

CONSTRUCCIÓN Y USO DE TESTS ADAPTATIVOS INFORMATIZADOS PARA LA EVALUACIÓN DEL APRENDIZAJE UNIVERSITARIO

Eduardo García Jiménez¹

Javier Gil Flores

María Teresa Padilla Carmona

Jesús García Vidal

Facultad de Ciencias de la Educación

RESUMEN

El presente trabajo es consecuencia de la preocupación por encontrar nuevas vías para la evaluación y calificación de los aprendizajes de los/as alumnos/as y desde esa perspectiva recoge el proceso de elaboración, aplicación y evaluación de dos Tests Adaptativos Informatizados para medir el rendimiento de alumnos de la Facultad de Ciencias de la Educación, dentro de la Licenciatura de Pedagogía, y en la asignaturas Diagnóstico Pedagógico y Estadística.

Entendemos que los tests adaptativos informatizados (TAIs) pueden ser una alternativa que afiance la idea de que es posible realizar una evaluación formativa apoyada en el conocimiento más preciso de la competencia real que el estudiante tiene en una determinada materia. Por otra parte, la construcción y administración informatizada de tales tests, que supone rentabilizar los avances metodológicos en el campo de la psicometría auspiciados desde la TRI, nos permitiría una mayor precisión en la medida del rendimiento y un conocimiento inmediato por parte del estudiante y profesor/a de los resultados de la evaluación..

ABSTRACT

This work results as a cuestion to find new ways to assess students learning, and in this way gets the process to elaborate, aplicate and assess of two Computerized Adaptive Tests to check the students achievement from Sciences of Education degree, in the subject of Educational Assessment and Educational Statistic.

We think that computerizate adaptative tests (CATs) may be an alternative to improve the students formative evaluation yilding such education in the real students competence in an educative area. In another hand, the computerized construccion and aplication of such tests, and it suposes the use of methodologics advances in the measurement area based from the IRT, could allow us a higher accuracy in the achievements measure and an immediate knowledge from the students and teachers of assessments results.

¹ Profesores del Departamento de Didáctica, Organización Escolar y Métodos de Investigación y Diagnóstico en Educación de la Universidad de Sevilla. También han contribuido a este proceso de innovación las profesoras-colaboradoras María de los Angeles López Romero, Olga Guijarro Cordobés y Mercedes Martín Berrido, y los/as alumnos/as internos/as: Pilar Fernández Cepeda, Carmen Fernández Camacho y Daniel Camargo Flores.

Los profesores que lo deseen pueden solicitarnos los TAls que se mencionan en este trabajo. Basta que utilicen la siguiente dirección de correo electrónico: E-mail: egarji@cica.es. Lo recibirán a través de la dirección de e-mail que nos indiquen.

INTRODUCCIÓN

Con la presentación de este informe sobre la **Construcción y uso de tests adaptativos informatizados para la evaluación del aprendizaje universitario* + concluye oficialmente el desarrollo de una experiencia innovadora en el ámbito de la evaluación. En esta experiencia han participado cuatro profesores que comparten dos asignaturas de la titulación de Pedagogía, Estadística y Diagnóstico Pedagógico, además de tres colaboradoras honorarias, cinco alumnos/as internos/as y, por supuesto, todos los alumnos y alumnas matriculados en dichas asignaturas durante el curso 97-98.

La innovación a la que está referida este informe es consecuencia de la preocupación por encontrar nuevas vías para la evaluación y calificación de los/as aprendizajes de los alumnos/as. En sí, la innovación realizada supone poner en marcha un modo alternativo de evaluación del rendimiento del alumnado utilizando tests adaptativos informatizados en diferentes asignaturas y apoyándose en la coordinación entre el profesorado. Desde esa perspectiva, este informe recoge el proceso de elaboración, aplicación y evaluación de dos Tests Adaptativos Informatizados para medir el rendimiento de estudiante.

Las primeras páginas de este documento están dirigidas, sin embargo, a introducir el concepto de test adaptativo informatizado y sus relaciones con otros elementos relativos a la evaluación y la medición educativas.

1. LOS TESTS ADAPTATIVOS INFORMATIZADOS

El test adaptativo o test a la medida se compone de diferentes conjuntos de ítems que se administran a diferentes estudiantes, según la situación de cada uno en el atributo medido (Renom, 1993:19). Estamos haciendo referencia, por tanto, a tests que aun midiendo el mismo contenido, no son idénticos para todos los estudiantes.

En la administración de un test adaptativo se emplean puntos de inicio (primer ítem) basados en informaciones previas sobre el estudiante y se establece una regla para la selección de los nuevos ítems que se irán presentando de modo que en cada respuesta se vaya concretando el nivel de capacidad del examinado. Ello supone efectuar algún tipo de predicción sobre los ítems aún no respondidos, a fin de seleccionar el que más informe sobre la posición del alumno o alumna. Este ciclo se repite hasta cumplir con cierta condición que da por terminada la presentación de nuevos ítems. Es entonces cuando debe asignarse una puntuación definitiva al alumno o a la alumna (Renom, 1993).

