	17.8% 67.2% 15%
4. Contexto de intervención: a. Sanitario b. Educativo c. Comunitario d. Clínico e. Organizacional f. Otros	5.6% 27.6% 21.9% 24.1% 4% 16.8%
5. Edad especificada: a. Sí b. No	17.8% 82.2%
6. Media de edad: a. 0-15 b. 16-30 c. 31-45 d. 46-60 e. 61-75 f. 76 ó más	47.55% 20.38% 19.24% 7.55% 4.53% 0.75%
7. Tipo de localidad: a. Urbano b. Rural c. Varios d. No hay datos	8.6% 3.7% 66.5% 21.2%
8. Aleatorización de usuarios a grupos: a. No y sin control b. No pero con control c. Sí d. No hay datos	1.3% 63.4% 15.7% 19.6%
9. Metodología/ diseño: a. Experimental-aleatorio b. Cuasi-experimental c. Pre-experimental d. No hay datos suficientes	15.7% 33% 32.2% 19.1%
10. Tamaño de la muestra: a. $n < 5$ b. $5 \leq n \leq 10$ c. $n > 10$ d. No hay datos suficientes	1.6% 4% 65.1% 29.3%
11. Mortalidad experimental: a. $\leq 30\%$ b. $> 30\%$ c. Sin mortalidad d. No hay datos suficientes	12.2% 57.5% 1.8% 28.5%

12. Seguimiento:	
a. < 6 meses	54.7%
b. 6-11 meses	0.9%
c. >12 meses	15.3%
d. No hay datos suficientes	29.1%
13. Momentos de medida:	
a. Posterior	48.7%
b. Pre/ posterior	31.4%
c. No hay datos suficientes	19.9%
14. Todas las medidas del pretest aparecen en el postest:	
a. No	1.1%
b. Algunas	25.4%
c. Todas	5.3%
d. No hay datos suficientes/ no procede	68.2%
15. Variables dependientes normalizadas:	
a. No estándar	2.1%
b. Semiestándar	71.9%
c. Estándar	6.4%
d. No hay datos suficientes	19.6%
16. Homogeneidad de intervención en las distintas condiciones:	
a. Iguales	78.1%
b. Distintos	3.1%
c. No hay datos suficientes	18.8%
17. Técnica de control:	
a. Ciego en usuarios	0.6%
b. Ciego en implementadores	2.1%
c. Ciego en ambos	6.2%
d. Otros	72.9%
e. No hay datos suficientes	18.2%
18. Tamaño del efecto y valor:	
a. Sí	17%
b. No	83%
19. Zona geográfica:	
a. Oceanía	9.8%
b. Europa comunitaria	21.7%
c. Sudamérica	0.4%
d. Canadá	2.4%
e. Estados Unidos	56.5%
f. Asia	1.3%
g. África	2.4%
h. Europa extracomunitaria	0.1%
i. Rusia	0.3%
j. No hay datos suficientes	5.1%

ANEXO V

Ítems que obtuvieron índices de congruencia iguales o mayores que 0.5 en distintas combinaciones de los aspectos/constructos valorados

ANEXO VI

**Resultados obtenidos en el segundo análisis bibliográfico
exploratorio para determinar las características que solían
presentar los estudios primarios en el ámbito de la formación
continua**

ÍTEM	N=95
CARACTERÍSTICAS METODOLÓGICAS	
1. Grupo control: <ul style="list-style-type: none"> ▪ Inactivo (0) ▪ Activo (1) 	71.43% 28.57%
2. Criterios de selección de la muestra (inclusión/exclusión): <ul style="list-style-type: none"> ▪ No especificados (0) ▪ Especificados (1) 	81.4% 18.6%
3. Azar: <ul style="list-style-type: none"> ▪ Pre-experimentales y cuasiexperimentales (sin control de vvee) (0) ▪ Cuasiexperimentales con control de vvee (balanceo, bloqueo, estratificación) (0.5) ▪ Experimentales (asignación aleatoria de las unidades a grupos) (1) 	52% 36% 12%
4. Tipo de metodología/ diseño: <ul style="list-style-type: none"> ▪ Pre-experimental (0) ▪ Cuasi-experimental (0.5) ▪ Serie temporal (obs. pre ≥ 30 y post ≥ 30) y discontinuidad en regresión (0.75) ▪ Experimental-aleatorio (1) 	45.1% 43.1% 0% 11.8%
5. Muestra: <ul style="list-style-type: none"> ▪ n<12 (0) ▪ n=[12-40] (0.5) ▪ n>40 (1) 	2% 9.8% 88.2%
6. Mortalidad global: <ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) ▪ $0 < N < 20\%$ (0.5) ▪ 0% (1) 	50% 33.3% 16.7%
7. Mortalidad diferencial: <ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) ▪ $0 < N < 20\%$ (0.5) ▪ 0% (1) 	100% 0% 0%
8. Exclusiones posteriores a la asignación aleatoria: <ul style="list-style-type: none"> ▪ $\geq 20\%$ (0) ▪ $0 < N < 20\%$ (0.5) ▪ 0% (1) 	80% 0% 20%
9. Seguimiento: <ul style="list-style-type: none"> ▪ No se da seguimiento (0) ▪ < 6 meses (0.3) ▪ [6-11] meses (0.6) ▪ ≥ 12 meses (1) 	63.8% 6.4% 14.9% 14.9%
10. Momentos de medida: <ul style="list-style-type: none"> ▪ Posterior (0) ▪ Previo y posterior (1) 	54.7% 45.3%
11. Las medidas del pre-test aparecen en el pos-test: <ul style="list-style-type: none"> ▪ Falta más de uno (0) ▪ Falta 1 (0.5) ▪ Todas aparecen en todos los momentos (1) 	0% 0% 100%
12. Variables dependientes normalizadas: <ul style="list-style-type: none"> ▪ Sólo autoinformes sin estandarizar (0) ▪ Ninguno normalizado, pero al menos uno es cuestionario o autoinforme estandarizado (0.5) ▪ Al menos un instrumento es objetivo o normalizado (p. e. registro fisiológico, pruebas baremadas) (1) 	2% 84% 14%
13. Enmascaramiento del evaluador: <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) 	50% 50%
14. Enmascaramiento del usuario: <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) 	100% 0%
15. Enmascaramiento del profesional que realiza la intervención: <ul style="list-style-type: none"> ▪ No (0) 	100%

<ul style="list-style-type: none"> ▪ Sí (1) 	0%
16. Homogeneidad de la intervención: <ul style="list-style-type: none"> ▪ No todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales (0) ▪ Todos los sujetos han recibido la misma intensidad de intervención, duración y profesionales (1) 	12% 88%
17. Definición del constructo: <ul style="list-style-type: none"> ▪ No (0) ▪ No operativa (0,5) ▪ Operativa (1) 	5.6% 64.2% 30.2%
18. Métodos estadísticos para inferir los valores perdidos: <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) 	96.3% 3.7%
19. Tamaño de efecto y valor: <ul style="list-style-type: none"> ▪ No (0) ▪ Sí (1) 	97.8% 2.2%
20. Índice de calidad: <ul style="list-style-type: none"> ▪ [0-5] ▪ [6-10] ▪ [11-15] ▪ [15-19] 	75.79% 24.21% 0% 0%
21. Índice estadístico calculado: <ul style="list-style-type: none"> ▪ α Cronbach ▪ t / F ▪ Descriptivos ▪ Regresión de Poisson ▪ Chi cuadrado ▪ Análisis factorial ▪ Generalizabilidad ▪ Modelos de ecuaciones estructurales ▪ Z de Wilcoxon ▪ U de Mann-Whitney ▪ Otros 	4.8% 16.8% 38.7% 2.4% 7.2% 2.4% 2.4% 2.4% 2.4% 2.4% 18.1%
22. Los resultados mostraron diferencias estadísticamente significativas: <ul style="list-style-type: none"> ▪ No ▪ Sí 	36% 64%
23. Índice de variabilidad facilitado: <ul style="list-style-type: none"> ▪ Desviación típica ▪ Varianza ▪ Error estándar 	84.2% 5.3% 10.5%
CARACTERÍSTICAS SUSTANTIVAS	
24. Número de participantes en cada grupo: <ul style="list-style-type: none"> ▪ [0-100] ▪ [101-200] ▪ [201-300] ▪ [301-400] ▪ [401-500] ▪ ≥ 501 	52.6% 9.2% 9.2% 6.6% 5.3% 17.1%
25. Número de grupos en el estudio: <ul style="list-style-type: none"> ▪ 1 ▪ 2 ▪ [3-10] 	65.4% 20.6% 14%
26. Exclusiones tras medidas posteriores: <ul style="list-style-type: none"> ▪ No ▪ Sí 	15% 85%
27. Rango de edad especificado: <ul style="list-style-type: none"> ▪ No ▪ Sí 	80.9% 19.1%
28. Media de edad: <ul style="list-style-type: none"> ▪ [28-32] ▪ [33-37] 	27.27% 36.37%

Anexos

<ul style="list-style-type: none"> ▪ [38-42] ▪ [43-45] 	<p>9.09%</p> <p>27.27%</p>
<p>29. Periodo de tratamiento:</p> <ul style="list-style-type: none"> ▪ ≤ 6 meses ▪ > 6 meses 	<p>82.4%</p> <p>17.6%</p>
<p>30. Intensidad del tratamiento: número de sesiones a la semana</p> <ul style="list-style-type: none"> ▪ [1-5] horas ▪ [6-10] horas ▪ [11-15] horas ▪ [16-20] horas ▪ [21-75] horas ▪ [1-5] días ▪ [6-10] días ▪ [11-15] días 	<p>15.79%</p> <p>15.79%</p> <p>5.26%</p> <p>10.53%</p> <p>10.53%</p> <p>5.26%</p> <p>31.58%</p> <p>5.26%</p>
<p>31. Unidades de intervención</p> <ul style="list-style-type: none"> ▪ Individual ▪ Grupal 	<p>14.8%</p> <p>85.2%</p>
<p>32. Área formativa</p> <ul style="list-style-type: none"> ▪ Habilidades sociales ▪ Prevención de riesgos laborales ▪ Personas mayores ▪ Elaboración de instrumentos ▪ Evaluación de la formación ▪ Formación ▪ Clínica ▪ Otros 	<p>11%</p> <p>6.6%</p> <p>8.8%</p> <p>8.8%</p> <p>26.3%</p> <p>19.8%</p> <p>11%</p> <p>7.7%</p>
<p>33. Campo de intervención</p> <ul style="list-style-type: none"> ▪ Salud ▪ Recursos humanos ▪ Sector secundario ▪ Educación ▪ Personas mayores ▪ Comunitaria ▪ Evaluación ▪ Revisión bibliográfica ▪ Sector terciario 	<p>33.4%</p> <p>7.06%</p> <p>7.06%</p> <p>12.12%</p> <p>8.07%</p> <p>6.06%</p> <p>18.17%</p> <p>1%</p> <p>7.06%</p>
<p>34. Tipo de publicación:</p> <ul style="list-style-type: none"> ▪ Revista ▪ Libro ▪ Tesis ▪ Congreso ▪ Otros 	<p>91.6%</p> <p>8.4%</p> <p>0%</p> <p>0%</p> <p>0%</p>

ANEXO VII

Escala de calidad utilizada para el análisis de las características de los programas de formación continua (versiones adaptadas para los diseños observacionales y encuestas, los cuasiexperimentales y los experimentales)

