

**PROGRAMAS DE FORMACIÓN EN RECONOCIMIENTO DE
NIVELES COMUNES DE REFERENCIA PARA PROFESORES DE
CENTROS BILINGÜES**

José Cuadrado Moreno
Universidad Pablo de Olavide, Sevilla, España
josecuadrado@upo.es

María Reyes Fernández
Universidad Pablo de Olavide, Sevilla, España
mreyfer@upo.es

Rafael Muñoz Rubio
Universidad Pablo de Olavide, Sevilla, España
munozrubio@gmail.com

This paper presents the design of a training programme for secondary education teachers in bilingual schools. The goal of this programme was to train participants to recognise the common reference levels (CRLs), as established in the Common European Framework of Reference (CEFR), in oral performances. The programme, based on North et al. (2009), consisted of two phases: familiarisation with the CEFR and standardisation/rating of oral samples. The conclusions are that, firstly, the participants' ability to identify and redistribute CRL descriptors depends on the number of both the descriptors and the communicative language activities they deal with. Secondly, there exists a direct relation between the number of recognition tasks and the degree of accuracy in CRL recognition. Thirdly, when classifying oral performances, the participants employed, apart from CRLs, additional criteria. Fourthly, the CRL recognition process was statistically controlled. The study shows that the majority of the sample of teachers needs further familiarisation with the CEFR and additional training in CRL recognition. Finally, several suggestions concerning CRL recognition training are made to educational authorities and teacher trainers.

Key words: Common European Framework of Reference, teacher training, speaking, reliability, familiarisation, rating.

Este artículo presenta el diseño de un programa de formación dirigido a profesores de educación secundaria en centros bilingües. El objetivo del programa era conseguir que los participantes reconocieran los niveles comunes de referencia (NCR), establecidos en el Marco Común Europeo de Referencia (MCER), en actuaciones orales. El programa, basado en North et al. (2009), constó de dos fases: familiarización con el MCER y estandarización/valoración de muestras orales. Se concluye, en primer lugar, que la capacidad de los participantes de identificar y reordenar los descriptores de los NCR depende del número de descriptores y de las actividades lingüísticas comunicativas con las que tenían que operar. En segundo lugar, existe una relación directa entre el número de actividades de reconocimiento y el grado de precisión en el reconocimiento del NCR. En tercer lugar, cuando clasificaban actuaciones orales, los participantes emplearon, además de los NCR, algún otro criterio. En cuarto lugar, el proceso de reconocimiento de los NCR estuvo estadísticamente controlado. El estudio indica que la mayoría de la muestra de profesores necesita más familiarización con el MCER y formación adicional en el reconocimiento de los NCR. Finalmente, se presentan varias sugerencias acerca de la formación en el reconocimiento de los NCR dirigidas a las autoridades educativas y los formadores de profesores.

Palabras clave: Marco Común Europeo de Referencia, formación del profesorado, producción oral, fiabilidad, familiarización, valoración.

1. Introducción

En los últimos años los legisladores españoles han introducido en sus normas de política lingüística referencias al *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación* [MCER] (Council of Europe 2001). El MCER ha sido, por ejemplo, un “referente clave” en el desarrollo del vigente currículo del área de lengua extranjera en la etapa de educación secundaria obligatoria (E.S.O.) (España 2007: 741-742). El legislador andaluz también ha especificado que

en el bachillerato se continúa el proceso de aprendizaje de la lengua extranjera y se procurará que al finalizarla los alumnos y las alumnas hayan consolidado las destrezas y habilidades necesarias para la comunicación, al menos en el nivel B1, según se describe este nivel en el Marco Común Europeo de Referencia para las Lenguas (Andalucía 2008: 118).

Por tanto, en Andalucía el conocimiento del MCER y su aplicación a la evaluación de la adquisición lingüística del alumnado son una exigencia normativamente impuesta a los profesores de E.S.O.

Por otro lado, en los oficialmente denominados “Centros Bilingües” de la Comunidad Autónoma de Andalucía suelen coexistir grupos que imparten el programa bilingüe y uno o más grupos que no participan en este programa (Andalucía 2006: 12). De hecho, es usual que en la mayoría de los centros andaluces que imparten la etapa de E.S.O. (sean o no centros bilingües) coexistan varios grupos del mismo curso, que pueden tener diferentes profesores para la materia de lengua extranjera. Esta situación constituye un riesgo para la fiabilidad de las puntuaciones que estos profesores proporcionan.

La exigencia de conocimiento de los niveles comunes de referencia establecidos en el MCER que las normas andaluzas imponen a los profesores de E.S.O., y la convicción de que un desconocimiento de estos niveles puede influir en la fiabilidad de las puntuaciones de los profesores de lengua extranjera en esa etapa, llevó a los autores a plantearse dos preguntas: (i) ¿Qué grado de conocimiento tiene este profesorado del MCER y de los denominados niveles comunes de referencia?, y (ii) ¿Qué influencia puede tener un hipotético desconocimiento de los niveles comunes de referencia en las puntuaciones que asigna este profesorado? Para responder estas preguntas se diseñó un programa de formación en reconocimiento de los niveles comunes de referencia en actuaciones orales, dirigido al profesorado de educación secundaria en centros bilingües, con la esperanza de que la participación en el programa permitiera que los profesores conocieran los niveles comunes de referencia establecidos en el MCER y, por consiguiente, que sus puntuaciones en las actividades de expresión e interacción oral que realizan ordinariamente en el aula tuvieran una elevada fiabilidad. Este artículo presenta el proceso de desarrollo de este programa de formación.

2. Revisión de la literatura

Los autores percibieron que el diseño de este programa de formación estaba relacionado con diferentes aspectos relacionados con la evaluación de la competencia en una lengua extranjera:

1. La fiabilidad de las puntuaciones que asignan los profesores de lengua extranjera en la etapa de educación secundaria obligatoria a las actuaciones orales.
2. El conocimiento del MCER y otras publicaciones basadas en este documento.
3. El control estadístico de los procesos.