La principal diferencia entre los tests adaptativos y los convencionales radica en que en éstos todos los individuos reciben la misma serie de ítems, mientras que en un test adaptativo «(...) diferentes conjuntos de cuestiones (ítems) se administran a diferentes individuos dependiendo del estatus individual del mismo sobre el rasgo medido» (Weiss, 1990: 454). Este hecho permite superar el dilema de tener que elegir entre la discriminación y la fidelidad (un test convencional puede proporcionar mediciones de gran fidelidad -precisión- en el nivel de dificultad establecido, pero tiene poca capacidad para discriminar a los individuos de otros

niveles de rasgo, o viceversa), seleccionando de un banco de ítems aquellos cuyo nivel de dificultad sea el adecuado para cada estudiante.

Los tests adaptativos pueden suponer, efectivamente, una respuesta a necesidades asociadas al diagnóstico de las dificultades en el aprendizaje del estudiante, la evaluación del rendimiento, la certificación de los aprendizajes, la orientación profesional...etc. Este tipo de tests supone un tratamiento diferenciado del problema de la medida en educación y de la vinculación entre evaluación y enseñanza.

Con todo, probablemente es en el desarrollo de una *evaluación idiosincrática y formativa* en donde los tests adaptativos suponen un mayor avance con respecto a las pruebas clásicas de medida objetiva, a menudo criticadas (con razón) por tener un fuerte carácter sumativo y, con frecuencia, normativo. Respecto al *carácter idiosincrático* de la evaluación, baste recordar que el propio concepto de test adaptativo parte del interés por una evaluación adaptada al individuo, que se interesa por la posición individual de un sujeto dado con respecto a la variable que se está midiendo (Weiss, 1990).

Respecto al segundo aspecto comentado, la posibilidad de vehicular una *evaluación formativa* que permita la toma de decisiones respecto al desarrollo posterior del programa de enseñanza y aprendizaje, pese a ser un aspecto aún inexplorado en buena parte, creemos que la posibilidad de elaborar ítems muy diversos en cuanto a su grado de abstracción y complejidad, en cuanto al formato de presentación de la información, el tipo de ayudas para la resolución, etc., unidos a la capacidad del ordenador para analizar relaciones entre conjuntos de respuestas muy diferentes entre sí, abre un campo de indudable interés para el/la profesor/a, que podrá basar sus juicios y decisiones en un análisis objetivo del efecto diferencial de las distintas variables implicadas en el diseño de los ítems dirigidos a una misma capacidad y/o contenido.

No obstante, construir y administrar un test adaptativo que siga la definición anterior supone una tarea ingente difícilmente abordable sin el uso de un ordenador. Un test adaptativo informatizado (TAI) es un tipo de test, que incorpora el ordenador para administrar tests a la medida o tests adaptativos en los que la elección del siguiente *ítem* a administrar está basada en la respuesta o respuestas previas del sujeto.

Un posible diseño de test adaptativo se inicia con la aplicación de un *ítem* que se espera sea de dificultad media. El segundo *ítem* y los subsiguientes a administrar están determinados por las respuestas del sujeto. Si el primer *ítem* es contestado correctamente, se selecciona y administra otro más difícil, mientras que una respuesta incorrecta conduce a la administración de un *ítem* más fácil. Tras cada respuesta se obtiene una estimación provisional de la capacidad del sujeto o de su nivel de funcionamiento y esta estimación se utiliza para elegir el *ítem* a aplicar a continuación. Este siguiente *ítem* se selecciona de modo tal que proporcione la máxima información acerca del sujeto, cuya capacidad se supone igual a la estimación basada en las respuestas anteriores. Se detiene la administración del test cuando las estimaciones alcanzan un nivel predeterminado de precisión o cuando se ha administrado un número máxi-

mo de *ítems* que puede considerarse como razonable (Linn,1988:156-157). No obstante, la presentación anterior ofrece algunas variaciones dependiendo del procedimiento o estrategia seleccionados en cada caso (p.e. binivel, ramificada, máxima información, etc.).

Los tests adaptativos computerizados o TAIs pueden ser utilizados, con las ventajas añadidas que este tipo de procedimiento supone con respecto a los procedimientos clásicos, en la evaluación de las capacidades que se desarrollan utilizando como soporte los contenidos referidos a hechos y conceptos. Este tipo de contenido es en gran medida evaluable a partir de pruebas objetivas, siendo la condición fundamental un diseño riguroso de los ítems, más la evaluación a través de los TACs puede trasladarse a otros contenidos -de tipo procedimental actitudinal- o, al menos, a algunos de sus componentes básicos.

La aplicación a situaciones particulares de determinados procedimientos, el grado de acierto en la elección de procedimientos para solucionar diferentes tipos de tareas, la generalización del procedimiento a contextos diferentes de los de enseñanza-aprendizaje, el grado de automatización del procedimiento, el grado de conocimiento (declarativo) sobre él, son todos aspectos que pueden ser evaluados a través de los TAIs. Incluso la corrección y precisión de las acciones que componen un procedimiento puede llegar a ser valorado a través de prueba adaptativas.