SISTEMA DE CATEGORÍAS ADAPTADO PARA LOS DISEÑOS OBSERVACIONALES Y DE ENCUESTA

ÍTEM	VALOR	CATEGORÍAS
CARACTERÍSTICAS METODOLÓGICAS		
1. Grupo de comparación	0	No
	1	Sí
	8	Ausencia de dato
	9	No aplicable (por el objeto de estudio)
2. Criterios de selección de la muestra (inclusión/exclusión)	0	No especificados
	1	Especificados
	9	No aplicable (por el objeto de estudio)
3. Azar	0	Selección de la muestra no aleatoria
	1	Selección aleatoria de la muestra
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
4. Diseño	0	1-2 medidas
	0.5	3-29 medidas
	1	30 ó más medidas (concretar valor)
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
5. Muestra	0	$N < 12$
	0.5	$12 \leq N \leq 40$
	1	$N > 40$
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
6. Mortalidad global (porcentaje de quienes empiezan el estudio, pero no lo concluyen)	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
7. Mortalidad diferencial	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un grupo)
8. Exclusiones posteriores a la agrupación de la muestra	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un grupo)
9. Seguimiento: tiempo de medida	0	Nada
	0.3	<6 meses
	0.6	[6-11] meses
	1	≥ 12 meses
	8	No se especifica el dato
10. Momentos de medida	0	Una medida
	1	Más de una medida
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
11. Medidas que aparecen en todos los momentos de registro	0	Más de una medida no aparece en todos los momentos de registro: también cuando ocurre esto en la única variable que se mide
	0.5	Una medida no aparece en todos los momentos de registro (siempre que se mida más de una variable)
	1	Todas las medidas son tomadas en todos los momentos de registro
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un momento de registro)

12. Variables dependientes normalizadas (uso de instrumentos normalizados)	0	Sólo autoinformes sin estandarizar (se considerarán estandarizados sólo cuando así se explicita)
	0.5	Ninguno normalizado pero al menos uno es cuestionario o autoinforme estandarizado (dicho explícitamente)
	1	Al menos uno es objetivo o normalizado; ej: registro fisiológico, registros de observación, pruebas baremadas y/o validadas
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
13. Enmascaramiento del evaluador (codificador)	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
14. Enmascaramiento del usuario	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
15. Enmascaramiento del evaluador interno	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
16. Homogeneidad en el proceso de registro	0	No mismo proceso, intensidad (nº sesiones), duración y/o profesionales
	1	Mismos proceso, intensidad, duración y profesionales
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
17. Definición de los constructos	0	Ninguno definido conceptual y/o empíricamente
	0.5	Al menos 1 definido conceptual y/o empíricamente
	1	Todos definidos conceptual y empíricamente
	8	No se especifica el dato
18. Métodos estadísticos para inferir los valores perdidos	0	Ninguno. Sólo se analizaron los datos completos
	1	Sí o análisis "por intención de tratar" (explicitar método)
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; no se dieron valores perdidos)
19. Tamaño de efecto y valor	0	No se especifica
	1	Se especifica (o índices derivables). Apuntar el índice concreto ; ej: d, Odds Ratio, media, desviación y n (los 3)
	9	No aplicable (por el objeto de estudio; sólo una medida en un grupo)
20. Índice de calidad	SUMA	De 0 a 19
21. Índice estadístico calculado	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
22. Diferencias estadísticamente significativas entre medidas (explicitar qué se compara)	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
23. Índice de variabilidad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo	Anotar valor concreto (la suma ha de dar el valor de la muestra)	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
25. Número de grupos en el estudio	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)

26. Exclusiones tras medidas posteriores	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
27. Rango de edad especificado	1	No se especifica
	2	Sí se especifica
	9	No aplicable (por el objeto de estudio)
28. Media de edad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
30. Intensidad del registro	Nº de horas/ periodo de estudio	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
31. Unidades de registro	1	Individual (o grupo considerado como unidad)
	2	Grupo
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
32. Área formativa	Anotar valor concreto	
33. Campo de intervención	Anotar valor concreto	
34. Tipo de publicación	1	Revista
	2	Libro
	3	Tesis
	4	Congreso
	5	Otras publicaciones
	6	Trabajos no publicados

*Los ítems en casillas sombreadas son aquellos que puntúan para el cálculo del índice global de calidad.

Los valores 8 y 9 no puntúan por tratarse de valores perdidos.

SISTEMA DE CATEGORÍAS ADAPTADO PARA LOS DISEÑOS CUASIEXPERIMENTALES

ÍTEM	VALOR	CATEGORÍAS
CARACTERÍSTICAS METODOLÓGICAS		
1. Grupo de comparación	0	No
	0.5	Inactivo (sin formación o formación tradicional)
	1	Activo (formación alternativa)
	8	Ausencia de dato
	9	No aplicable (por el objeto de estudio)
2. Criterios de selección de la muestra (inclusión/exclusión)	0	No especificados
	1	Especificados
	8	Ausencia de dato
	9	No aplicable (por el objeto de estudio)
3. Azar	0	Sin control de variables extrañas ni delimitación de los criterios de inclusión y asignación
	1	Con control de variables extrañas o delimitación de los criterios de inclusión y asignación
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
4. Diseño	0	Pre-experimental: pre-post sin grupo control
	0.5	Pre-post con grupo control no equivalente con medidas entre 2 y 29
	0.75	Series temporales (30 o más medidas en pre y post conjuntamente)
	1	Discontinuidad en la regresión
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
5. Muestra	0	$N < 12$
	0.5	$12 \leq N \leq 40$
	1	$N > 40$
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
6. Mortalidad global (porcentaje de quienes empiezan el estudio, pero no lo concluyen)	0	$\geq 20\%$
	0.5	$0 < N < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
7. Mortalidad diferencial	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; un solo grupo)
8. Exclusiones posteriores a la asignación de la muestra a las distintas condiciones	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo grupos)
9. Seguimiento: tiempo de medida una vez finalizada la intervención	0	Nada
	0.3	<6 meses
	0.6	[6-11] meses
	1	≥ 12 meses
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
10. Momentos de medida	0	Sólo posterior
	1	Previo y posterior
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)

11. Medidas que aparecen en todos los momentos de registro	0	Más de una medida no aparece en todos los momentos de registro; también cuando ocurre esto en la única variable que se mide
	0.5	Una medida no aparece en todos los momentos de registro (siempre que se mida más de una variable)
	1	Todas las medidas son tomadas en todos los momentos de registro
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un momento de registro)
12. Variables dependientes normalizadas (uso de instrumentos normalizados)	0	Sólo autoinformes sin estandarizar (se considerarán estandarizados sólo cuando así se explicita)
	0.5	Ninguno normalizado pero al menos uno es cuestionario o autoinforme estandarizado (dicho explícitamente)
	1	Al menos uno es objetivo o normalizado; ej: registro fisiológico, registros de observación, pruebas baremadas y/o validadas
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
13. Enmascaramiento del evaluador (codificador)	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
14. Enmascaramiento del usuario	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
15. Enmascaramiento del formador	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
16. Homogeneidad de intervención	0	No mismos proceso, intensidad (nº sesiones), duración y/o profesionales
	1	Mismos proceso, intensidad, duración y profesionales
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
17. Definición de los constructos en referencia a la/s variable/s independiente/s y la/s dependiente/s	0	Ninguno definido conceptual y/o empíricamente
	0.5	Al menos 1 definido conceptual y/o empíricamente
	1	Todos definidos conceptual y empíricamente
	8	No se especifica el dato
18. Métodos estadísticos para inferir los valores perdidos	0	Ninguno. Sólo se analizaron los datos completos
	1	Sí o análisis "por intención de tratar" (explicitar método)
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; no se dieron valores perdidos)
19. Tamaño de efecto y valor	0	No se especifica
	1	Se especifica (o índices derivables). Apuntar el índice concreto ; ej: d, Odds Ratio, media, desviación y n (los 3)
	9	No aplicable (por el objeto de estudio; sólo una medida en un grupo)
20. Índice de calidad	SUMA	De 0 a 19
21. Índice estadístico calculado	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
22. Diferencias estadísticamente significativas entre medidas (explicitar qué se compara)	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)

23. Índice de variabilidad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo	Anotar valor concreto (la suma ha de dar el valor de la muestra)	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
25. Número de grupos en el estudio	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
26. Exclusiones tras medidas posteriores	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
27. Rango de edad especificado	1	No se especifica
	2	Sí se especifica
	9	No aplicable (por el objeto de estudio)
28. Media de edad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
29. Periodo de estudio	1	≤ 6 meses
	2	> 6 meses
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
30. Intensidad de la intervención	nº de horas/ periodo de tiempo	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
31. Unidades de registro	1	Individual (o grupo considerado como unidad)
	2	Grupal
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
32. Área formativa	Anotar valor concreto	
33. Campo de intervención: destinatarios	Anotar valor concreto	
34. Tipo de publicación	1	Revista
	2	Libro
	3	Tesis
	4	Congreso
	5	Otras publicaciones
	6	Trabajos no publicados

*Los ítems en casillas sombreadas son aquellos que puntúan para el cálculo del índice global de calidad.

Los valores 8 y 9 no puntúan por tratarse de valores perdidos.

SISTEMA DE CATEGORÍAS ADAPTADO PARA LOS DISEÑOS EXPERIMENTALES

ÍTEM	VALOR	CATEGORÍAS
CARACTERÍSTICAS METODOLÓGICAS		
1. Grupo control	0	No
	0.5	Inactivo (sin formación o formación tradicional)
	1	Activo (formación alternativa)
	8	Ausencia de dato
	9	No aplicable (por el objeto de estudio)
2. Criterios de selección de la muestra (inclusión/exclusión)	0	No especificados
	1	Especificados
	9	No aplicable (por el objeto de estudio)
3. Azar	0	Asignación aleatoria de las personas a los grupos y/o selección aleatoria de la muestra inadecuados
	1	Asignación aleatoria de las personas a los grupos y selección aleatoria de la muestra adecuados
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
4. Diseño	0	Un solo momento de medida
	1	Al menos dos momentos de medida
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
5. Muestra	0	$N < 12$
	0.5	$12 \leq N \leq 40$
	1	$N > 40$
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
6. Mortalidad global (porcentaje de quienes empiezan el estudio, pero no lo concluyen)	0	$\geq 20\%$
	0.5	$0 < N < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
7. Mortalidad diferencial	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un grupo)
8. Exclusiones posteriores a la asignación de la muestra a las distintas condiciones	0	$\geq 20\%$
	0.5	$0 < n < 20\%$
	1	0%
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un grupo)
9. Seguimiento: tiempo de medida una vez finalizada la intervención	0	Nada
	0.3	<6 meses
	0.6	[6-11] meses
	1	≥ 12 meses
	8	No se especifica el dato
10. Momentos de medida	0	Sólo posterior
	1	Previo y posterior
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
11. Medidas que aparecen en todos los momentos de registro	0	Más de una medida no aparece en todos los momentos de registro; también cuando ocurre esto en la única variable que se mide
	0.5	Una medida no aparece en todos los momentos de registro (siempre que se mida más de una variable)
	1	Todas las medidas son tomadas en todos los momentos de registro
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; sólo un momento de registro)