El proceso de puntuación de una actuación oral está directamente relacionado con lo que en evaluación educativa y psicológica se denomina la fiabilidad, que las *Standards for Educational and Psychological Testing* definen como “the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable, and repeatable for an individual test taker” (American Educational Research Association, American Psychological Association, y National Council on Measurement in Education 1999: 180). La importancia de la fiabilidad de las puntuaciones para la prueba es tal que Bachman y Palmer (1996: 19) la describen como una propiedad crítica de la prueba. Un aspecto relacionado con la fiabilidad de las puntuaciones en una prueba es la consistencia de las valoraciones realizadas por uno o más examinadores, cuando cada uno de ellos puntúa independientemente la misma prueba. Otro aspecto también relacionado con la fiabilidad es la consistencia interna de estas puntuaciones, la cual puede ser determinada con una única convocatoria mediante un parámetro como el coeficiente alfa de Cronbach, que proporciona “the average of all the possible split-half coefficients for a given test” (Cronbach 1951: 300) y cuyos valores oscilan entre 0 y 1. Basándose en las propuestas para la interpretación de los coeficientes de fiabilidad de Murphy y Davidshofer (1994: 104), se asignaron las siguientes interpretaciones a los valores del coeficiente alfa de Cronbach: inaceptable (valores inferiores a 0,6), baja (valores entre 0,61 y 0,75), intermedia (valores entre 0,76 y 0,89) y elevada (valores superiores a 0,9). Aparte de ser un indicador de la consistencia interna de una prueba, el coeficiente alfa de Cronbach también ha sido empleado como un estimador de la unidimensionalidad de una prueba, es decir, como un estimador de la existencia de una única propiedad latente que subyace a la prueba. De hecho, Cronbach (1951: 320) afirmó que “for a test to be interpretable, however, it is no essential that all items be factorially similar. What is required is that a large proportion of the test variance should be attributable the principal factor running through the test”. Según Cronbach, el coeficiente alfa “estimates the proportion of the test variance due to all common factors

among the items” e indica “how much the test score depends upon the general and group, rather than on the item specific, factors” (*ibid.*).

Para los autores, la familiarización con el MCER y con los niveles comunes de referencia debía ser un objetivo principal del programa de formación (North 2000; Little 2006). El MCER es un documento publicado por el Consejo de Europa en 2001, que Little (2006: 167) describe como sigue:

The CEFR is a descriptive scheme that can be used to analyse L2 learners' needs, specify L2 learning goals, guide the development of L2 learning materials and activities, and provide orientation for the assessment of L2 learning outcomes.

En su capítulo 3 el MCER presenta seis niveles “appropriate to the organisation of language learning and the public recognition of achievement” (Council of Europe 2001: 22-23), que denomina niveles comunes de referencia A1, A2, B1, B2, C1 y C2. El MCER tuvo el acierto de describir explícitamente los niveles comunes de referencia en términos generales mediante párrafos holísticos (Cuadro 1, p. 24) y párrafos autodescriptivos (Cuadro 2, pp. 26-27). Además, en los capítulos 4 y 5 del MCER se incluyeron cuadros que asignan un descriptor a cada nivel común de referencia en una serie de actividades y competencias lingüísticas comunicativas. El enfoque utilizado para la construcción de los cuadros fue conductista en el sentido de que los descriptores de los cuadros describen conductas comunicativas (Little 2006: 168). Estos descriptores habían sido recopilados de otras escalas y seleccionados por profesores en actividades cuyo objetivo era construir escalas basadas en la *percepción* que estos profesores tenían de los diferentes niveles de competencia, pero no en una descripción rigurosa de los niveles de competencia de los aprendientes (North 2000: 573).

El MCER ha sido desarrollado por una infinidad de documentos (por ejemplo, Alderson 2002; Cambridge ESOL 2003; Corrigan 2007; Broeder y Martyniuk 2008). Entre estos documentos de desarrollo destaca, para nuestros fines, North *et al.* (2009), al que nos referiremos como el *Manual*. El objetivo explícito del *Manual* es

to help the providers of examinations to develop, apply and report transparent, practical procedures in a cumulative process of continuing improvement in order to situate their examination(s) in relation to the Common European Framework (CEFR) (North *et al.* 2009: 1).

Para relacionar una prueba con el MCER, el *Manual* recomienda seguir un proceso constituido por cinco fases (North *et al.* 2009: 9):

1. Familiarización (*familiarisation*), cuyo objetivo es “to ensure that participants in the linking process have a detailed knowledge of the CEFR, its levels and illustrative descriptors” (p. 10).
2. Especificación (*specification*), que constituye “a self-audit of the coverage of the examination (content and tasks types) profiled in relation to the categories presented in CEFR Chapter 4 and CEFR Chapter 5” (p. 10).
3. Formación en estandarización/establecimiento de muestras normalizadas (*training in standardisation/benchmarking*), cuyas actividades tienen como objetivo “(a) to help panellists to implement a common understanding of the CEFR levels; (b) to verify that such a common understanding is achieved, and (c) to maintain that standard over time” (p. 37).
4. Establecimiento de puntuaciones límite (*standard setting*) entre los diferentes niveles comunes de referencia.

5. Validación (*validation*), que está relacionada con “the body of evidence put forward to convince the test users that the whole process and its outcomes are trustworthy” (p. 90).

Dada la relevancia de las fases de familiarización y de formación en estandarización/establecimiento de muestras normalizadas para el programa que fue diseñado, a continuación desarrollamos un poco más extensamente estas dos fases. El *Manual* recomienda realizar las siguientes actividades durante la fase de familiarización en aproximadamente tres horas (North *et al.* 2009: 23):

1. Presentación breve por parte del coordinador (30 minutos).
2. Actividad introductoria y discusión (45 minutos).
3. Actividad cualitativa, incluyendo trabajo en grupo (45 minutos).
4. Preparación para la valoración (45 minutos).
5. Conclusión (15 minutos).