1.1.LOS TESTS ADAPTATIVOS INFORMATIZADOS Y LA TRI

La construcción y administración adaptativa de tests requiere una cobertura psicométrica diferente a la proporcionada por la teoría clásica de los tests. Esta teoría resulta inadecuada porque hace depender las mediciones de un dominio de los instrumentos utilizados y de los sujetos implicados en la medición. Estas limitaciones son particularmente graves cuando como ocurre con los TAIs- cada sujeto responde a un test diferente y la dificultad de los ítems es independiente de la competencia del grupo de sujetos utilizado. De igual modo, las puntuaciones con números enteros, que son los más corrientemente utilizados en los tests de rendimiento, resultan inadecuadas para un test en el que la dificultad de los ítems administrado varía de un sujeto a otro.

El desarrollo de tests adaptativos requiere un tipo de respuestas psicométricas que sólo puede aportar una teoría de medición como la Teoría de Respuesta al Ítem (TRI). Las expectativas que trata de llenar esta teoría podrían resumirse en:

1. Obtener mediciones que no varíen en función del instrumento utilizado, que sean invariantes respecto a los tests empleados.
2. Disponer de instrumentos de medida cuyas propiedades no dependan de los objetos medidos, sean invariantes respecto de los sujetos evaluados (Muñiz, 1990:16).

Lo que viene a aportar la TRI es, por tanto, la posibilidad de equiparar convenientemente puntuaciones obtenidas con distintos tests que miden el mismo dominio, así como pautas adecuadas para adaptar el test al sujeto.

Un TAI, como ya se ha apuntado, supone decidir qué ítems integran el test y expresar los resultados en una métrica común a otros tests diferentes utilizados con otros alumnos o alumnas para medir el mismo contenido. La TRI constituye el marco psicométrico adecuado para resolver ambas cuestiones, pues al apoyarse en bancos de ítems con parámetros estimados sobre cada elemento (dificultad, discriminación o respuesta al azar) permite elegir los adecuados según el nivel de capacidad del alumno o alumna. Utilizando la función de información se elegirán aquellos ítems que proporcionen más información en relación con el nivel de capacidad del sujeto, además los resultados vendrán expresados en la misma escala (la métrica del banco), independientemente del test o conjunto de ítems utilizado con cada alumno o alumna.

En esta aplicación de la TRI, sin embargo, resulta potencialmente importante el grado en que sea unidimensional el banco calibrado de ítems a partir del cual son seleccionados los ítems. Esto es así porque la evaluación de la capacidad en una escala común, cuando los sujetos han respondido a diferentes ítems, depende del supuesto de la invarianza de las estimaciones de la capacidad a lo largo de los diferentes conjuntos de ítems, lo cual, al menos con los modelos actualmente utilizables, requiere unidimensionalidad (Linn, 1988:158).

2. EL PROCESO DE CONSTRUCCIÓN Y APLICACIÓN DE LOS T.A.I.S

La construcción de un TAI está sujeta al tipo de decisiones que se adopten como respuesta a tres cuestiones básicas:

1. ¿Cómo elegir el ítem que inicia el test?
2. ¿Cómo elegir el siguiente ítem que va a ser administrado después de conocer la respuesta del alumno a un ítem cualquiera?
3. ¿Cómo concluir el test? (Wainer, 1990:14).

La construcción y aplicación de un TAI precisa de una serie de elementos comúnmente aceptados. Renom (1993) resume las aportaciones de diferentes autores sobre el proceso de construcción y aplicación de un TAI en torno a siete bloques de elementos fundamentales: Adecuación del contenido, condiciones de aplicación, calibración del banco de ítems, algoritmo de aplicación de ítems, validez, fiabilidad y control de factores humanos y situacionales.

2.1. EL DISEÑO DE UNA ESTRATEGIA DE TRABAJO EN COMÚN

Nuestro proyecto de innovación educativa pretende responder, desde una misma iniciativa, a las necesidades e intereses asociados a asignaturas y profesores/as diferentes. Partimos de la idea de que un cambio en los recursos y modos de evaluación para tener ciertas posibilidades de éxito y, desde luego, para convertirse en una práctica habitual, no debe presentarse ante la comunidad universitaria como una experiencia individual que todos asocien a determinado/a profesor/a y asignatura, sino que necesita del apoyo decidido de un grupo de docentes que trabajen desde diferentes materias.

Las diferentes fases de este trabajo ha supuesto adoptar una estrategia común de trabajo que se ha traducido en decisiones relativas a la construcción de tablas de especificaciones, formato de los ítems, el diseño de la recogida de datos, el estudio exploratorio de los datos, calibración de los ítems, el desarrollo y gestión informatizada de los bancos o la evaluación de los TAI.

En ese proceso de toma de decisiones se acordó elaborar diferentes bancos de ítems para las asignaturas de Estadística y Diagnóstico Pedagógico. Nuestra intención era que cada uno de estos bancos abordase un dominio educativo presidido por cierta unidad conceptual y temporal, de modo que se asegurase cierta unidimensionalidad y que la aplicación de los TAI al alumnado siguiese un proceso paralelo al desarrollo del curso. Así, se diseñaron dos bancos de ítems para la asignatura Diagnóstico Pedagógico, correspondientes a otros tantos bloques de contenidos y exámenes parciales, y dos bancos para la asignatura Estadística que estaban asociados a otros dos bloques y exámenes parciales. No obstante, con la presentación de este informe se hace referencia únicamente a dos bancos de ítems, uno por cada una de las asignaturas antes mencionadas.