12. Variables dependientes normalizadas (uso de instrumentos normalizados)	0	Sólo autoinformes sin estandarizar (se considerarán estandarizados sólo cuando así se explicita)
	0.5	Ninguno normalizado pero al menos uno es cuestionario o autoinforme estandarizado (dicho explícitamente)
	1	Al menos uno es objetivo o normalizado; ej: registro fisiológico, registros de observación, pruebas baremadas y/o validadas
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
13. Enmascaramiento del evaluador (codificador)	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
14. Enmascaramiento del usuario	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
15. Enmascaramiento del formador	0	No: conoce la hipótesis del estudio
	1	Sí: desconoce la hipótesis del estudio
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
16. Homogeneidad de intervención o proceso de registro	0	No mismos proceso, intensidad (nº sesiones), duración y/o profesionales
	1	Mismos proceso, intensidad, duración y profesionales
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
17. Definición de los constructos en referencia a la/s variable/s independiente/s y la/s dependiente/s	0	Ninguno definido conceptual y/o empíricamente
	0.5	Al menos 1 definido conceptual y/o empíricamente
	1	Todos definidos conceptual y empíricamente
	8	No se especifica el dato
18. Métodos estadísticos para inferir los valores perdidos	0	Ninguno. Sólo se analizaron los datos completos
	1	Sí o análisis “por intención de tratar” (explicitar método)
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio, no se dieron valores perdidos)
19. Tamaño de efecto y valor	0	No se especifica
	1	Se especifica (o índices derivables). Apuntar el índice concreto ; ej: d, Odds Ratio, media, desviación y n (los 3)
	9	No aplicable (por el objeto de estudio; sólo una medida en un grupo)
20. Índice de calidad	SUMA	De 0 a 19
21. Índice estadístico calculado	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
22. Diferencias estadísticamente significativas entre medidas (explicitar qué se compara)	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
23. Índice de variabilidad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)

CARACTERÍSTICAS SUSTANTIVAS		
24. Número de participantes en cada grupo	Anotar valor concreto (la suma ha de dar el valor de la muestra)	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
25. Número de grupos en el estudio	Anotar valor concreto (la suma ha de dar el valor de la muestra)	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
26. Exclusiones tras medidas posteriores	1	No
	2	Sí
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
27. Rango de edad especificado	1	No se especifica
	2	Sí se especifica
	9	No aplicable (por el objeto de estudio)
28. Media de edad	Anotar valor concreto	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio; una sola medida)
29. Periodo de estudio	1	≤ 6 meses
	2	> 6 meses
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
30. Intensidad del tratamiento	Nº horas/ periodo de tiempo	
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
31. Unidades de intervención o de registro	1	Individual (o grupo considerado como unidad)
	2	Grupal
	8	No se especifica el dato
	9	No aplicable (por el objeto de estudio)
32. Área formativa	Anotar valor concreto	
33. Campo de intervención: destinatarios	Anotar valor concreto	
34. Tipo de publicación	1	Revista
	2	Libro
	3	Tesis
	4	Congreso
	5	Otras publicaciones
	6	Trabajos no publicados

* Los ítems en casillas sombreadas son aquéllos que puntúan para el cálculo del índice global de calidad. Los valores 8 y 9 no puntúan por tratarse de valores perdidos.

ANEXO VIII

Resultados obtenidos en la priorización de acciones formativas en Diputación de Sevilla: comparación entre métodos

PRIORIZACIÓN EN DIPUTACIÓN DE SEVILLA 2004 (N=9)				
ÁREA A	DUNN-RANKIN	ORDEN. RANGOS	INT. SUCESIVOS	INT. APAR. IGUALES
Curso A	1 (56.41)	1 (22)	1 (1.056)	1 (2)
Curso B	1 (56.41)	1 (22)	2 (0.939)	2 (1.833)
Curso C	4 (43.59)	4 (17)	3 (0.352)	3 (1)
Curso D	3 (48.718)	3 (19)	3 (0.352)	3 (1)
ÁREA B				
Curso A	1 (66.667)	1 (26)	1 (1.169)	1 (2.125)
Curso B	1 (66.667)	1 (26)	1 (1.169)	1 (2.125)
Curso C	3 (46.254)	3 (18)	3 (0.362)	3 (1)
Curso D	4 (33.333)	4 (13)	4 (*)	4 (0.643)
ÁREA C				
Curso A	7 (62.906)	7 (368)	10 (1.414)	10 (45)
Curso B	21 (49.402)	21 (289)	21 (1.172)	21 (37)
Curso C	4 (65.983)	4 (386)	1 (1.754)	1 (55)
Curso D	14 (57.607)	14 (337)	17 (1.296)	17 (41)
Curso E	6 (64.103)	6 (375)	3 (1.658)	3 (52.25)
Curso F	1 (76.068)	1 (445)	9 (1.422)	9 (45.25)
Curso G	32 (41.709)	32 (244)	45 (0.623)	45 (20)
Curso H	39 (37.436)	39 (219)	31 (0.904)	31 (28)
Curso I	2 (69.402)	2 (406)	4 (1.648)	4 (52)
Curso J	28 (45.812)	28 (268)	24 (1.015)	24 (32)
Curso K	14 (57.607)	14 (337)	10 (1.414)	10 (45)
Curso L	34 (41.197)	34 (241)	40 (0.718)	40 (23)
Curso M	9 (61.538)	9 (360)	15 (1.381)	15 (44)
Curso N	30 (43.761)	30 (256)	22 (1.05)	22 (33)
Curso O	11 (61.026)	11 (357)	7 (1.479)	7 (47)
Curso P	3 (68.376)	3 (400)	10 (1.414)	10 (45)
Curso Q	17 (54.359)	17 (318)	23 (1.024)	23 (32.25)
Curso R	35 (41.026)	35 (240)	25 (0.981)	25 (31)
Curso S	50 (32.137)	50 (188)	58 (0.365)	58 (12.25)
Curso T	5 (65.299)	5 (382)	5 (1.561)	5 (49.25)
Curso U	26 (46.325)	26 (271)	37 (0.751)	37 (24)
Curso V	32 (41.709)	32 (244)	28 (0.973)	28 (30.75)
Curso W	25 (47.692)	25 (279)	18 (1.232)	18 (39)
Curso X	22 (48.547)	22 (284)	16 (1.372)	16 (43.75)
Curso Y	10 (61.368)	10 (359)	8 (1.446)	8 (46)
Curso Z	12 (59.316)	12 (347)	10 (1.414)	10 (45)
Curso AA	49 (32.308)	49 (189)	35 (0.786)	35 (25)
Curso AB	46 (33.504)	46 (196)	48 (0.518)	48 (17)
Curso AC	62 (23.419)	62 (137)	51 (0.451)	51 (15)
Curso AD	36 (40.855)	36 (239)	32 (0.878)	32 (28)
Curso AE	56 (30.085)	56 (176)	43 (0.654)	43 (21)
Curso AF	38 (40.171)	38 (235)	39 (0.725)	39 (23.25)
Curso AG	52 (31.453)	52 (184)	47 (0.525)	47 (17.25)
Curso AH	54 (30.94)	54 (181)	45 (0.623)	45 (20)
Curso AI	64 (20.513)	64 (120)	53 (0.419)	53 (14)
Curso AJ	65 (12.479)	65 (73)	64 (0.185)	64 (6.75)
Curso AK	58 (26.496)	58 (155)	59 (0.356)	59 (12)
Curso AL	46 (33.504)	46 (196)	61 (0.323)	61 (11)
Curso AM	45 (34.359)	45 (201)	48 (0.518)	48 (17)
Curso AN	59 (25.299)	59 (148)	50 (0.486)	50 (16)
Curso AO	57 (28.376)	57 (166)	62 (0.253)	62 (9)
Curso AP	13 (57.778)	13 (338)	14 (1.39)	14 (44.25)
Curso AQ	48 (32.65)	48 (191)	53 (0.419)	53 (14)
Curso AR	63 (23.07)	63 (135)	55 (0.413)	55 (13.75)
Curso AS	42 (35.043)	42 (205)	36 (0.778)	36 (24.75)
Curso AT	41 (36.068)	41 (211)	57 (0.381)	57 (12.75)

Innovaciones metodológicas en la evaluación de la formación continua

Curso AU	37 (40.513)	37 (237)	59 (0.356)	59 (12)
Curso AV	22 (48.547)	22 (284)	19 (1.223)	19 (38.75)
Curso AW	29 (44.444)	29 (260)	20 (1.195)	20 (37.75)
Curso AX	24 (48.863)	24 (280)	25 (0.981)	25 (31)
Curso AY	44 (34.872)	44 (204)	41 (0.71)	41 (22.75)
Curso AZ	31 (43.59)	31 (255)	34 (0.82)	34 (26)
Curso BA	27 (46.154)	27 (270)	29 (0.948)	29 (30)
Curso BB	19 (50.085)	19 (293)	33 (0.848)	33 (27)
Curso BC	7 (62.906)	7 (368)	2 (1.684)	2 (53)
Curso BD	53 (31.282)	53 (183)	51 (0.451)	51 (15)
Curso BE	16 (56.068)	16 (328)	6 (1.552)	6 (49)
Curso BF	40 (37.094)	40 (217)	37 (0.751)	37 (24)
Curso BG	51 (31.624)	51 (185)	42 (0.686)	42 (22)
Curso BH	42 (35.043)	42 (205)	43 (0.654)	43 (21)
Curso BI	19 (50.083)	19 (293)	25 (0.981)	25 (31)
Curso BJ	61 (23.59)	61 (138)	55 (0.413)	55 (13.75)
Curso BK	60 (23.932)	60 (140)	62 (0.253)	62 (9)
Curso BL	18 (53.162)	18 (311)	29 (0.948)	29 (30)
Curso BM	54 (30.94)	54 (181)	65 (0.015)	65 (1)
ÁREA E				
Curso A	2 (71.111)	2 (64)	2 (*)	2 (1.083)
Curso B	1 (76.667)	1 (69)	1(3.738)	1 (5)
ÁREA F				
Curso A	1 (74.444)	1 (67)	1 (0)	1 (1.5)
Curso B	2 (73.333)	2 (66)	2 (*)	2 (1.25)
ÁREA G				
Curso A	2 (57.778)	2 (130)	1 (1.168)	1 (22)
Curso B	4 (46.667)	4 (105)	4 (1.088)	4 (21)
Curso C	3 (48.444)	3 (109)	1 (1.168)	1 (22)
Curso D	11 (35.556)	11 (80)	13 (0.370)	13 (8)
Curso E	1 (64)	1 (144)	3 (1.146)	3 (21.75)
Curso F	5 (43.556)	5 (98)	6 (0.951)	6 (18.75)
Curso G	6 (39.556)	6 (89)	10 (0.598)	10 (12)
Curso H	10 (37.335)	10 (84)	8 (0.847)	8 (16)
Curso I	24 (14.222)	24 (32)	18 (0.174)	18 (4.75)
Curso J	8 (39.111)	8 (88)	9 (0.790)	9 (15)
Curso K	19 (20.444)	19 (46)	24 (*)	24 (1.375)
Curso L	6 (39.556)	6 (89)	5 (0.970)	5 (19)
Curso M	9 (37.778)	9 (85)	7 (0.914)	7 (18)
Curso N	12 (29.333)	12 (66)	11 (0.429)	11 (9)
Curso O	20 (19.556)	20 (44)	17 (0.250)	17 (6)
Curso P	24 (14.222)	24 (32)	22 (0.022)	22 (2)
Curso Q	21 (19.111)	21 (43)	21 (0.079)	21 (3)
Curso R	17 (21.333)	17 (48)	24 (*)	24 (1.375)
Curso S	22 (17.333)	22 (39)	22 (0.022)	22 (2)
Curso T	14 (26.222)	14 (59)	20 (0.137)	20 (4)
Curso U	13 (26.667)	13 (60)	11 (0.429)	11 (9)
Curso V	17 (21.333)	17 (48)	14 (0.355)	14 (7.75)
Curso W	15 (24)	15 (54)	15 (0.310)	15 (7)
Curso X	16 (22.222)	16 (50)	15 (0.310)	15 (7)
Curso Y	23 (16)	23 (36)	18 (0.174)	18 (4.75)
ÁREA H				
Curso A	1 (66.67)	1 (12)	1 (0)	1 (2.083)
Curso B	2 (44.44)	2 (8)	2 (*)	2 (1.083)
ÁREA I				
Curso A	3 (33.333)	3 (12)	3 (0.223)	3 (1.875)
Curso B	3 (33.333)	3 (12)	4 (0.149)	4 (1.75)
Curso C	1 (58.333)	1 (21)	1 (0.913)	1 (3.625)
Curso D	2 (52.778)	2 (19)	2 (0.807)	2 (3.25)
ÁREA K				