La Tabla 1 presenta el calendario que North *et al.* (2009: 45) proponen para un programa de formación cuyo objetivo es estandarizar los juicios de una serie de evaluadores con respecto a muestras de actuación oral.

Fases
Fase 1: Familiarización (60 minutos)
Fase 2: Práctica con muestras normalizadas
Paso 1: Aclaración con aproximadamente tres actuaciones normalizadas (60 minutos)
Paso 2: Práctica controlada con aproximadamente tres actuaciones normalizadas (60 minutos)
Paso 3: Fase libre con aproximadamente tres actuaciones normalizadas (60 minutos)
Fase 3: Puntuación de muestras locales
Puntuación individual y discusión en grupo de aproximadamente tres actuaciones (60 minutos)
Puntuación individual de aproximadamente cinco actuaciones más (60 minutos)

Tabla 1. Calendario para un programa de formación en estandarización de juicios acerca de actuaciones orales (North et al. 2009: 45).

El *Manual* ha constituido la base de múltiples programas de formación cuyo objetivo ha sido estandarizar los juicios de evaluadores acerca de los niveles comunes de referencia, como Lepage y North (2005) o Breton, North y Lepage (2008). Sin embargo, un dato que se echa de menos en la literatura sobre estos programas es la existencia de un parámetro a partir del cual se pueda inferir que el participante ha adquirido la competencia suficiente para poder determinar, con un elevado grado de fiabilidad, el nivel común de referencia de una determinada actuación oral.

El último aspecto con el que se relacionó el programa de formación diseñado fue con el control estadístico de los procesos, cuyos orígenes están en la preocupación actual por mejorar la calidad de los servicios ofrecidos a los usuarios y en la consiguiente necesidad de disponer de métodos que permitan descubrir si una determinada propiedad relacionada con la calidad del servicio evoluciona positiva o negativamente. El estudio de la evolución de estas propiedades relacionadas con la calidad de un proceso se ha convertido en el objeto de un nuevo campo, denominado control estadístico de los procesos, que, según Stoumbos *et al.* (2000: 992) se ocupa de “some statistical methods used extensively to monitor and improve the quality and productivity of manufacturing processes and service operations”. El objetivo de los especialistas en el control estadístico de los procesos es conseguir procesos estadísticamente controlados, es decir, procesos donde “the probability distribution representing the quality characteristic is constant over time” (Woodall 2000: 341). Una de las herramientas más utilizadas en el análisis estadístico de los procesos es el gráfico de control (*control chart*), un gráfico “used to detect any change in a process that may affect the quality of the output” (Stoumbos *et al.* 2000: 992). Según Triola (2006: 742), se debe concluir que un proceso está estadísticamente descontrolado si en el gráfico de control que representa el proceso posee una o más de las siguientes características:

1. Existe un esquema repetitivo, una tendencia o un ciclo que no es aleatorio.
2. Existe un punto que está fuera de la zona entre los límites de control superior e inferior.
3. Existen ocho puntos consecutivos sobre o bajo la línea central.

La Figura 1 ilustra el uso de los gráficos de control para determinar si un proceso está estadísticamente controlado. El gráfico de la Figura 1 muestra la evolución del retraso en una serie de envíos. Puesto que uno de los puntos que representa un envío está situado por encima de la línea de control superior, se debe concluir, de acuerdo con la regla (2) de Triola (2006: 742), que el proceso está estadísticamente descontrolado.

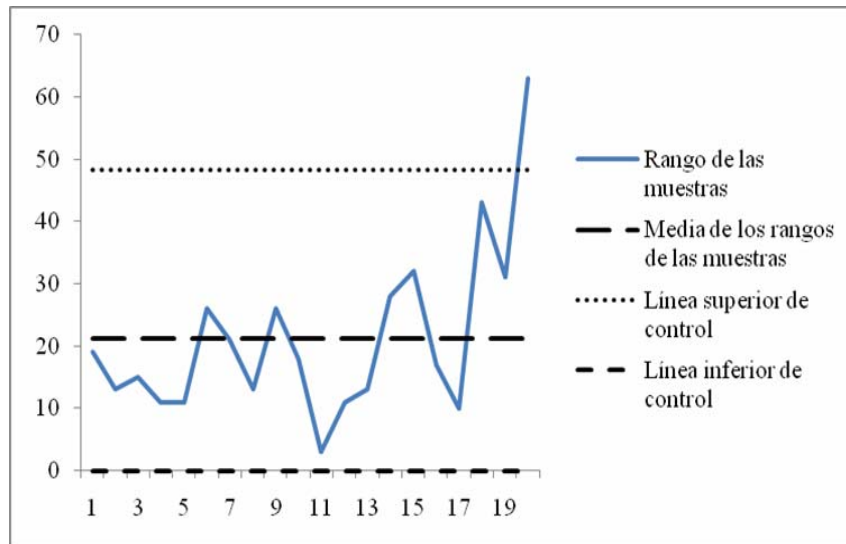


Figura 1. Ejemplo de gráfico de control (gráfico de rangos).

Dado que un programa de formación del profesorado de lengua inglesa de centros bilingües cuyo objetivo es identificar los niveles comunes de referencia de una serie de actuaciones orales es un proceso, los autores consideraron pertinente aplicar los métodos procedentes del control estadístico de los procesos al análisis del grado de aprovechamiento final de los participantes en ese curso de formación.