2.2. REDACCIÓN Y ADMINISTRACIÓN DE ÍTEMS

La primera de las tareas abordadas en la construcción de los bancos ha sido la definición del dominio educativo correspondiente a la materia objeto de evaluación. Para ello, se hizo un repartido de tablas de especificaciones que tuviesen en cuenta las unidades de contenido más significadas en cada caso, así como el tipo de capacidad que debiera exhibir el estudiante tras el proceso de aprendizaje. En este sentido, intentamos construir ítems que midiesen en esencia la capacidad del alumno/a para comprender el significado de ciertos conceptos (su naturaleza, propiedades y relaciones con otros conceptos) y procedimientos (finalidad, posibilidad, limitaciones, etc.). De igual modo, en cada una de las tablas de especificaciones se fijaron tres niveles de dificultad (baja, media y alta) con objeto de que los contenidos a abordar estuviesen representados en el banco mediante ítems con diferentes niveles de dificultad. Pretendíamos, sencillamente, que al contestar a un TAI cualquiera el alumno encontrase ítems referidos a distintas unidades de contenido que se ajustasen a su dominio de la capacidad medida.

La redacción de los ítems fue la segunda de las tareas acometidas en este proceso. Decidimos que todos los ítems tuviesen un mismo formato en todos los bancos y que respondiesen a una misma presentación en las pruebas de lápiz y papel. Los ítems serían dicotómicos y de elección de respuesta, concretamente de opción múltiple con cuatro alternativas. Se siguieron unas reglas comunes a la hora de redactarlos y se decidió que en la corrección se tendrían en cuenta los errores. De acuerdo con tales decisiones se han redactado más de trescientos ítems en cada una de las asignaturas.

La administración de los ítems a partir de pruebas de lápiz y papel se hizo coincidir temporalmente con los exámenes parciales de cada una de las asignaturas, de modo que esa administración

nistración es, o bien el propio examen, o bien un Aejercicio de ensayo@ para los/as alumnos/as que se realiza uno o dos días antes del examen real. Dado que cada uno de los bancos cuenta con un número de items tal que no puede administrarse en una sola sesión, dividimos el banco en dos partes que comparten diez items-ancla. En la asignatura de Diagnóstico se han administrado dos pruebas de 80 y 60 items, y en la de Estadística una prueba de 85 y otra de 40 items cada una. En el primer caso, las dos pruebas corresponden a exámenes reales, mientras que en el segundo caso - Estadística- la primera prueba era un ensayo y la segunda el examen real.

2.3. ANÁLISIS EXPLORATORIO DE LOS DATOS

A partir de los resultados obtenidos tras la administración de los items, se ha procedido a la calibración del banco. Esta tarea conlleva el estudio de las características psicométricas de los elementos, que en el caso del enfoque de la TRI asumido en este estudio, supone estimar los parámetros de acuerdo con un determinado modelo logístico de respuesta al ítem. No obstante, antes de proceder a la estimación de parámetros, hemos realizado una primera selección de items, basándonos en los índices clásicos de dificultad y discriminación, y en el comportamiento de los distractores. Pretendemos con ello reducir el conjunto inicial de items, excluyendo los que presentan un funcionamiento inapropiado, y favorecer de este modo una mayor calidad psicométrica de las pruebas. La incidencia de items defectuosos es especialmente importante en los tests adaptativos dado que éstos suelen tener un tamaño relativamente pequeño. De ahí que en la construcción del banco se suelen dedicar considerables esfuerzos para asegurar la alta calidad de los items del test (Hambleton, Zaal y Pieters, 1991).

La selección de items se ha basado en los criterios de la teoría clásica, cuya aplicación conduce a una mejora en el ajuste a los modelos de la TRI, como hemos podido comprobar en trabajos anteriores (Gil, García y Rodríguez, 1995). Concretamente, hemos excluido del conjunto inicial aquellos reactivos que presentan un índice de dificultad extremo, es decir, una proporción de aciertos superior a 0.90 ó inferior a 0.10; un insuficiente poder de discriminación, que se traduciría en un valor para la correlación ítem-total, medida por el coeficiente de correlación biserial puntual, inferior a 0.30; y una discordancia entre la respuesta especificada como correcta y la que, desde un punto de vista psicométrico, funciona como tal, situación que se daría cuando la mayor correlación con la puntuación total corresponde a un distractor. Al considerar la discriminación del ítem, se han conservado algunos en los que ésta descendía hasta el valor 0.28, siempre y cuando las correlaciones con todos los distractores fueran negativas.

En el caso de las pruebas de la asignatura Estadística, los filtros basados en la dificultad y capacidad de discriminación de los items han llevado a descartar 22. Los correspondientes índices aparecen en las tablas 1 y 2, donde además se ha marcado aquellos items que no cumplen los criterios considerados. A estos habría que añadir 3 items más que se descartaron por mal funcionamiento de la respuesta especificada como correcta, restando un total de 90 elementos en el banco.