Anexos

Curso A	8 (28.395)	8 (23)	7 (0.392)	7 (3)
Curso B	6 (38.272)	6 (31)	7 (0.392)	7 (3)
Curso C	4 (44.444)	4 (36)	2 (1.245)	2 (6.25)
Curso D	2 (49.383)	2 (40)	4 (0.95)	4 (5)
Curso E	1 (65.432)	1 (53)	1 (1.548)	1 (8.625)
Curso F	7 (33.355)	7 (27)	6 (0.859)	6 (4.667)
Curso G	3 (45.699)	3 (37)	3 (1.193)	3 (6)
Curso H	5 (39.506)	5 (32)	4 (0.95)	4 (5)
Curso I	9 (20.988)	9 (17)	9 (*)	9 (1.375)

PRIORIZACIÓN EN DIPUTACIÓN DE SEVILLA 2005 (N=21)				
ÁREA A	DUNN-RANKIN	ORDEN. RANGOS	INT. SUCESIVOS	INT. APAR. IGUALES
Curso A	2 (61.905)	2 (26)	2 (*)	2 (2.4)
Curso B	1 (64.286)	1 (27)	1 (0.375)	1 (2.7)
ÁREA B				
Curso A	2 (50.476)	2 (53)	2 (0.791)	2 (3.083)
Curso B	3 (44.762)	3 (47)	3 (0.614)	3 (2.75)
Curso C	1 (62.857)	1 (66)	1 (1.313)	1 (4)
Curso D	5 (39.048)	5 (41)	4 (0.241)	4 (2)
Curso E	4 (40.952)	4 (43)	4 (0.241)	4 (2)
ÁREA C				
Curso A	30 (29.21)	30 (227)	33 (0.27)	33 (7.5)
Curso B	11 (45.817)	11 (356)	9 (1.252)	9 (24)
Curso C	1 (68.726)	1 (534)	1 (2.132)	1 (34.167)
Curso D	15 (44.53)	15 (346)	13 (1.023)	13 (20.5)
Curso E	8 (51.737)	8 (402)	15 (0.956)	15 (19.5)
Curso F	18 (41.828)	18 (325)	17 (0.891)	17 (18.5)
Curso G	24 (36.036)	24 (280)	20 (0.738)	20 (16)
Curso H	36 (18.533)	36 (144)	35 (0.161)	35 (5.5)
Curso I	9 (50.45)	9 (392)	5 (1.468)	5 (27)
Curso J	3 (54.183)	3 (421)	2 (1.647)	2 (29.5)
Curso K	17 (43.243)	17 (336)	15 (0.956)	15 (19.5)
Curso L	34 (25.483)	34 (198)	32 (0.325)	32 (8.5)
Curso M	2 (55.598)	2 (432)	4 (1.585)	4 (28.5)
Curso N	20 (36.68)	20 (285)	27 (0.49)	27 (11.5)
Curso O	12 (45.56)	12 (354)	11 (1.187)	11 (23)
Curso P	14 (44.788)	14 (348)	12 (1.087)	12 (21.5)
Curso Q	5 (53.024)	5 (412)	6 (1.382)	6 (25.833)
Curso R	21 (36.551)	21 (284)	25 (0.589)	25 (13.5)
Curso S	25 (33.848)	25 (263)	27 (0.49)	27 (11.5)
Curso T	26 (33.719)	26 (262)	26 (0.562)	26 (13)
Curso U	10 (45.946)	10 (357)	13 (1.023)	13 (20.5)
Curso V	27 (33.591)	27 (261)	22 (0.707)	22 (15.5)
Curso W	28 (33.462)	28 (260)	27 (0.49)	27 (11.5)
Curso X	7 (52.252)	7 (406)	3 (1.616)	3 (29)
Curso Y	13 (45.431)	13 (353)	10 (1.22)	10 (23.5)
Curso Z	5 (53.024)	5 (412)	7 (1.284)	7 (24.5)
Curso AA	33 (27.671)	33 (215)	33 (0.27)	33 (7.5)
Curso AB	35 (25.354)	35 (197)	36 (0.119)	36 (4.5)
Curso AC	23 (36.422)	23 (283)	20 (0.738)	20 (16)
Curso AD	29 (29.472)	29 (229)	23 (0.677)	23 (15)
Curso AE	4 (53.153)	4 (413)	7 (1.284)	7 (24.5)
Curso AF	21 (36.551)	21 (284)	19 (0.863)	19 (18)
Curso AG	31 (29.086)	31 (226)	30 (0.427)	30 (10.5)
Curso AH	37 (16.731)	37 (130)	37 (0.109)	37 (4.25)
Curso AI	19 (39.382)	19 (306)	24 (0.648)	24 (14.5)
Curso AJ	32 (28.571)	32 (222)	31 (0.38)	31 (9.5)
Curso AK	16 (43.887)	16 (341)	17 (0.891)	17 (18.5)
ÁREA E				
Curso A	1 (85.714)	1 (36)	1 (0)	1 (1)
Curso B	2 (42.857)	2 (18)	2 (*)	2 (2)
ÁREA F				
Curso A	3 (52.381)	3 (55)	3 (1.155)	3 (3.375)
Curso B	1 (59.048)	1 (62)	1 (1.410)	1 (3.875)
Curso C	5 (27.619)	5 (29)	5 (*)	5 (1.444)
Curso D	4 (46.667)	4 (49)	4 (0.866)	4 (2.857)
Curso E	2 (54.286)	2 (57)	2 (1.373)	2 (3.8)
ÁREA G				

Anexos

Curso A	8 (41.905)	8 (132)	9 (1.132)	9 (8)
Curso B	1 (60)	1 (189)	2 (2.089)	2 (12.5)
Curso C	7 (42.222)	7 (133)	7 (1.227)	7 (8.5)
Curso D	2 (54.603)	2 (172)	3 (2.012)	3 (12.167)
Curso E	12 (33.651)	12 (106)	13 (0.604)	13 (4.5)
Curso F	4 (48.571)	4 (153)	4 (1.537)	4 (10.167)
Curso G	8 (41.905)	8 (132)	7 (1.227)	7 (8.5)
Curso H	13 (28.254)	13 (89)	12 (0.699)	12 (5.167)
Curso I	10 (39.683)	10 (125)	9 (1.132)	9 (8)
Curso J	2 (54.603)	2 (172)	1 (2.216)	1 (13)
Curso K	14 (25.079)	14 (79)	14 (0.369)	14 (3.5)
Curso L	15 (23.175)	15 (73)	15 (0.16)	15 (2.5)
Curso M	11 (34.603)	11 (109)	11 (0.992)	11 (7.167)
Curso N	6 (44.444)	6 (140)	5 (1.507)	5 (10)
Curso O	5 (44.762)	5 (141)	6 (1.414)	6 (9.5)
ÁREA I				
Curso A	1 (63.492)	1 (40)	1 (0.747)	1 (2.333)
Curso B	3 (49.206)	3 (31)	3 (*)	3 (1.5)
Curso C	2 (58.73)	2 (37)	2 (0.512)	2 (2.071)
ÁREA J				
Curso A	1 (69.048)	1 (29)	1 (0)	1 (1.55)
Curso B	2 (66.667)	2 (28)	2 (*)	2 (1.45)
ÁREA K				
Curso A	4 (39.286)	4 (33)	4 (*)	4 (1.364)
Curso B	1 (73.81)	1 (62)	1 (*)	1 (3.636)
Curso C	2 (66.667)	2 (56)	2 (1.232)	2 (3.05)
Curso D	3 (44.048)	3 (37)	3 (0.389)	3 (2)

PRIORIZACIÓN EN DIPUTACIÓN DE SEVILLA 2006 (N=374)				
ÁREA A	DUNN-RANKIN	ORDEN. RANGOS	INT. SUCESIVOS	INT. APAR. IGUALES
Curso B	2 (14.225)	2 (532)	3 (0.687)	3 (4.55)
Curso C	1 (15.321)	1 (573)	3 (0.687)	3 (4.55)
Curso D	3 (13.583)	3 (508)	6 (0.621)	6 (4.263)
Curso E	5 (12.54)	5 (469)	1 (0.703)	1 (4.625)
Curso F	7 (11.497)	7 (430)	8 (0.487)	8 (3.667)
Curso G	9 (10.936)	9 (409)	10 (0.384)	10 (3.182)
Curso H	4 (13.235)	4 (495)	5 (0.668)	5 (4.471)
Curso I	6 (12.059)	6 (451)	2 (0.691)	2 (4.571)
Curso J	8 (11.337)	8 (424)	7 (0.497)	7 (3.714)
Curso K	10 (10.829)	10 (405)	11 (0.329)	11 (2.917)
Curso L	12 (9.412)	12 (352)	12 (*)	12 (1.473)
Curso M	11 (10.775)	11 (403)	9 (0.398)	9 (3.25)
ÁREA B				
Curso B	5 (6.417)	5 (240)	6 (0.549)	6 (4.333)
Curso C	3 (6.818)	3 (255)	3 (0.637)	3 (4.8)
Curso D	8 (5.294)	8 (198)	8 (0.404)	8 (3.571)
Curso E	6 (6.123)	6 (229)	4 (0.621)	4 (4.714)
Curso F	11 (4.465)	11 (167)	11 (0.219)	11 (2.625)
Curso G	1 (7.62)	1 (285)	1 (0.746)	1 (5.375)
Curso H	7 (5.374)	7 (201)	7 (0.462)	7 (3.875)
Curso I	9 (4.733)	9 (177)	9 (0.243)	9 (2.75)
Curso J	4 (6.791)	4 (254)	4 (0.565)	4 (4.417)
Curso K	12 (4.091)	12 (153)	14 (*)	14 (1.444)
Curso L	14 (3.77)	14 (141)	14 (*)	14 (1.444)
Curso M	19 (2.54)	19 (95)	19 (*)	19 (1.229)
Curso N	16 (3.556)	16 (133)	14 (*)	14 (1.444)
Curso O	17 (3.503)	17 (131)	18 (*)	18 (1.297)
Curso P	15 (3.743)	15 (140)	12 (0.097)	12 (2)
Curso Q	13 (4.037)	13 (151)	13 (0.049)	13 (1.75)
Curso R	18 (3.075)	18 (115)	17 (*)	17 (1.323)
Curso S	2 (7.139)	2 (267)	2 (0.729)	2 (5.286)
Curso T	10 (4.626)	10 (173)	10 (0.227)	10 (2.667)
ÁREA C				
Curso B	16 (2.674)	16 (100)	16 (*)	16 (1.37)
Curso C	16 (2.674)	16 (100)	15 (*)	15 (1.409)
Curso D	23 (2.353)	23 (88)	18 (*)	18 (1.3)
Curso E	7 (4.171)	7 (156)	8 (0.382)	8 (3.5)
Curso F	11 (3.877)	11 (145)	6 (0.41)	6 (3.833)
Curso G	25 (2.112)	25 (79)	18 (*)	18 (1.3)
Curso H	28 (1.952)	28 (73)	25 (*)	25 (1.214)
Curso I	71 (0.829)	71 (31)	44 (*)	44 (1.125)
Curso J	42 (1.524)	42 (57)	41 (*)	41 (1.145)
Curso K	9 (3.957)	9 (148)	11 (0.287)	11 (2.5)
Curso L	1 (5.348)	1 (200)	3 (0.498)	3 (4.833)
Curso M	6 (4.305)	6 (161)	7 (0.403)	7 (3.75)
Curso N	16 (2.674)	16 (100)	22 (*)	22 (1.241)
Curso O	8 (4.011)	8 (150)	9 (0.35)	9 (3.167)
Curso P	2 (5.214)	2 (195)	4 (0.49)	4 (4.75)
Curso Q	12 (3.128)	12 (117)	12 (0.287)	12 (1.833)
Curso R	74 (0.802)	74 (30)	65 (*)	65 (1.071)
Curso S	61 (0.963)	61 (36)	65 (*)	65 (1.071)
Curso T	46 (1.283)	46 (48)	44 (*)	44 (1.125)
Curso U	99 (0.428)	99 (16)	88 (*)	88 (1.041)
Curso V	65 (0.909)	65 (34)	58 (*)	58 (1.088)
Curso W	113 (0.294)	113 (11)	104 (*)	104 (1.026)
Curso X	36 (1.658)	36 (62)	35 (*)	35 (1.167)