A partir de la revisión de la literatura, se pueden establecer una serie de preguntas para la investigación:

1. ¿Están familiarizados los profesores de lengua inglesa con los niveles comunes de referencia?
2. ¿Son fiables las puntuaciones que los profesores de lengua inglesa de centros bilingües asignan a las actuaciones orales? A su vez, esta pregunta (2) se puede subdividir en dos subpreguntas:
 - 2.1. ¿Cuál es el grado de consistencia entre las diferentes puntuaciones que un grupo de profesores de lengua inglesa de centros bilingües han asignado a las mismas actuaciones orales de un grupo de alumnos de un centro bilingüe?
 - 2.2. ¿Cuál es el grado de consistencia interna dentro de las puntuaciones que un grupo de profesores de lengua inglesa de centros bilingües han asignado a las mismas actuaciones orales de un grupo de alumnos de un centro bilingüe?
3. ¿El proceso de identificación de niveles en actuaciones orales que realizan los profesores de lengua inglesa está estadísticamente controlado?

3. Metodología

Con el fin de responder a estas preguntas de investigación, los autores diseñaron un programa de formación en estandarización/puntuación de muestras locales de actuaciones orales en lengua inglesa, dirigido a profesores de centros bilingües (de la etapa de la educación secundaria obligatoria) de toda la provincia de Sevilla. El programa se realizó en el Centro de Profesores de Sevilla durante el mes de noviembre de 2009. Puesto que el número de solicitantes procedentes de centros bilingües fue inferior a las plazas ofertadas, la entidad organizadora permitió que se inscribieran otros profesores de educación secundaria obligatoria de centros sin programas bilingües, de bachillerato y de escuelas oficiales de idiomas. El número final de participantes fue de 16, aunque algunas actividades sólo fueron realizadas por 15 participantes.

El programa de formación siguió, en la medida de lo posible, las indicaciones del *Manual*, de tal manera que se organizó como sigue:

1. Familiarización (4 horas).
2. Práctica con muestras normalizadas.
 - a. Paso 1: Aclaración (3 horas).
 - b. Paso 2: Práctica controlada (1 hora y 30 minutos).
 - c. Paso 3: Fase libre (1 hora y 30 minutos).

Puesto que el programa de formación estaba dirigido al profesorado de educación secundaria obligatoria de centros bilingües, se decidió que uno de los objetivos del programa fuera que los participantes pudieran reconocer el nivel común de referencia en muestras locales de alumnos de educación secundaria obligatoria. Los autores se encontraron con el problema de que la legislación educativa andaluza que regula la etapa de la E.S.O. no incluye ninguna referencia a objetivos del aprendizaje de una lengua extranjera expresado en términos de los niveles comunes de referencia. Teniendo en cuenta que en Andalucía el legislador optó por establecer el nivel común de referencia mínimo B1 al finalizar el Bachillerato (Orden de 24 de julio de 2006: 118), los autores decidieron establecer como objetivo mínimo para la etapa de la E.S.O. la adquisición de un nivel común de referencia A2 en la lengua inglesa.

3.1. Fase de familiarización

Durante la fase de familiarización los participantes realizaron las siguientes actividades:

1. Presentación del programa de formación por el coordinador (30 minutos).
2. Presentación del MCER (30 minutos).
3. Autoevaluación del nivel de competencia lingüística en la lengua inglesa.
4. Actividad de aprendizaje consistente en la reordenación de los descriptores del “Cuadro 2. Niveles comunes de referencia: cuadro de autoevaluación” (Consejo de Europa 2002: 30-31) en niveles comunes de referencia.

5. Actividad de aprendizaje donde los participantes reordenaban los descriptores del cuadro “Expresión escrita en general” (*ibid.*, p. 64) en niveles comunes de referencia.

6. Actividad de aprendizaje en la que los participantes volvían a reordenar los descriptores del cuadro “Expresión oral en general” (*ibid.*, p. 61) en niveles comunes de referencia.

Las hipótesis nulas h_0 que se deseaban demostrar con la realización de las actividades de familiarización (4), (5) y (6) eran que no existía una diferencia entre la distribución por niveles de los descriptores de los cuadros del MCER y la distribución que de estos descriptores realizaría cada participante (o, dicho con otras palabras, que los participantes ordenarían los descriptores del mismo modo en que aparecen en los cuadros del MCER). Para contrastar estas hipótesis nulas, se utilizó la prueba (no paramétrica) de los signos (Cochran 1937; Wilcoxon 1945; Triola 2006: 678-686), ya que, hasta donde los autores saben, no se dispone de datos acerca del grado de familiarización del profesorado de E.S.O. con el MCER

3.2. Formación en estandarización/puntuación de muestras locales

Durante la realización de las actividades de aprendizaje de las fases 2 (práctica con muestras normalizadas) y 3 (puntuación de muestras locales) los participantes dispusieron de una copia del “Cuadro 3 Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada” del MCER (Consejo de Europa 2002: 32-33), con el fin de que lo pudieran consultar cuando lo desearan durante la realización de esas actividades de aprendizaje.

A continuación, se realizaron las siguientes actividades:

a. Actividad 1 (*aclaración*), durante la cual se presentaron seis videos, cada uno de los cuales mostraba a una pareja de usuarios de la lengua inglesa de cada uno de los niveles comunes de referencia. A continuación, los participantes y uno de los autores justificaban el nivel de la pareja de usuarios a partir de la información contenida en el “Cuadro 3 Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada”.

b. Actividad 2 (*práctica controlada*), durante la cual los participantes vieron tres vídeos, cada uno de los cuales presentaba una pareja de usuarios de los niveles A1, A2 o B1. Los participantes debían asignar un nivel común de referencia a cada pareja. Tras recoger las respuestas de cada actividad, la formadora identificaba el nivel común de referencia de cada pareja y justificaba la asignación realizada.

c. Actividad 3 (*actividad libre*), en la que los participantes debían identificar el nivel común de referencia de cuatro actuaciones orales de alumnos de dos institutos de educación secundaria de Montequinto (Sevilla), grabados por parejas. Los niveles comunes de referencia de los alumnos que aparecían en las grabaciones eran los siguientes:

- Grabación 1: A2 y A1.
- Grabación 2: A2 y A2.