Tabla 1. Indicadores de dificultad y discriminación para los items de la primera prueba de Estadística

Item	p	r_{ho}	Item	p	r_{ho}	Item	p	r_{ho}
1 *	.37	.24	30	.86	.31	58	.66	.36
2	.58	.40	31	.75	.35	59	.59	.38
3	.87	.38	32	.74	.33	60	.47	.40
4 *	.72	.20	33	.59	.39	61	.78	.49
5	.61	.47	34	.52	.45	62	.34	.46
6	.28	.29	35	.68	.34	63	.81	.43
7	.46	.41	36	.77	.31	64	.90	.37
8 *	.77	.28	37	.60	.41	65	.64	.40
9	.45	.40	38	.49	.40	66	.69	.43
10	.39	.33	39*	.24	.22	67	.46	.40
11	.87	.31	40*	.85	.22	68	.48	.40
12	.33	.32	41	.49	.38	69	.42	.43
13	.65	.37	42	.47	.36	70 *	.46	.15
14 *	.34	.04	43	.56	.33	71	.44	.38
15 *	.39	.03	44	.78	.35	72	.35	.38
16	.70	.30	45	.89	.39	73	.44	.40
17	.58	.36	46	.65	.31	74	.71	.38
18	.50	.34	47	.42	.41	75 *	.55	.26
19 *	.71	.16	48	.42	.43	76	.80	.34
20 *	.39	.21	49*	.16	.20	77	.36	.34
21	.74	.37	50*	.50	.19	78 *	.48	.21
22	.58	.34	51	.40	.38	79	.61	.42
23 *	.14	.14	52	.72	.28	80	.45	.31
24	.65	.41	53	.49	.30	81	.54	.49
25 *	.17	.26	54	.46	.35	82	.19	.28
26	.45	.36	55	.54	.38	83 *	.46	.20
27 *	.17	.15	56	.37	.43	84 *	.44	.22
28	.72	.46	57 *	.72	.25	85	.55	.43
29 *	.12	.11						

Tabla 2. Indicadores de dificultad y discriminación para los ítems de la segunda prueba de Estadística.

Item	p	r_{bp}	Item	p	r_{bp}	Item	p	r_{bp}
1	.46	.48	15	.52	.52	28	.82	.30
2	.73	.44	16	.33	.48	29	.40	.30
3	.58	.33	17 *	.37	.25	30	.53	.39
4	.83	.35	18	.58	.53	31	.80	.30
5	.91	.31	19	.49	.47	32	.57	.45
6 *	.77	.18	20	.38	.35	33	.59	.42
7	.73	.45	21	.86	.28	34	.56	.41
8 *	.87	.26	22	.46	.35	35	.60	.28
9 *	.63	.13	23	.41	.52	36	.81	.34
10	.49	.33	24	.64	.29	37	.53	.52
11	.59	.40	25	.39	.28	38	.81	.28
12	.65	.43	26 *	.57	.26	39	.31	.47
13	.67	.28	27	.63	.36	40	.31	.42
14 *	.32	.26						

En el caso de las pruebas de la asignatura Diagnóstico, el análisis de la dificultad y de la capacidad de discriminación de los ítems ha eliminado 48 ítems. Los índices relativos a dichos indicadores aparecen en las tablas 3 y 4, en las que se ha marcado aquellos ítems que no cumplen los criterios considerados. Con estos ítems descartamos 2 más por mal funcionamiento de la clave de respuesta especificada en cada uno de ellos. De este modo quedaron en el banco un total de 85 ítems.

Considerando este conjunto de reactivos hemos examinado su distribución, en cada área de contenido, a lo largo de los diferentes niveles de dificultad definidos por el porcentaje de acierto (ver tablas 5 y 6). La presencia de ítems de todas las áreas temáticas en la zona central de la tabla es una primera conclusión a la que podemos llegar. A pesar de que en todos los temas, contamos con ítems de mayor o menor grado de dificultad, determinadas áreas cuentan con una mayor presencia de ítems difíciles, concretamente las referidas a *Correlaciones y regresión+ (en Estadística) y a *Reglamento de los I.E.S.+ y *Diversificación curricular+ (en Diagnóstico Pedagógico), a diferencia de lo que ocurre respecto a temas tales como las Escalas de medida+, *Distribuciones de frecuencias+ o *Medidas de tendencia central+ (en Estadística) y a la *LOGSE+ y a la *LOPEGCE+, en las que los ítems de menor dificultad predominan claramente.

Tabla 3. Indicadores de dificultad y discriminación para los ítems de la primera prueba de la asignatura Diagnóstico Pedagógico.