Anexos

Curso Y	61 (0.963)	61 (36)	35 (*)	35 (1.167)
Curso Z	5 (4.813)	5 (180)	1 (0.529)	1 (5.167)
Curso AA	4 (4.973)	4 (186)	5 (0.452)	5 (4.333)
Curso AB	3 (5.187)	3 (194)	2 (0.514)	2 (5)
Curso AC	10 (3.93)	10 (147)	10 (0.335)	10 (3)
Curso AD	25 (2.112)	25 (79)	22 (*)	22 (1.241)
Curso AE	34 (1.765)	34 (66)	35 (*)	35 (1.167)
Curso AF	71 (0.829)	71 (31)	58 (*)	58 (1.088)
Curso AG	59 (0.989)	59 (37)	50 (*)	50 (1.106)
Curso AH	56 (1.043)	56 (39)	41 (*)	41 (1.145)
Curso AI	39 (1.578)	39 (59)	30 (*)	30 (1.19)
Curso AJ	71 (0.829)	71 (31)	65 (*)	65 (1.071)
Curso AK	31 (1.872)	31 (70)	30 (*)	30 (1.19)
Curso AL	56 (1.043)	56 (39)	50 (*)	50 (1.106)
Curso AM	65 (0.909)	65 (34)	58 (*)	58 (1.088)
Curso AN	87 (0.535)	87 (20)	65 (*)	65 (1.071)
Curso AO	53 (1.07)	53 (40)	50 (*)	50 (1.106)
Curso AP	97 (0.455)	97 (17)	104 (*)	104 (1.026)
Curso AQ	113 (0.294)	113 (11)	119 (*)	119 (1.013)
Curso AR	89 (0.508)	89 (19)	88 (*)	88 (1.041)
Curso AS	108 (0.348)	108 (13)	88 (*)	88 (1.041)
Curso AT	94 (0.481)	94 (18)	104 (*)	104 (1.026)
Curso AU	59 (0.989)	59 (37)	78 (*)	78 (1.056)
Curso AV	112 (0.321)	112 (12)	104 (*)	104 (1.026)
Curso AW	24 (2.139)	24 (80)	25 (*)	25 (1.214)
Curso AX	51 (1.096)	51 (41)	50 (*)	50 (1.106)
Curso AY	84 (0.588)	84 (22)	65 (*)	65 (1.071)
Curso AZ	105 (0.374)	105 (14)	78 (*)	78 (1.056)
Curso BA	56 (1.043)	56 (39)	50 (*)	50 (1.106)
Curso BB	19 (2.567)	19 (96)	20 (*)	20 (1.269)
Curso BC	21 (2.46)	21 (92)	22 (*)	22 (1.241)
Curso BD	22 (2.406)	22 (90)	25 (*)	25 (1.214)
Curso BE	40 (1.551)	40 (58)	44 (*)	44 (1.125)
Curso BF	43 (1.337)	43 (50)	44 (*)	44 (1.125)
Curso BG	14 (3.021)	14 (113)	16 (*)	16 (1.37)
Curso BH	27 (1.979)	27 (74)	25 (*)	25 (1.214)
Curso BI	13 (3.102)	13 (116)	14 (*)	14 (1.452)
Curso BJ	15 (2.968)	15 (111)	13 (*)	13 (1.5)
Curso BK	47 (1.23)	47 (46)	41 (*)	41 (1.145)
Curso BL	53 (1.07)	53 (40)	50 (*)	50 (1.106)
Curso BM	44 (1.31)	44 (49)	35 (*)	35 (1.167)
Curso BN	40 (1.551)	40 (58)	30 (*)	30 (1.19)
Curso BO	23 (1.952)	23 (73)	30 (*)	30 (1.19)
Curso BP	33 (1.818)	33 (68)	20 (*)	20 (1.269)
Curso BQ	97 (0.455)	97 (17)	104 (*)	104 (1.026)
Curso BR	38 (1.604)	38 (60)	35 (*)	35 (1.167)
Curso BS	31 (1.872)	31 (70)	35 (*)	35 (1.167)
Curso BT	36 (1.658)	36 (62)	44 (*)	44 (1.125)
Curso BU	77 (0.722)	77 (27)	78 (*)	78 (1.056)
Curso BV	28 (1.952)	28 (73)	30 (*)	30 (1.19)
Curso BW	35 (1.684)	35 (63)	44 (*)	44 (1.125)
Curso BX	44 (1.31)	44 (49)	50 (*)	50 (1.106)
Curso BY	76 (0.775)	76 (29)	65 (*)	65 (1.071)
Curso BZ	63 (0.936)	63 (35)	78 (*)	78 (1.056)
Curso CA	50 (1.123)	50 (42)	65 (*)	65 (1.071)
Curso CB	68 (0.856)	68 (32)	58 (*)	58 (1.088)
Curso CC	74 (0.802)	74 (30)	78 (*)	78 (1.056)
Curso CD	117 (0.187)	117 (7)	104 (*)	104 (1.026)
Curso CE	89 (0.508)	89 (19)	104 (*)	104 (1.026)
Curso CF	78 (0.695)	78 (26)	78 (*)	78 (1.056)

Innovaciones metodológicas en la evaluación de la formación continua

Curso CG	94 (0.481)	94 (18)	104 (*)	104 (1.026)
Curso CH	83 (0.615)	83 (23)	88 (*)	88 (1.041)
Curso CI	108 (0.348)	108 (13)	88 (*)	88 (1.041)
Curso CJ	84 (0.588)	84 (22)	88 (*)	88 (1.041)
Curso CK	48 (1.203)	48 (45)	58 (*)	58 (1.088)
Curso CL	121 (0.134)	121 (5)	104 (*)	104 (1.026)
Curso CM	78 (0.695)	78 (26)	65 (*)	65 (1.071)
Curso CN	108 (0.348)	108 (13)	78 (*)	78 (1.056)
Curso CO	89 (0.508)	89 (19)	78 (*)	78 (1.056)
Curso CP	53 (1.07)	53 (40)	50 (*)	50 (1.106)
Curso CQ	84 (0.588)	84 (22)	88 (*)	88 (1.041)
Curso CR	101 (0.401)	101 (15)	78 (*)	78 (1.056)
Curso CS	67 (0.882)	67 (33)	65 (*)	65 (1.071)
Curso CT	117 (0.187)	117 (7)	104 (*)	104 (1.026)
Curso CU	130 (0)	130 (0)	129 (*)	129 (1)
Curso CV	51 (1.096)	51 (41)	65 (*)	65 (1.071)
Curso CW	48 (1.203)	48 (45)	58 (*)	58 (1.088)
Curso CX	101 (0.401)	101 (15)	104 (*)	104 (1.026)
Curso CY	105 (0.374)	105 (14)	104 (*)	104 (1.026)
Curso CZ	124 (0.08)	124 (3)	119 (*)	119 (1.013)
Curso DA	115 (0.267)	115 (10)	104 (*)	104 (1.026)
Curso DB	81 (0.642)	81 (24)	88 (*)	88 (1.041)
Curso DC	99 (0.428)	99 (16)	88 (*)	88 (1.041)
Curso DD	89 (0.508)	89 (19)	88 (*)	88 (1.041)
Curso DE	117 (0.187)	117 (7)	104 (*)	104 (1.026)
Curso DF	129 (0.027)	129 (1)	129 (*)	129 (1)
Curso DG	121 (0.134)	121 (5)	119 (*)	119 (1.013)
Curso DH	127 (0.053)	127 (2)	119 (*)	119 (1.013)
Curso DI	120 (0.16)	120 (6)	119 (*)	119 (1.013)
Curso DJ	68 (0.856)	68 (32)	65 (*)	65 (1.071)
Curso DK	80 (0.668)	80 (25)	88 (*)	88 (1.041)
Curso DL	130 (0)	130 (0)	129 (*)	129 (1)
Curso DM	89 (0.508)	89 (19)	88 (*)	88 (1.041)
Curso DN	68 (0.856)	68 (32)	65 (*)	65 (1.071)
Curso DO	81 (0.642)	81 (24)	65 (*)	65 (1.071)
Curso DP	63 (0.936)	63 (35)	58 (*)	58 (1.088)
Curso DQ	101 (0.401)	101 (15)	88 (*)	88 (1.041)
Curso DR	130 (0)	130 (0)	129 (*)	129 (1)
Curso DS	108 (0.348)	108 (13)	88 (*)	88 (1.041)
Curso DT	124 (0.08)	124 (3)	119 (*)	119 (1.013)
Curso DU	124 (0.08)	124 (3)	119 (*)	119 (1.013)
Curso DV	130 (0)	130 (0)	129 (*)	129 (1)
Curso DW	105 (0.374)	105 (14)	88 (*)	88 (1.041)
Curso DX	130 (0)	130 (0)	129 (*)	129 (1)
Curso DY	116 (0.241)	116 (9)	119 (*)	119 (1.013)
Curso DZ	130 (0)	130 (0)	129 (*)	129 (1)
Curso EA	130 (0)	130 (0)	129 (*)	129 (1)
Curso EB	123 (0.107)	123 (4)	119 (*)	119 (1.013)
Curso EC	87 (0.535)	87 (20)	88 (*)	88 (1.041)
Curso ED	94 (0.481)	94 (18)	78 (*)	78 (1.056)
Curso EE	101 (0.401)	101 (15)	104 (*)	104 (1.026)
Curso EF	20 (2.513)	20 (94)	25 (*)	25 (1.214)
Curso EG	127 (0.053)	127 (2)	119 (*)	119 (1.013)
ÁREA D				
Curso B	1 (15.561)	1 (582)	1 (0.528)	1 (4)
Curso C	2 (11.845)	2 (443)	2 (0.400)	2 (3.179)
Curso D	3 (7.754)	3 (290)	3 (0.089)	3 (1.875)
Curso E	4 (6.658)	4 (249)	4 (*)	4 (1.311)
ÁREA E				
Curso B	1 (9.358)	1 (350)	1 (1)	1 (4.214)