Con el fin de responder la pregunta de investigación (2), que está relacionada con la fiabilidad de las puntuaciones en la prueba, para estimar la consistencia interna de la prueba se calculó el alfa de Cronbach (1951: 300).

4. Análisis

4.1. Análisis de los datos obtenidos durante la fase de familiarización

La hipótesis nula h_0 de que no existían diferencias entre la distribución por niveles de los descriptores de los cuadros del MCER y la distribución que realizaron los participantes fue contrastada mediante el método tradicional (Triola 2006: 401) aplicando la prueba de los signos ($\alpha = 0.05$). En las Tablas 2, 3 y 4 aparecen los estadísticos de la prueba obtenidos y los valores críticos para la contrastación de la hipótesis nula correspondiente en cada participante. En las Tablas 2, 3 y 4 el asterisco indica que con el estadístico de la prueba correspondiente es imposible obtener para ese participante un valor en la región crítica y que, por tanto, se debe aceptar, necesariamente, la hipótesis nula en el caso de ese participante. En estas mismas tablas un valor entre paréntesis representa un valor en la región crítica, es decir, que la hipótesis nula en ese participante debe ser rechazada.

Participante	Estadístico de la prueba	Valor crítico a la izquierda	Valor crítico a la derecha
1	(14)	2	11
2	(19)	4	15
3	(15)	3	12
4	(9)	1	8
5	(17)	4	13
6	(11)	1	10
7	2	*	*
8	4	*	*
9	(8)	0	8
10	(11)	1	10
11	(14)	2	12
12	(15)	3	12
13	(11)	1	10
14	(8)	0	8
15	(9)	1	8

Tabla 2. Estadísticos de la prueba de los signos y valores críticos para los datos procedentes de la actividad de familiarización 4.

Participante	Estadístico de la prueba	de la	Valor crítico a la izquierda	Valor crítico a la derecha
1	2		*	*
2	2		*	*
3	2		*	*
4	2		*	*
5	2		*	*
6	0		*	*
7	2		*	*
8	2		*	*
9	0		*	*
10	2		*	*
11	3		*	*
12	2		*	*
13	2		*	*
14	0		*	*
15	3		*	*

Tabla 3. Estadísticos de la prueba de los signos y valores críticos para los datos procedentes de la actividad de familiarización 5.

Participante	Estadístico de la prueba	Valor crítico a la izquierda	Valor crítico a la derecha
1	1	*	*
2	1	*	*
3	1	*	*
4	1	*	*
5	1	*	*
6	-	-	-
7	1	*	*
8	1	*	*
9	1	*	*
10	1	*	*
11	1	*	*
12	1	*	*
13	1	*	*
14	1	*	*
15	1	*	*

Tabla 4. Estadísticos de la prueba de los signos y valores críticos para los datos obtenidos en la actividad de familiarización 6.

Del examen de estas tablas se pueden extraer tres conclusiones principales con respecto a los datos obtenidos acerca de la familiarización:

Existen suficientes pruebas que permiten rechazar la mayoría (para ser precisos, el 86,66%) de las hipótesis nulas de que los participantes son capaces de reordenar los descriptores del “Cuadro 2. Niveles comunes de referencia: cuadro de autoevaluación”. El origen de este resultado puede estar en dos hechos: (i) el Cuadro 2 consta de treinta y seis descriptores, y (ii) la naturaleza de la actividad de formación 4, que obligaba a que los participantes colocaran todos los descriptores, por lo que un error en la colocación de un descriptor implicaba automáticamente que otro descriptor fuera también mal colocado.

No existen suficientes datos que permitan rechazar las hipótesis de que los participantes pueden reconstruir el cuadro “Expresión escrita en general” del MCER.

No existen suficientes datos que permitan rechazar las hipótesis de que los participantes pueden reconstruir el cuadro “Expresión oral en general” del MCER.

Por consiguiente, del análisis de los datos obtenidos durante la fase de familiarización se puede concluir que los profesores de lengua inglesa de nuestra muestra fueron significativamente capaces de identificar y reordenar los descriptores de los niveles comunes de referencia de actividades lingüísticas comunicativas aisladas y descritas mediante un único criterio. Sin embargo, los participantes en el estudio fueron significativamente incapaces de reordenar los descriptores de los niveles comunes de referencia de seis actividades lingüísticas comunicativas diferentes cuando sus descriptores fueron presentados simultáneamente. Es decir, parece que el grado de acierto en las actividades de familiarización estuvo relacionado con el número de criterios (y el consiguiente número de descriptores) que los profesores debían manejar durante la realización de la actividad de familiarización.

4.2. Análisis de los datos obtenidos en la fase de estandarización/puntuación de muestras locales

La Figura 2 presenta el porcentaje de niveles comunes de referencia correctamente identificados por cada participante 1-16 en las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales, donde los participantes observaban a un total de diez interlocutores.