Ítem	p	r_{bo}	Ítem	p	r_{bo}	Ítem	p	r_{bo}
1	.83	.33	28 *	.19	.03	55	.58	.39
2	.42	.37	29	.45	.34	56	.56	.53
3 *	.29	-.34	30	.71	.57	57	.47	.44
4 *	.17	-.18	31	.39	.55	58 *	.14	-.09
5	.82	.36	32 *	.40	.10	59 *	.20	-.08
6	.51	.33	33 *	.24	.23	60	.57	.64
7 *	.91	.56	34 *	.42	.27	61	.51	.30
8 *	.91	.59	35 *	.04	-.09	62	.59	.31
9 *	.93	.27	36	.54	.31	63	.63	.54
10	.83	.48	37 *	.50	.24	64	.80	.43
11	.52	.09	38	.72	.48	65	.56	.55
12 *	.36	-.11	39	.79	.39	66	.62	.58
13	.61	.43	40	.56	.43	67	.67	.55
14 *	.77	.27	41 *	.36	.08	68	.72	.52
15	.41	.49	42	.67	.44	69	.79	.51
16 *	.92	.60	43	.67	.29	70	.72	.45
17 *	.92	.61	44 *	.63	.24	71	.61	.49
18 *	.93	.32	45	.60	.52	72	.37	.39
19 *	.71	.23	46	.73	.42	73 *	.80	.26
20	.76	.42	47 *	.44	.11	74	.62	.56
21	.61	.30	48	.52	.45	75	.73	.28
22	.84	.41	49	.71	.32	76	.40	.37
23	.61	.37	50	.71	.51	77 *	.63	-.01
24	.78	.46	51	.67	.50	78	.51	.55
25	.27	.34	52 *	.24	.26	79 *	.41	.18
26	.73	.55	53 *	.57	.19	80	.52	.51
27	.77	.43	54	.76	.48			

Tabla 4. Indicadores de dificultad y discriminación para los ítems de la segunda prueba de la asignatura Diagnóstico Pedagógico.

Item	p	r_{bo}	Item	p	r_{bo}	Item	p	r_{bo}
1	.34	.47	21	.49	.29	41	.81	.46
2	.56	.31	22	.25	.29	42	.80	.48
3	.64	.40	23	.73	.31	43	.65	.33
4 *	.13	.20	24	.66	.39	44	.54	.39
5	.56	.55	25	.85	.54	45	.56	.40
6 *	.03	.04	26	.40	.46	46	.49	.55
7	.20	.42	27	.44	.40	47 *	.92	.60
8	.36	.36	28	.71	.54	48	.39	.40
9 *	.28	.18	29	.40	.36	49 *	.63	.17
10	.72	.47	30 *	.49	.05	50	.58	.32
11	.53	.52	31	.62	.41	51 *	.43	.27
12	.36	.38	32	.96	.82	52 *	.50	-.12
13 *	.93	.24	33	.75	.54	53 *	.24	.18
14 *	.09	.30	34 *	.97	.85	54 *	.39	.11
15	.40	.41	35 *	.91	.36	55	.58	.48
16 *	.24	.14	36	.66	.32	56	.73	.50
17 *	.21	.14	37	.23	.31	57	.58	.47
18	.40	.50	38 *	.66	.19	58	.51	.39
19 *	.86	.16	39 *	.23	.07	59 *	.30	.16
20	.44	.37	40	.66	.43	60	.36	.34

Tabla 5. Distribución de items según dificultad y áreas de contenido, una vez filtrados a partir de 1 criterios clásicos en la asignatura Estadística.

CONTENIDOS	0.10- 0.20	0.21- 0.40	0.41- 0.60	0.61- 0.80	0.81- 0.90
Escalas de medida			2	2	2
Distribución de frecuencias			1	2	1
Representaciones gráficas			2	4	1
Medidas de tendencia central		1	3	3	2
Medidas de dispersión			3	1	2
Medidas de posición			3	2	2
Medidas de forma			4	3	
Puntuaciones individuales			3	1	1
Curva normal			2	3	
Correlación de Pearson		3	4	4	
Regresión simple	1	4	5	1	
Otros coeficientes de correlación		4	8		
Total	1	12	40	26	11

Tabla 6. Distribución de items según dificultad y áreas de contenido, una vez filtrados a partir de 10 criterios clásicos en la asignatura Diagnóstico Pedagógico (Legislación Educativa).

CONTENIDOS	.10 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .90
LOGSE			4	3	3
LODE		1	1	1	
LOPEGCE		1	1	6	2
Reglamento I.E.S.		2	2	4	
Derechos y deberes alumnos	1		1		
.....
Curriculum Educación Infantil				2	1
Curriculum Educación Primaria			1		
Curriculum ESO		1	2		
.....
R.D. Educación Especial		1	5	2	
Adaptaciones curriculares		1	3	7	
Programas de garantía social			1	3	1
Diversificación Curricular		4	2	6	
.....
Equipos de Apoyo Externo		2	5	1	
Departamentos de Orientación		1	4	1	
Total	1	14	32	36	7