Anexos

Curso C	4 (8.048)	4 (301)	3 (0.913)	3 (3.955)
Curso D	3 (8.075)	3 (302)	4 (0.751)	4 (3.5)
Curso E	2 (9.064)	2 (339)	2 (0.947)	2 (4.056)
ÁREA F				
Curso B	1 (8.556)	1 (320)	1 (0.756)	1 (6.625)
Curso C	5 (5.642)	5 (211)	5 (0.191)	5 (3)
Curso D	3 (6.845)	3 (256)	4 (0.464)	4 (5)
Curso E	8 (3.369)	8 (126)	9 (*)	9 (1.279)
Curso F	2 (8.128)	2 (304)	2 (0.75)	2 (6.6)
Curso G	6 (4.492)	6 (168)	6 (0.016)	6 (1.625)
Curso H	4 (6.791)	4 (254)	3 (0.634)	3 (6)
Curso I	26 (1.331)	26 (50)	19 (*)	19 (1.146)
Curso J	13 (2.166)	13 (81)	12 (*)	12 (1.216)
Curso K	7 (4.385)	7 (164)	7 (*)	7 (1.481)
Curso L	9 (3.102)	9 (116)	10 (*)	10 (1.257)
Curso M	26 (1.337)	26 (50)	29 (*)	29 (1.089)
Curso N	15 (2.032)	15 (76)	27 (*)	27 (1.102)
Curso O	24 (1.551)	24 (58)	27 (*)	27 (1.102)
Curso P	24 (1.551)	24 (58)	19 (*)	19 (1.146)
Curso Q	30 (1.123)	30 (42)	29 (*)	29 (1.089)
Curso R	29 (1.203)	29 (45)	24 (*)	24 (1.116)
Curso S	17 (1.898)	17 (71)	15 (*)	15 (1.163)
Curso T	32 (0.508)	32 (19)	32 (*)	32 (1.041)
Curso U	33 (0)	33 (0)	33 (*)	33 (1)
Curso V	10 (2.995)	10 (112)	8 (*)	8 (1.355)
Curso W	12 (2.299)	12 (86)	15 (*)	15 (1.163)
Curso X	10 (2.995)	10 (112)	10 (*)	10 (1.257)
Curso Y	20 (1.738)	20 (65)	19 (*)	19 (1.146)
Curso Z	16 (1.925)	16 (72)	13 (*)	13 (1.197)
Curso AA	28 (1.31)	28 (49)	24 (*)	24 (1.116)
Curso AB	23 (1.658)	23 (62)	22 (*)	22 (1.131)
Curso AC	31 (0.615)	31 (23)	31 (*)	31 (1.052)
Curso AD	21 (1.711)	21 (64)	22 (*)	22 (1.131)
Curso AE	19 (1.765)	19 (66)	24 (*)	24 (1.116)
Curso AF	18 (1.872)	18 (70)	14 (*)	14 (1.179)
Curso AG	22 (1.684)	22 (63)	15 (*)	15 (1.163)
Curso AH	14 (2.059)	14 (77)	15 (*)	15 (1.163)
ÁREA G				
Curso B	18 (3.235)	18 (121)	14 (0.245)	14 (2.25)
Curso C	29 (1.952)	29 (73)	27 (*)	27 (1.241)
Curso D	41 (0.561)	41 (21)	41 (*)	41 (1.051)
Curso E	40 (0.722)	40 (27)	38 (*)	38 (1.081)
Curso F	27 (2.406)	27 (90)	29 (*)	29 (1.194)
Curso G	4 (6.096)	4 (228)	8 (0.419)	8 (4.417)
Curso H	34 (1.364)	34 (51)	38 (*)	38 (1.081)
Curso I	18 (3.235)	18 (121)	22 (*)	22 (1.36)
Curso J	3 (6.711)	3 (251)	3 (0.672)	3 (6)
Curso K	21 (3.155)	21 (118)	18 (0.162)	18 (1.583)
Curso L	15 (3.717)	15 (139)	22 (*)	22 (1.36)
Curso M	35 (1.257)	35 (47)	36 (*)	36 (1.097)
Curso N	5 (5.668)	5 (212)	4 (0.546)	4 (5.25)
Curso O	2 (6.738)	2 (252)	2 (0.707)	2 (6.2)
Curso P	38 (1.203)	38 (45)	33 (*)	33 (1.132)
Curso Q	33 (1.471)	33 (55)	33 (*)	33 (1.132)
Curso R	8 (5.321)	8 (199)	5 (0.508)	5 (5)
Curso S	6 (5.535)	6 (207)	6 (0.488)	6 (4.875)
Curso T	10 (4.626)	10 (173)	11 (0.262)	11 (3.333)
Curso U	7 (5.455)	7 (204)	7 (0.457)	7 (4.667)
Curso V	31 (1.818)	31 (68)	32 (*)	32 (1.172)
Curso W	9 (4.652)	9 (174)	9 (0.322)	9 (3.75)

Innovaciones metodológicas en la evaluación de la formación continua

Curso X	1 (7.059)	1 (264)	1 (0.861)	1 (7)
Curso Y	21 (3.155)	21 (118)	20 (*)	20 (1.396)
Curso Z	24 (2.914)	24 (109)	26 (*)	26 (1.327)
Curso AA	28 (2.326)	28 (87)	27 (*)	27 (1.241)
Curso AB	20 (3.182)	20 (119)	20 (*)	20 (1.396)
Curso AC	17 (3.262)	17 (122)	17 (0.153)	17 (1.625)
Curso AD	16 (3.369)	16 (126)	13 (0.257)	13 (2.333)
Curso AE	39 (1.016)	39 (38)	36 (*)	36 (1.097)
Curso AF	14 (3.85)	14 (144)	12 (0.196)	12 (2.857)
Curso AG	11 (4.278)	11 (160)	10 (0.3)	10 (3.6)
Curso AH	23 (3.075)	23 (115)	19 (*)	19 (1.477)
Curso AI	30 (1.898)	30 (71)	29 (*)	29 (1.194)
Curso AJ	26 (2.647)	26 (99)	22 (*)	22 (1.36)
Curso AK	32 (1.765)	32 (66)	29 (*)	29 (1.194)
Curso AL	36 (1.23)	36 (46)	35 (*)	35 (1.114)
Curso AM	25 (2.807)	25 (105)	22 (*)	22 (1.36)
Curso AN	13 (3.957)	13 (148)	16 (0.172)	16 (1.75)
Curso AO	36 (1.23)	36 (46)	38 (*)	38 (1.081)
Curso AP	12 (4.198)	12 (157)	15 (0.208)	15 (2)
ÁREA H				
Curso B	2 (8.93)	2 (334)	4 (0.576)	4 (4.4)
Curso C	1 (10)	1 (375)	1 (0.909)	1 (5.875)
Curso D	3 (8.422)	3 (315)	2 (0.666)	2 (4.8)
Curso E	5 (7.353)	5 (275)	6 (0.443)	6 (3.813)
Curso F	7 (6.096)	7 (228)	7 (0.351)	7 (3.375)
Curso G	6 (6.711)	6 (251)	5 (0.542)	5 (4.25)
Curso H	4 (7.701)	4 (288)	3 (0.612)	3 (4.556)
ÁREA I				
Curso B	8 (10)	8 (374)	7 (0.926)	7 (4.5)
Curso C	10 (8.69)	10 (325)	10 (0.72)	10 (3.667)
Curso D	1 (11.684)	1 (437)	1 (1.178)	1 (5.643)
Curso E	3 (11.203)	3 (419)	4 (1.109)	4 (5.333)
Curso F	2 (11.417)	2 (11.417)	2 (1.146)	2 (5.5)
Curso G	9 (9.893)	9 (9.893)	5 (1.013)	5 (4.9)
Curso H	6 (10.481)	6 (10.481)	7 (0.926)	7 (4.5)
Curso I	5 (10.481)	5 (10.481)	6 (1.008)	6 (4.875)
Curso J	4 (11.016)	4 (11.016)	3 (1.124)	3 (5.4)
Curso K	7 (10.107)	7 (10.107)	9 (0.891)	9 (4.357)
ÁREA J				
Curso B	1 (8.4)	1 (314)	1 (1.173)	1 (6.917)
Curso C	2 (7.09)	2 (265)	2 (1.353)	2 (5)
Curso D	3 (6.79)	3 (254)	4 (1.384)	4 (4)
Curso E	4 (6.68)	4 (250)	3 (1.135)	3 (4.083)
AREA K				
Curso B	9 (5.588)	9 (209)	12 (0.258)	12 (2.6)
Curso C	6 (6.364)	6 (238)	6 (0.61)	6 (4.4)
Curso D	5 (6.444)	5 (241)	7 (0.583)	7 (4.25)
Curso E	7 (5.802)	7 (217)	9 (0.522)	9 (3.917)
Curso F	3 (6.952)	3 (260)	3 (0.722)	3 (5)
Curso G	12 (4.973)	12 (186)	11 (0.36)	11 (3.083)
Curso H	14 (4.091)	14 (153)	13 (0.04)	13 (1.667)
Curso I	13 (4.171)	13 (156)	14 (0.02)	14 (1.583)
Curso J	2 (7.086)	2 (265)	2 (0.749)	2 (5.143)
Curso K	11 (5.348)	11 (200)	8 (0.538)	8 (4)
Curso L	10 (5.428)	10 (203)	10 (0.59)	10 (3.8)
Curso M	8 (5.615)	8 (210)	5 (0.641)	5 (4.571)
Curso N	16 (2.941)	16 (110)	16 (*)	16 (0.229)
Curso O	15 (3.449)	15 (129)	15 (*)	15 (0.297)
Curso P	4 (6.631)	4 (248)	3 (0.722)	3 (5)
Curso Q	1 (7.353)	1 (275)	1 (0.887)	1 (5.857)

ANEXO IX

Número de empates obtenidos por los distintos métodos de escalamiento

AÑO	ÁREA DE CONOCIMIENTO	DR/OR	INTERV.
		Nº CURSOS EMPATADOS	
2004	A	1	1
	B	1	1
	C	8	15
	E	0	0
	F	0	0
	G	3	6
	H	0	0
	I	1	0
	K	0	2
2005	A	0	0
	B	0	1
	C	2	8
	E	0	0
	F	0	0
	G	2	2
	I	0	0
	J	0	0
	K	0	0
2006	A	0	1
	B	0	3
	C	57	109
	D	0	0
	E	0	0
	F	3	10
	G	3	11
	H	0	0
	I	0	1
	J	0	0
	K	0	1
TOTAL		81	172

ANEXO X

**Cuestionario disponible durante la realización de cada acción
formativa**

CUESTIONARIO DURANTE

NO ESPERES A QUE ACABE LA ACCIÓN FORMATIVA. Con la intención de mejorar, haznos llegar cualquier opinión, por insignificante que pueda parecer, sobre los apartados que aparecen a continuación. Muchas gracias por tu colaboración.