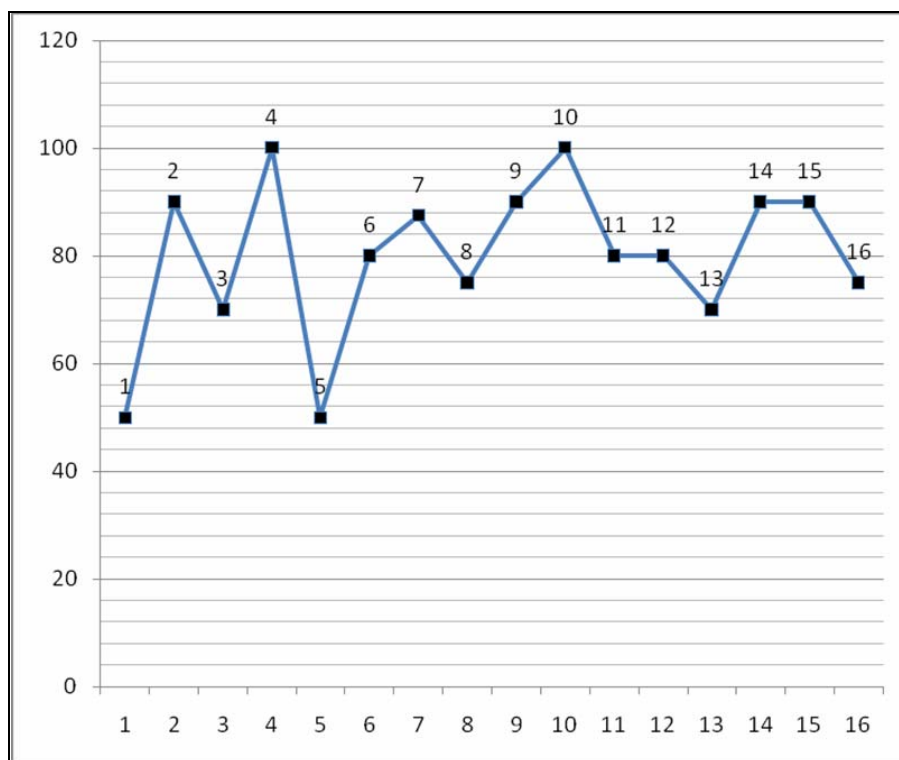


Figura 2. Porcentaje de niveles comunes de referencia identificados correctamente por cada participante 1-16 en las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales.

Se puede observar que durante las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales (i) únicamente los participantes 4 y 10 (el 12,50% de la muestra) fueron capaces de identificar correctamente el nivel común de referencia de los diez interlocutores observados, (ii) cuatro participantes (el 25% de la muestra) fueron capaces de identificar el 75% de los niveles comunes de referencia de los interlocutores observados, y (iii) dos participantes (el 12,50% de la muestra) sólo pudieron identificar el 50% de los niveles comunes de referencia de los interlocutores observados.

La Figura 3 presenta el porcentaje de niveles comunes de referencia correctamente identificados por cada participante 1-16 en la actividad 3 (actividad libre) de la fase de formación en estandarización/puntuación de muestras locales.

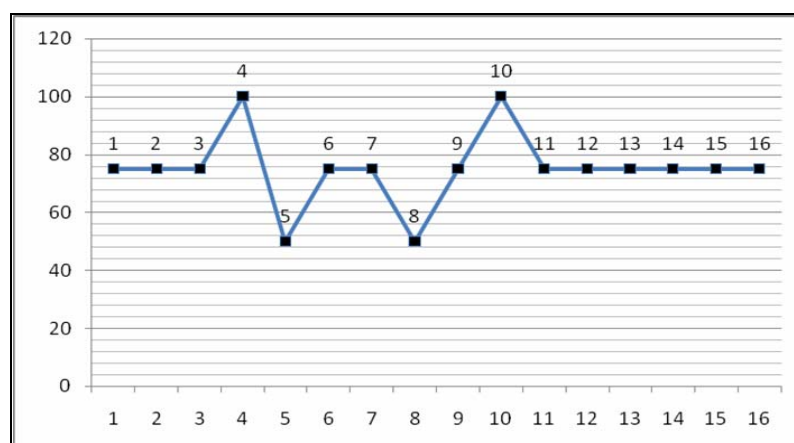


Figura 3. Porcentaje de niveles comunes de referencia correctamente identificados por cada participante 1-16 en la actividad 3 (actividad libre).

En el caso de la actividad 3 se observa que (i) únicamente los participantes 4 y 10 (el 12,50% de la muestra) fueron capaces de identificar correctamente el nivel común de referencia de los cuatro interlocutores observados, (ii) doce participantes (el 75% de la muestra) fueron capaces de identificar el 75% de los niveles comunes de referencia de los interlocutores observados, y (iii) dos participantes (el 12,50% de la muestra) sólo pudieron identificar el 50% de los niveles comunes de referencia de los interlocutores observados.

La experiencia del programa de formación, por consiguiente, mostró una tendencia en los participantes a una identificación correcta del nivel común de referencia a medida que aumentaba el número de actividades de aprendizaje realizadas durante la fase de formación en estandarización/puntuación de muestras locales.

En este punto se plantea la cuestión del establecimiento de una puntuación límite que puede ser utilizada en los programas de formación en reconocimiento de niveles comunes de referencia como base para decidir acerca de la expedición de certificados de capacidad de reconocimiento de niveles comunes de referencia. Los autores realizan las siguientes propuestas:

Considerar el porcentaje de identificaciones correctas como una indicación del grado de competencia en identificación de niveles comunes de referencia.

En los casos en los que la puntuación que asigna un examinador sea la base de decisiones importantes (como la puntuación en la subprueba de expresión oral del tercer ejercicio de la fase general de la futura prueba de acceso a la universidad, regulada en España 2008: 46934), considerar un porcentaje del 90% como puntuación límite mínima para que los examinadores obtengan un certificado de capacidad de reconocimiento de niveles comunes de referencia, aunque un porcentaje del 95% debería ser el porcentaje mínimo deseable. El origen de esta propuesta está en la sugerencia de Nunnally (1978: 246), que propuso emplear un índice de fiabilidad mínimo de 0,90 para decisiones importantes basadas en puntuaciones en una prueba, aunque para Nunnally (1978: 246) el índice de fiabilidad mínimo deseable en este caso debería ser 0,95. Hay que destacar el hecho de que si se hubiera establecido en el programa objeto de esta investigación un porcentaje mínimo de identificaciones correctas para obtener una certificación de capacidad de reconocimiento de niveles comunes de referencia, sólo los participantes 4 y 10 (el 12,5% de la muestra) la hubieran obtenido.