2.4. CALIBRACIÓN Y EQUIPARACIÓN DE LOS ÍTEMS

Con los 90 y 85 ítems seleccionados, respectivamente, en los bancos de Estadística y Diagnóstico Pedagógico se llevó a cabo una estimación de parámetros, adoptando un modelo logístico de 3 parámetros, y estimando el valor de éstos según el método de máxima verosimilitud marginal, debido a Bock y Lieberman (1970), aplicado mediante el programa XCALIBRE (Assessment Systems Corporation, 1995). Los resultados de esta operación aparecen recogidos en las tablas 7 y 8 (Estadística), 9 y 10 (Diagnóstico Pedagógico). Como puede observarse, ningún ítem presenta parámetros de dificultad extremos (fuera del intervalo $[-2.95, 2.95]$), y los niveles de discriminación resultaron buenos, encontrándose todos los parámetros a por encima del valor 0.48. Asimismo, encontramos valores aceptables en el parámetro de azar c , que no se alejaron mucho del 0.25 esperado. El estudio de la bondad de ajuste de los datos a la Curva Característica del Ítem para cada elemento, basado en el análisis de los residuos estandarizados, no detectó ningún caso de desajuste significativo, situándose los valores obtenidos dentro de la banda $[-2, +2]$ que, asumiendo la distribución normal de los residuos, se consideran índices de ajuste aceptables (Muñiz, 1990).

Tabla 7. Parámetros a , b y c estimados para los ítems de la primera prueba (Estadística).

Ítem	a	b	c	Ítem	a	b	c
2	0.6722	0.1340	0.2462	46	0.5996	-0.1714	0.2502
3	0.8497	-1.5450	0.2423	47	0.8457	0.9009	0.2304
5	0.8533	-0.0602	0.2318	48	0.8445	0.8388	0.2416
6	0.7776	1.8559	0.2215	51	0.7091	1.0562	0.2416
7	0.6903	0.6329	0.2496	52	0.5400	-0.6710	0.2542
9	0.7269	0.7486	0.2422	53	0.6124	0.6805	0.2582
10	0.6642	1.1682	0.2366	54	0.6857	0.9291	0.2530
11	0.6893	-1.7544	0.2461	55	0.7076	0.4745	0.2438
12	0.7869	1.5056	0.2279	56	0.7849	1.1396	0.2189
13	0.6807	-0.1392	0.2467	58	0.6112	-0.3224	0.2486
16	0.5821	-0.5314	0.2537	59	0.6855	0.1147	0.2534
17	0.6807	0.1954	0.2413	60	0.7765	0.7227	0.2427
18	0.6509	0.5871	0.2515	61	0.8331	-1.0113	0.2448
21	0.6936	-0.7587	0.2427	62	0.8748	1.1451	0.2106
22	0.6309	0.2712	0.2507	63	0.7756	-1.1617	0.2432
24	0.7314	-0.1776	0.2468	64	0.7973	-1.9479	0.2452
26	0.7218	0.8649	0.2427	65	0.7197	-0.2371	0.2459
28	0.7584	-0.6669	0.2412	66	0.7792	-0.4493	0.2384
30	0.6695	-1.6311	0.2480	67	0.7412	0.6896	0.2337
31	0.6394	-0.8419	0.2455	68	0.7612	0.6409	0.2424
32	0.6059	-0.8373	0.2439	69	0.8475	0.9198	0.2254
33	0.7035	0.1230	0.2431	71	0.7258	0.9225	0.2266
34	0.7661	0.4005	0.2458	72	0.8313	1.2588	0.2314
35	0.6212	-0.4317	0.2524	73	0.7979	0.8899	0.2332
36	0.6066	-1.0153	0.2530	74	-0.7097	-0.5308	0.2478
37	0.6896	0.0547	0.2466	76	0.6462	-1.1866	0.2463
38	0.8312	0.6364	0.2410	77	0.7437	1.3302	0.2433
41	0.7778	0.6861	0.2431	79	0.8352	0.0212	0.2437
42	0.7040	0.8501	0.2424	80	0.6331	0.9942	0.2530
43	0.6253	0.3906	0.2522	81	0.7909	0.2290	0.2331
44	0.6868	-1.0502	0.2420	82	0.7827	2.5385	0.2026
45	0.8673	-1.6349	0.2438	85	0.7368	0.2224	0.2389

Tabla 8. Parámetros a , b y c estimados para los items de la segunda prueba (Estadística).

Item	a	b	c	Item	a	b	c
1	0.8389	0.6336	0.2347	24	0.5380	-0.1332	0.2602
2	0.7108	-0.8172	0.2535	25	0.6222	1.5317	0.2678
3	0.5627	0.3160	0.2633	27	0.6349	-0.0399	0.2567
4	0.7172	-1.3532	0.2533	28	0.5911	-1.4626	0.2538
5	0.7523	-2.0449	0.2533	29	0.6314	1.4667	0.2502
7	0.8315	-0.6869	0.2523	30	0.6450	0.4728	0.2548
11	0.5921	-0.0066	0.2563	31	0.5451	-1.3517	0.2577
12	0.6760	-0.2840	0.2563	32	0.7491	0.0969	0.2456
13	0.5253	-0.4494	0.2637	33	0.6856	0.1186	0.2509
15	0.8326	0.2142	0.2420	34	0.6106	0.2007	0.2597
16	0.8579	1.2263	0.2105	35	0.5323	0.2138	0.2653
18	0.6990	-0.1337	0.2485	36	0.6043	-1.3936	0.2567
19	0.7188	0.3545	0.2491	37	0.7974	0.2178	0.2337
20	0.6014	1.1157	0.2546	38	0.5402	-1.4741	0.2569
21	0.5650	-1.9798	0.2555	39	0.8258	1.2790	0.2245
22	0.6266	0.8700	0.2562	40	0.7413	1.3839	0.2514
23	0.8238	0.7359	0.2363				