NOMBRE DEL CURSO:

CÓDIGO: _____

Fecha en que se cumplimenta este impreso: _____

1. *Objetivos del curso:*

2. *Contenidos del curso:*

3. *El ambiente* (compañerismo, participación, etc.):

4. *Los formadores* (interés, motivación, etc.):

5. *Los materiales y/o recursos* (mobiliario, condiciones ambientales, recursos didácticos, apuntes escritos, etc.):

6. *Otras sugerencias:*

ANEXO XI

**Listado de verificación (check list)
para formadores y coordinadores en los tres momentos:
antes, durante y después de la acción formativa**

LISTA DE VERIFICACIÓN PARA FORMADORES**ANTES DE LA ACCIÓN FORMATIVA**

Estar dado de alta en el registro de formadores	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Presentar el currículo en la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Tramitar la documentación administrativa de la acción formativa	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Adaptar el programa del curso al nivel del grupo	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Elaborar indicadores de transferencia de la formación impartida al puesto de trabajo	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Elaborar la propuesta del programa de la acción formativa a desarrollar	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Entregar la propuesta de programa al coordinador con una antelación no inferior a 15 días	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Comunicar a los coordinadores los recursos didácticos necesarios (aula, materiales, recursos audiovisuales, etc.)	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Preparar el material de apoyo a entregar al alumnado	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Estar disponible para proporcionar los medios necesarios para el buen desarrollo de la acción formativa	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO

DURANTE LA ACCIÓN FORMATIVA

Transmitir cualquier cambio en la programación al coordinador de la acción formativa	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Localizar y solucionar imprevistos en la medida de lo posible	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Promover las prácticas individuales y en grupo	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Favorecer una metodología orientada hacia el autoaprendizaje	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Tener en cuenta las características de los puestos de trabajo a la hora de programar los objetivos y contenidos	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO

DESPUÉS DE LA ACCIÓN FORMATIVA

Participar si fuera pertinente en la reunión final en la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
--	-----------------------------	-----------------------------

LISTA DE VERIFICACIÓN PARA COORDINADORES

ANTES DE LA ACCIÓN FORMATIVA

Presentar el currículo en la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Participar en la programación y evaluación del Plan de Formación	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Conocer los objetivos y contenidos de la acción formativa	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Gestionar los recursos didácticos (aula, materiales, recursos audiovisuales, etc.)	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Recopilar el material de apoyo y repartirlo entre el alumnado	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Recopilar los instrumentos de evaluación (pruebas de conocimiento, cuestionario de satisfacción, cuestionario durante)	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Entregar la documentación y materiales a la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO

DURANTE LA ACCIÓN FORMATIVA

Presentar la acción formativa y a los formadores	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Estar disponible durante todas las sesiones	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Poner a disposición del alumnado los cuestionarios durante	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Entregar al alumnado y recoger las pruebas de conocimiento	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Llevar el control de asistencia de alumnado y profesorado	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Hacer seguimiento del buen estado de los medios didácticos	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Realizar una gestión adecuada de incidencias	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Transmitir cualquier cambio en la programación a la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Localizar y solucionar imprevistos en la medida de lo posible	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO

DESPUÉS DE LA ACCIÓN FORMATIVA

Entregar y recoger el cuestionario de satisfacción de la acción formativa	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO
Entregar las hojas de firma del profesorado y alumnado a la Sección de Formación Continua	<input type="checkbox"/> SÍ	<input type="checkbox"/> NO

ANEXO XII

Cuestionario de satisfacción

CUESTIONARIO DE EVALUACIÓN DE ACCIONES FORMATIVAS

CÓDIGO DE LA ACCIÓN FORMATIVA:

ACCIÓN FORMATIVA:

COORDINADOR/A:

**Este cuestionario es anónimo; rogamos que responda con sinceridad y objetividad.
Su opinión será de gran utilidad para mejorar la calidad de la formación futura.**

VALORE DE 1 A 5 (1: muy en desacuerdo; 2: en desacuerdo; 3: ni de acuerdo ni en desacuerdo; 4: de acuerdo; 5: muy de acuerdo)

OBJETIVOS CONTENIDO	1	A su juicio se han alcanzado los objetivos previstos	1	2	3	4	5
	2	Los temas se han tratado con la profundidad que permitía la duración del curso	1	2	3	4	5
	3	La duración del curso se ha adecuado a los objetivos y contenidos	1	2	3	4	5
METODOLOGÍA AMBIENTE	4	La metodología se ha adecuado a los objetivos y contenidos	1	2	3	4	5
	5	La metodología ha permitido la participación activa de los asistentes	1	2	3	4	5
	6	Esta acción formativa me ha facilitado compartir experiencias profesionales con otros compañeros	1	2	3	4	5
	7	La acción formativa ha sido realista y práctica	1	2	3	4	5
	8	La materiales didácticos entregados han sido de calidad	1	2	3	4	5
	9	Los recursos utilizados han sido adecuados para facilitar el proceso formativo	1	2	3	4	5
UTILIDAD. VALOR GLOBAL	10	La acción formativa recibida es útil para mi puesto de trabajo	1	2	3	4	5
	11	La acción formativa recibida es útil para mi formación personal	1	2	3	4	5
	12	La acción formativa recibida merece una valoración óptima	1	2	3	4	5
COMENTARIOS SUGERENCIAS	13	Señale al menos un aspecto para mejorar la calidad de esta acción formativa: 1.- 2.-					
	14	Plantee otras acciones formativas en las que le gustaría participar para mejorar el desarrollo de su trabajo:					
Propuesta de acción formativa			Nivel Bajo/Medio/Alto		Modalidad Presencial/Distancia		
Destinatarios							
Utilidad para su puesto de trabajo							
Objetivos							
Contenidos							

NOMBRE DEL COORDINADOR:					
1	El coordinador ha estado disponible durante la realización del curso	1	2	3	4 5
2	Ha llevado correctamente el control de asistencia	1	2	3	4 5
3	Ha solucionado debidamente los imprevistos del curso	1	2	3	4 5
4	El coordinador merece una valoración óptima	1	2	3	4 5
5	Otros comentarios respecto al coordinador				

NOMBRE DEL PROFESOR:					
1	El profesor domina la materia	1	2	3	4 5
2	Tiene capacidad para transmitir con claridad sus ideas	1	2	3	4 5
3	Las enseñanzas transmitidas se adecuan a los objetivos y contenidos del curso	1	2	3	4 5
4	Es ameno en sus intervenciones	1	2	3	4 5
5	Motiva y despierta interés en los asistentes	1	2	3	4 5
6	El profesor merece una valoración óptima	1	2	3	4 5
7	Otros comentarios respecto al profesor				

ANEXO XIII

**Modelo de informe referente a la calidad métrica de una
prueba de evaluación de conocimientos dirigida al
profesorado que la elaboró**

CALIDAD MÉTRICA DE LA PRUEBA PARA EVALUAR CONOCIMIENTOS ACCIÓN FORMATIVA: MICROSOFT ACCESS 2000 BÁSICO (200601)

En la siguiente tabla, se presentan los índices métricos obtenidos en la prueba posttest así como los valores que se consideran óptimos.

	Posttest	Valores óptimos
Fiabilidad	0,594	Mayor de 0,6
Discriminación	0,318	Mayor de 0,3
Dificultad	0,875	Óptimo: 0,625

1. Fiabilidad.

La fiabilidad es la consistencia de la medida. La prueba ha obtenido una puntuación algo baja (0,594), aunque muy cercana a su valor óptimo (0,6).

2. Discriminación.

La discriminación es la capacidad de la prueba para diferenciar a las personas con niveles más altos de conocimiento de aquellas con niveles más bajos. La puntuación obtenida es buena (0,318), considerándose a partir de 0,3 un valor aceptable.

3. Dificultad.

El nivel de dificultad de la prueba puede oscilar entre 0 (máxima dificultad) y 1 (máxima facilidad). El test obtiene una dificultad promedio en torno al valor considerado como óptimo (0,875, siendo el óptimo 0,625).

4. Adecuación de los ítems.

En la tabla que se presenta a continuación, se muestran en negrita los ítems que han obtenido unos buenos resultados globales y que se podrían mantener tal como han sido formulados porque han sido acertados por aquellos que más nivel de conocimiento han mostrado, mientras que no lo han acertado quienes obtuvieron puntuaciones más bajas en la prueba; para el resto, se proponen mejoras.

Ítem	Valoración
1	No cambiar nada
2	Cambiar la opción "d" (no fue elegida por nadie; demasiado fácil de discriminar)
3	Cambiar la opción "b" (sin ser la correcta, es elegida por quienes mayores

	puntuaciones obtuvieron en la prueba; confunde)
4	Eliminar (todos lo aciertan; demasiado fácil)
5	Cambiar la opción "d" (sólo fue elegida por una persona; demasiado fácil de discriminar)
6	Eliminar (de 10 personas, lo acertaron 9; demasiado fácil)
7	Eliminar (todos fallaron; demasiado difícil)
8	Eliminar (de 10 personas, fallaron 9; demasiado difícil)
9	No cambiar nada
10	No cambiar nada

En conclusión, creemos que mejorando los aspectos señalados en este punto y considerando sus aportaciones y la experiencia que Vd. como formador tiene en la materia que imparte, se podría mejorar la calidad de la prueba aplicable en próximos cursos o ediciones a impartir.

En **<http://www.dipusevilla.es/formacioncontinua>** puede ver los resultados tanto de la prueba de conocimientos como de los cuestionarios de satisfacción de su curso.

Si tiene alguna duda, puede contactar con nosotros llamando al 954550862 o escribiendo a la dirección de correo electrónico formacioncontinua@dipusevilla.es.

ANEXO XIV

Posibles instrumentos para evaluar transferencia (Chacón, Holgado, López y Sanduvete, 2006)

INSTRUMENTO 1

CUESTIONARIO DE EVALUACIÓN DE LA TRANSFERENCIA EN EL CENTRO DE FORMACIÓN DEL PAS

El objetivo de este cuestionario es conocer si pudo usted aplicar en su trabajo lo aprendido en el curso. Estos datos servirán para intentar mejorar las acciones formativas que se programan y para detectar otras posibles necesidades formativas existentes.

Valore de 1 a 5 (1: nada de acuerdo; 5: totalmente de acuerdo)

1. ¿Aplica en su puesto de trabajo lo que aprendió en el curso?	1	2	3	4	5
2. ¿Cree que los resultados de su trabajo han mejorado por haber participado en el curso?	1	2	3	4	5
3. ¿Dispone en su entorno de trabajo de los recursos y condiciones necesarios para poder aplicar la formación?	1	2	3	4	5
4. Añada cualquier comentario que estime de interés.					

INSTRUMENTO 2

CUESTIONARIO DE EVALUACIÓN DE LA TRANSFERENCIA EN DIPUTACIÓN DE SEVILLA

Correo electrónico: Para evaluar la utilidad de lo que aprendió en la acción formativa denominada _____ en la que Vd. participó, le pedimos que cumplimente con sinceridad un cuestionario. Le agradecemos de antemano su participación; sólo le supondrá varios minutos de su tiempo. La información que recibamos será tratada de forma confidencial. Su opinión es fundamental para mejorar la calidad de futuras acciones formativas.

Puede acceder al cuestionario en el siguiente enlace: PONER VÍNCULO QUE LE LLEVE AL CUESTIONARIO.

CUESTIONARIO ANEXO

ACCIÓN FORMATIVA
FECHA DE REALIZACIÓN de la acción formativa
Fecha de cumplimentación de este cuestionario

VALORE DE 1 A 5 (1:nada de acuerdo, 5: totalmente de acuerdo)

CUESTIONES	1	2	3	4	5
1. He aplicado la formación recibida					
2. He mejorado el rendimiento en mi trabajo					
3. He puesto en práctica alguna innovación (ha aumentado mi iniciativa)					
4. Ha mejorado mi grado de motivación en el trabajo					
5. El entorno de trabajo me facilita la aplicación de la formación recibida					
6. Dispongo de los recursos necesarios para aplicar los conocimientos adquiridos en el puesto de trabajo					

7. Razones principales por las que realizó el curso:

- Mejora del desempeño en el puesto de trabajo.
- Promoción profesional.
- Desarrollo e interés personal.
- Otros (especificar).....

8. OBSERVACIONES (Añada cualquier comentario que considere oportuno respecto a la utilidad de la acción formativa).

--

INSTRUMENTO 3A

ENTREVISTA A RESPONSABLE

A. DESCRIPTIVOS

1. ¿Cuántos años lleva desempeñando el mismo puesto de trabajo?
2. ¿Cuáles son las principales funciones y responsabilidades del personal de la unidad/departamento de la que es responsable?

B. ANTES DE LA ACCIÓN FORMATIVA

3. ¿Ha tenido alguna intervención en el diseño de las acciones formativas?