4.3. Análisis de la fiabilidad de las puntuaciones de los participantes durante la fase de formación en estandarización/puntuación de muestras locales e interpretación

En la Tabla 5 aparecen los valores del coeficiente alfa de Cronbach de las valoraciones que los participantes realizaron de cada interlocutor que aparecía en las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales y la interpretación que se puede realizar de cada uno de estos coeficientes alfa de Cronbach basándose en Murphy y Davidshofer (1994: 104).

Actividad	Interlocutor	Coefficiente alfa de Cronbach	Grado de fiabilidad
1	1	0,769	Bajo
	2	0,568	Inaceptable
	3	0,789	Bajo
2	4	0,151	Inaceptable
3	5	0,671	Bajo
	6	0,817	Intermedio
4	7	0,906	Elevado
	8	0,477	Inaceptable
5	9	-0,677	
	10	0,648	Bajo

Tabla 5. Grado de fiabilidad de las valoraciones de cada interlocutor en las actividades 2 y 3 e interpretación.

Del análisis de la Tabla 5 se puede concluir que en la inmensa mayoría de los casos es baja o inaceptable la consistencia interna de las valoraciones realizadas por los participantes de cada uno de los interlocutores. La causa probable de esta mayoritaria inconsistencia en las valoraciones puede radicar en que el concepto que cada participante asociaba a cada nivel común de referencia en una actividad de expresión e interacción orales, y que constituía la base que guiaba la valoración de cada interlocutor, variaba a lo largo de las actividades. Esta supuesta diferencia fue el origen probable de esos coeficientes alfa de Cronbach bajos o inaceptables.

La Tabla 6 presenta el coeficiente alfa de Cronbach de las valoraciones realizadas por los interlocutores de (a) el conjunto de interlocutores que aparecieron en la actividad 2, (b) el conjunto de interlocutores observados en la actividad 3, y (c) el conjunto de todos los interlocutores que aparecieron en las actividades 2 y 3.

Actividades	Coficiente alfa de Cronbach	Grado de fiabilidad
Actividad 2	0,721	Bajo
Actividad 3	0,175	Inaceptable
Actividades 2 y 3	0,669	Bajo

Tabla 6. Grado de fiabilidad de las valoraciones de los interlocutores en las actividades 2 y 3 e interpretación.

Es destacable en la Tabla 6 el bajo valor de la consistencia interna de las valoraciones realizadas en la actividad 3 (actividad libre), en la que los participantes debían valorar muestras locales. Es decir, precisamente en la actividad donde se debería haber alcanzado una consistencia interna más elevada es en la que la consistencia interna es más baja. La razón probable de esta discrepancia es la probable diferencia entre el concepto que cada participante asociaba a cada nivel común de referencia en una actividad de expresión e interacción orales y los conceptos que el resto de participantes asociaba a esos mismos niveles comunes de referencia. O, dicho de otra manera, los conceptos que constituían la base de las valoraciones de los participantes no estaban estandarizados y no se correspondían con los descriptores de los niveles comunes de referencia para las actividades de expresión e interacción orales establecidos en el MCER.

En resumen, las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales no fueron pruebas unidimensionales ya que, probablemente, los participantes emplearon diferentes constructos durante la evaluación y el constructo utilizado iba variando conforme iban realizando las actividades.

4.4. El control de los procesos estadísticos en la fase de estandarización/puntuación de muestras locales

La Figura 4 es un gráfico de control \bar{x} , que representa las proporciones de las actuaciones identificadas correctamente por los participantes.

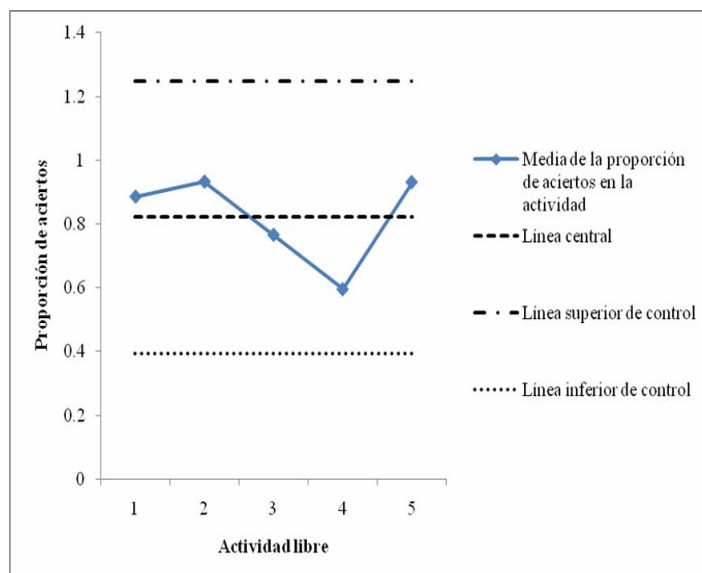


Figura 4. Gráfico de control (gráfico \bar{x}) de las actividades identificadas correctamente por los participantes en la fase de formación en estandarización/puntuación de muestras locales.

Puesto que el proceso representado en la Figura 4 no cumple ninguno de los criterios mencionados anteriormente para determinar cuándo un proceso está estadísticamente descontrolado, se debe concluir que durante la fase de estandarización/puntuación de muestras locales el proceso de formación en reconocimiento de niveles comunes de referencia en actuaciones orales en nuestra muestra de profesores está controlado estadísticamente.

Aunque el tamaño de la muestra en relación a la población de profesores de lengua inglesa en centros bilingües de E.S.O. no es representativa, esta conclusión suscita dos interrogantes:

¿Pueden ser estos resultados generalizados a la población de profesores de lengua inglesa en centros bilingües?

¿La adición de actividades de aprendizaje a la actividad 3 (fase libre) en la estandarización/puntuación de muestras locales puede conducir a un proceso de formación estadísticamente descontrolado?

5. Conclusiones

Entre las conclusiones de este estudio se pueden citar las siguientes:

Los profesores de lengua inglesa de nuestra muestra fueron significativamente capaces de identificar y reordenar los descriptores de los niveles comunes de referencia de actividades lingüísticas comunicativas aisladas y descritas mediante un único criterio. Sin embargo, los participantes fueron significativamente incapaces de reordenar los descriptores de los niveles comunes de referencia de seis actividades lingüísticas comunicativas diferentes cuando estos descriptores fueron presentados simultáneamente.

Existió una tendencia en los participantes a una identificación correcta del nivel común de referencia a medida que aumentaba el número de actividades de aprendizaje realizadas durante la fase de formación en estandarización/puntuación de muestras locales.

Las actividades 2 y 3 de la fase de formación en estandarización/puntuación de muestras locales no fueron pruebas unidimensionales ya que, probablemente, durante la realización de esas actividades cada participante utilizó un concepto de nivel común de referencia diferente al del resto de participantes y este concepto se iba modificando a medida que se realizaban las actividades.

El proceso de identificación de los niveles comunes de referencia fue estadísticamente estable.

Basándose en estas conclusiones, los autores proponen una serie de sugerencias a la administración educativa, los formadores y los diseñadores de programas de formación en reconocimiento de niveles comunes de referencia:

La administración educativa debería incluir para el área de lengua extranjera, en los currículos de los niveles no universitarios, objetivos explícitos consistentes en la adquisición de un nivel común de referencia mínimo y un nivel común de referencia máximo.

La administración educativa debería establecer un procedimiento de certificación de capacidad de reconocimiento de niveles comunes de referencia, con un porcentaje del 90% de identificaciones correctas como puntuación límite mínima para que los participantes obtengan un certificado de capacidad de reconocimiento de niveles comunes de referencia, aunque un porcentaje del 95% debería ser el porcentaje mínimo deseable. Estas puntuaciones límites deberían también ser utilizadas por los diseñadores de programas de formación similares.

Sólo los profesionales con esta certificación podrían tomar parte en procedimientos administrativos en los que se toma una decisión importante basada en la identificación de niveles comunes de referencia.

Referencias bibliográficas

- Alderson, J. C., ed. 2002. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*. Strasbourg: Council of Europe.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. 1999. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Andalucía. 2006. “Orden de 24 de julio de 2006, por la que se regulan determinados aspectos sobre la organización y el funcionamiento de los Centros Bilingües”. *Boletín Oficial de la Junta de Andalucía* 156: 11-15.
- Andalucía. 2008. “Orden de 5 de agosto de 2008, por la que se desarrolla el currículo correspondiente al Bachillerato en Andalucía”. *Boletín Oficial de la Junta de Andalucía* 169: 98-222.
- Bachman, L. F., y Palmer, A. S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Breton, G., North, B., y Lepage, S. 2008. Cross-language Benchmarking Seminar to Calibrate Examples of Spoken Production in English, French, German, Italian and Spanish with regard to the Six Levels of the Common European Framework of Reference for Languages (CEFR). Strasbourg: Council of Europe.
- Broeder, P., y Martyniuk, W. 2008. “Language education in Europe: The Common European Framework of Reference”. *Encyclopedia of Language and Education, Volume 4: Second and Foreign Language Education*. Eds. N. van Deusen-Scholl y N. H. Hornberger. New York: Springer Science y Business Media LLC. 209–226.

- Cambridge ESOL. 2003. *Samples of Oral Production Illustrating, for English, the Levels of the Common European Framework of Reference for Languages*. Cambridge y Strasbourg: Cambridge ESOL y Council of Europe.
- Cochran, W. G. 1937. "The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient". *Journal of the Royal Statistical Society* 100: 69-73.
- Consejo de Europa. 2002. *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación*. Madrid: Anaya.
- Corrigan, M. 2007. *Seminar to Calibrate Examples of Spoken Performance*. Strasbourg: Council of Europe.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Cronbach, L. J. 1951. "Coefficient alpha and the internal structure of tests". *Psychometrika* 16: 297-334.
- España. 2007. "Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria". *Boletín Oficial del Estado* 5: 677-773.
- España. 2008. "Real Decreto 1892/2008, de 14 de noviembre, por el que se regulan las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades públicas españolas". *Boletín Oficial del Estado* 283: 46932-46946.
- Lepage, S., y North, B. 2005. *Guide for the Organisation of a Seminar to Calibrate Examples of Spoken Performance in line with the Scales of the Common European Framework of Reference for Languages*. Strasbourg: Council of Europe.

- Little, D. 2006. "The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact". *Language Learning* 39: 167-190.
- Murphy, K. R., y Davidshofer, C. O. 1994. *Psychological Testing: Principles and Applications*. Englewood Cliffs, New Jersey: Prentice Hall.
- North, B. 2000. *The Development of a Common Framework Scale of Language Proficiency*. New York: Peter Lang.
- North, B., Figueras, N., Takala, S., Van Avermaet, P., y Verhelst, N. 2009. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*. Strasbourg: Council of Europe.
- Nunnally, J. C. 1978. *Psychometric Theory*. New York: McGraw-Hill.
- Stoumbos, Z. G., Reynolds, M. R. Jr., Ryan, T. P., y Woodall, W. H. 2000. "The state of statistical process control as we proceed into the 21st century". *Journal of the American Statistical Association* 95: 992-998.
- Triola, M. 2006. *Elementary Statistics*. Boston: Pearson.
- Wilcoxon, F. 1945. "Individual comparisons by ranking methods". *Biometrics Bulletin* 1: 80-83.
- Woodall, W. H. 2000. "Controversies and contradictions in statistical process control". *Journal of Quality Technology* 32: 341-350.

Primera versión recibida: mayo de 2010.
Versión final aceptada: octubre de 2010.