Sí	No
Cuál _____	

4. Cuando alguien del personal de su unidad/departamento solicita una de esas acciones formativas, ¿conoce los objetivos, contenidos o duración de la actividad?
5. ¿Están relacionados los objetivos de los cursos con el trabajo que se realiza en su unidad/departamento?
6. ¿Conoce cuáles son los motivos que llevan al personal de su unidad/departamento a solicitar estas acciones?

C. DURANTE LA ACCIÓN FORMATIVA

7. Cuando alguien del personal de la unidad/departamento recibe una acción formativa, ¿habla con él/ella sobre sus expectativas de aplicación de las habilidades aprendidas en las acciones formativas en cuestión?

D. TRAS LA ACCIÓN FORMATIVA: FORMULARIO DE TRANSFERENCIA

8. Indique si el personal de su unidad/departamento dispone de herramientas de apoyo y tecnología adecuadas en el puesto de trabajo para aplicar los conocimientos aprendidos.

1 Sí 2 No 3 Algunos sí, otros no. _____%

9. Indique el grado de aplicabilidad práctica de la acción formativa al puesto de trabajo por parte del personal de su unidad.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

10. Seleccione o piense en cuatro colaboradores; 2 que crea que hayan aplicado la formación al puesto de trabajo y 2 que no.

PERSONAL					
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5
	1	2	3	4	5

11. Aspectos como la motivación o clima laboral han mejorado tras la acción formativa.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

12. La participación en estas acciones formativas ha aumentando la eficacia y el mejor desempeño del trabajo de la unidad/departamento de la que es responsable.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

13. ¿Podría decirme aspectos que hayan mejorado en su unidad/departamento a raíz de la acción formativa?

14. Si fuera participante en el diseño de la evaluación, ¿cómo haría para registrar resultados aplicados al puesto de trabajo tras una acción formativa?

E. NECESIDADES DE FORMACIÓN

15. Indique algunas áreas en las que crea que es necesario implementar acciones formativas para la mejora del servicio de su unidad/departamento.
16. En su unidad/departamento, ¿pueden aparecer nuevas necesidades que no hayan aparecido antes? ¿Se le ocurre cómo se podrían detectar?
17. ¿Qué necesidades detectaba en su unidad/departamento antes de iniciar la acción formativa?
18. Las acciones formativas en las que ha participado el personal de su unidad/departamento, ¿se adaptan a la realidad de los puestos de trabajo?
19. El personal de su unidad, ¿le transmite las necesidades formativas que tiene?
20. ¿Transmite las necesidades que presenta su unidad/departamento a sus responsables?

F. MONITORIZACIÓN DE LA TRANSFERENCIA

21. Tras un curso determinado, ¿estimaría conveniente planificar un seguimiento posterior del curso para comprobar si realmente la formación ha sido útil?
22. ¿Estaría dispuesto/a a participar en un estudio piloto en el que, junto con los formadores de distintos cursos y el personal de evaluación, se programen una serie de prácticas a realizar en el mismo puesto de trabajo?
23. ¿Cree que en dicho estudio se podrían registrar los resultados que mencionó anteriormente con respecto a la transferencia? (claro está, siempre desde la más rigurosa confidencialidad de la procedencia de estos datos).

G. OTROS COMENTARIOS Y SUGERENCIAS

24. ¿Qué ideas, sugerencias y propuestas de mejora podría formular para el desarrollo de las futuras acciones formativas?

INSTRUMENTO 3B

ENTREVISTA AL PERSONAL PARTICIPANTE

A. RECORDATORIO DEL CURSO

1. ¿Recuerda el curso que realizó en el mes de _____?
2. ¿Qué contenidos fueron los que se impartieron en el curso? ¿Le gustó?
3. ¿Cómo se enteró de la existencia de este curso?

B. DESCRIPTIVOS

4. ¿Cuántos años lleva desempeñando el mismo puesto?
5. ¿Cuáles son las principales funciones y responsabilidades de su trabajo?

C. ANTES DE LA ACCIÓN FORMATIVA

6. ¿Ha tenido alguna intervención en el diseño de las acciones formativas?
7. Cuando solicita un curso de formación, ¿conoce los objetivos, contenidos, duración, etc. que la actividad en concreto pretende cubrir?
8. En relación con las funciones que realiza, ¿en qué grado está de acuerdo con los objetivos propuestos para los cursos como el que ha realizado?
9. ¿Cuáles son los motivos que le llevaron a solicitar estas acciones?

D. DURANTE LA ACCIÓN FORMATIVA

10. Cuando recibe una acción formativa, ¿habla con el responsable de su unidad/departamento sobre las expectativas de aplicación de las habilidades aprendidas en las acciones formativas en cuestión?

E. TRAS LA ACCIÓN FORMATIVA: FORMULARIO DE TRANSFERENCIA

11. Dispone de herramientas de apoyo y tecnología adecuadas en el puesto de trabajo para aplicar los conocimientos aprendidos.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

12. Indique el grado de aplicabilidad práctica de la acción formativa a su puesto de trabajo.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

13. Aspectos como la motivación, conocimientos, destrezas y eficacia en su puesto de trabajo o clima laboral, han aumentado tras la acción formativa.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

14. Los aspectos anteriores, han aumentado el rendimiento en el trabajo de la unidad en la que trabaja.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

15. De algún modo, se le reconoce explícita o implícitamente los logros conseguidos en el trabajo tras la acción formativa.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

16. Indique la influencia de los cursos de formación a los que ha asistido en el mejor desempeño del puesto de trabajo.

1 (Nada)	2 (Casi nada)	3 (A veces)	4 (Casi siempre)	5 (Siempre)
-------------	------------------	----------------	---------------------	----------------

17. ¿Podría decirme aspectos que hayan mejorado en el puesto de trabajo a raíz de la acción formativa?

18. Si fuera participante en el diseño de la evaluación, ¿cómo haría para registrar resultados aplicados al puesto de trabajo tras una acción formativa?

F. NECESIDADES DE FORMACIÓN

19. Indique algunas áreas en las que crea que es necesaria implementar acciones formativas para la mejora del servicio de la unidad/departamento en el que trabaja.

20. En su unidad/departamento, ¿pueden aparecer nuevas necesidades que no hayan aparecido antes? ¿Cómo plantearía su detección?

21. ¿Qué necesidades detectaba en su puesto de trabajo antes de iniciar el curso de formación _____?

22. ¿En qué grado se ofrecen tareas realistas y relacionadas con el puesto de trabajo en las acciones formativas?

23. Las acciones formativas en las que ha participado, ¿se han adaptado a la realidad de los puestos de trabajo?

24. ¿Transmite las necesidades formativas que tiene?

G. MONITORIZACIÓN DE LA TRANSFERENCIA

25. Tras un curso determinado, ¿estimaría conveniente planificar un seguimiento posterior del mismo para comprobar si realmente la formación ha sido útil?

26. ¿Estaría dispuesto a participar en un estudio piloto en el que, junto con los formadores de distintos cursos y el responsable de su unidad/departamento, se programen una serie de prácticas a realizar en el mismo puesto de trabajo?

27. ¿Cree que en dicho estudio piloto se podría registrar los resultados que mencionó anteriormente con respecto a la transferencia, claro está siempre desde la más rigurosa confidencialidad de la procedencia de los datos?

H. OTROS COMENTARIOS Y SUGERENCIAS

28. ¿Qué ideas, sugerencias y propuestas de mejora podría formular para el desarrollo de futuras acciones formativas?

INSTRUMENTO 4A

CUESTIONARIO DE EVALUACIÓN DE LA TRANSFERENCIA A CUMPLIMENTAR PREVIAMENTE A LA ACCIÓN FORMATIVA POR EL PARTICIPANTE

Nombre del Curso _____

Fechas de impartición _____

Datos de la empresa:

Nombre _____

Domicilio _____

Teléfono _____

Fax _____

Datos del participante:

Nombre _____

Puesto _____

Fecha _____

1. ¿Sabe de qué trata este curso?

1 (Nada)	2 (Poco)	3 (Bastante)	4 (Mucho)
-------------	-------------	-----------------	--------------

2. En caso de que lo conozca, ¿por qué medio ha conocido el contenido del curso?

Se lo han comunicado personalmente (su organización)	.
Por carta o circular	.
Por el tablón de anuncios	.
A través de los propios compañeros	.
A través del sindicato	.
Otros: especificar	.

3. ¿Por qué cree que resulta necesario asistir al curso?

Lo necesito para mi trabajo actual	.
Lo necesito para un nuevo trabajo o puesto que se me va a asignar	.
De cara a mi promoción	.
Para ampliar conocimientos sin una repercusión concreta en el trabajo	.

Como compensación/estímulo a mi trabajo	.
Como beneficio social	.
Por la incorporación de nuevas tecnologías	.
Por la reconversión de actividades de la empresa	.

4. Indique con una X si utiliza o practica los conocimientos, habilidades o actividades que se mencionan a continuación. En caso de que los utilice, puntúe de 1 a 5 el grado en que considera que se cumple el indicador, siendo el mínimo 1 y el máximo 5; en caso de que no los utilice, indique la razón.

a) Conocimiento, habilidad o actividad 1. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

b) Conocimiento, habilidad o actividad 2. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

c) Conocimiento, habilidad o actividad 3. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

d) Conocimiento, habilidad o actividad 4. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

Etc.....

INSTRUMENTO 4B

CUESTIONARIO DE EVALUACIÓN DE LA TRANSFERENCIA A CUMPLIMENTAR FINALIZADA LA ACCIÓN FORMATIVA POR EL PARTICIPANTE

Nombre del Curso _____

Fechas de impartición _____

Datos de la empresa:

Nombre _____

Domicilio _____

Teléfono _____

Fax _____

Datos del participante:

Nombre _____

Puesto _____

Fecha _____

El objetivo de este cuestionario es evaluar la transferencia de conocimientos del curso al puesto de trabajo. Es decir, si en su trabajo actual aplica y le resultan de utilidad los conocimientos adquiridos y las prácticas realizadas durante el curso.

1. Indique con una X si utiliza o practica los conocimientos, habilidades o actividades que se mencionan a continuación. En caso de que los utilice, puntúe de 1 a 5 el grado en que considera que se cumple el indicador, siendo el mínimo 1 y el máximo 5; en caso de que no los utilice, indique la razón.

a) Conocimiento, habilidad o actividad 1. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino		<input type="checkbox"/> no lo necesito todavía		<input type="checkbox"/> no es aplicable

b) Conocimiento, habilidad o actividad 2. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino		<input type="checkbox"/> no lo necesito todavía		<input type="checkbox"/> no es aplicable

c) Conocimiento, habilidad o actividad 3. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

d) Conocimiento, habilidad o actividad 4. ¿Lo utiliza o practica?

<input type="checkbox"/> Sí	1	2	3	4	5
<input type="checkbox"/> No, porque	<input type="checkbox"/> no lo domino	<input type="checkbox"/> no lo necesito todavía	<input type="checkbox"/> no es aplicable		

Etc.....

2. Indique qué ventajas ha supuesto el curso para su trabajo (mayor seguridad, nuevas aplicaciones, más eficiencia, etc.). Procure concretar para qué lo utiliza y la mejora específica que haya notado.

INSTRUMENTO 5

EVALUACIÓN DE LA TRANSFERENCIA

1. Valore de 1 (nada) a 5 (mucho) el grado en que ha sido útil la acción formativa para su puesto de trabajo.

2. Si ha respondido 5 en la pregunta 1, por favor, explicita en una línea la razón.

3. Si no ha respondido 5 en la pregunta 1, por favor, explicita en una línea la razón.

4. Indique cualquier comentario que considere respecto a la utilidad de la acción formativa en su puesto.