Tabla 9. Parámetros a , b y c estimados para los items de la primera prueba (Diagnóstico Pedagógico)

Items	a	b	c	Items	a	b	c
1	0.5210	-1.5087	0.2522	48	0.7388	0.5432	0.2372
2	0.6836	1.2871	0.2403	49	0.5016	-0.5447	0.2547
4	0.5932	0.5864	0.2414	50	0.6863	-0.4922	0.2457
5	0.5554	-1.3684	0.2530	51	0.6817	-0.2735	0.2432
6	0.5949	0.8985	0.2583	54	0.6490	-0.7674	0.2491
10	0.6471	-1.3536	0.2501	55	0.5711	0.3141	0.2496
13	0.5700	0.1489	0.2472	56	0.6728	0.2976	0.2338
15	0.7545	1.1545	0.2280	57	0.7177	0.8719	0.2378
20	0.5793	-0.8401	0.2522	60	0.9008	0.2535	0.2359
21	0.5385	0.2120	0.2603	61	0.5661	0.8723	0.2557
22	0.5659	-1.5337	0.2534	62	0.5410	0.3724	0.2555
23	0.5597	0.1795	0.2551	63	0.7121	-0.0385	0.2396
24	0.5806	-1.0101	0.2502	64	0.5628	-1.2230	0.2535
25	0.7347	2.2328	0.2184	65	0.7634	0.2666	0.2310
26	0.7050	-0.5944	0.2454	66	0.8139	-0.0099	0.2378
27	0.5831	-0.9466	0.2513	67	0.7468	-0.2501	0.2446
29	0.6447	1.2382	0.2519	68	0.7039	-0.5616	0.2430
30	0.7554	-0.4720	0.2442	69	0.6537	-1.0291	0.2482
31	0.8521	1.0757	0.2137	70	0.6201	-0.5797	0.2467
36	0.5296	0.7521	0.2581	71	0.6366	0.0638	0.2420
38	0.6387	-0.6125	0.2474	72	0.6850	1.5029	0.2300
39	0.5448	-1.1522	0.2527	74	0.7874	0.0441	0.2410
40	0.6113	0.4794	0.2506	75	0.5341	-0.6589	0.2518
42	0.6527	-0.2169	0.2485	76	0.6367	1.4676	0.2389
43	0.4795	-0.1918	0.2566	78	0.7841	0.5051	0.2259
45	0.6974	0.0855	0.2382	80	0.7662	0.5655	0.2374
46	0.5726	-0.6923	0.2469				

Tabla 10. Parámetros a , b y c estimados para los items de la segunda prueba (Diagnóstico Pedagógico)

Items	a	b	c	Items	a	b	c
1	0.6186	1.5741	0.2450	31	0.5467	0.0352	0.2567
2	0.5018	0.4464	0.2570	33	0.6459	-0.7368	0.2532
3	0.5286	0.0078	0.2604	36	0.5604	-0.1917	0.2556
5	0.5752	0.2604	0.2588	37	0.6616	2.3876	0.2347
7	0.6717	2.4242	0.2269	40	0.5509	-0.2146	0.2565
8	0.5942	1.6900	0.2482	41	0.5666	-1.2412	0.2581
10	0.6003	-0.5864	0.2562	42	0.6000	-1.1882	0.2587
11	0.5926	0.4290	0.2517	43	0.4849	-0.2977	0.2629
12	0.5857	1.4295	0.2527	44	0.5500	0.6366	0.2571
15	0.5921	1.2731	0.2522	45	0.5563	0.3034	0.2554
18	0.6129	1.2368	0.2472	46	0.5633	0.6575	0.2539
20	0.6332	1.0664	0.2535	48	0.5437	1.4688	0.2620
21	0.5275	0.8278	0.2635	50	0.5369	0.2294	0.2632
22	0.6250	2.3072	0.2482	55	0.5993	0.2229	0.2550
23	0.5038	-0.7262	0.2616	56	0.5629	-0.6730	0.2569
24	0.5505	-0.2039	0.2601	57	0.5316	0.2893	0.2639
25	0.6154	-1.6772	0.2582	58	0.5724	0.6124	0.2576
26	0.5429	1.0365	0.2615	60	0.5699	1.6245	0.2583
28	0.6383	-0.5281	0.2521				

Los parámetros resultantes de esta estimación fueron trasladados a una métrica común, siguiendo el procedimiento de equiparación de parámetros basado en la media y la desviación típica, observándose en Estadística un valor medio para el parámetro de dificultad situado en el valor -0.2352 (ver tablas 11 y 12), y en Diagnóstico Pedagógico un valor medio para el parámetro de dificultad de 0.1776 (ver tablas 13 y 14).

El examen de la distribución que los items del banco presentan a lo largo de los diferentes valores de la capacidad medida (ver tablas 15 y 16), indica efectivamente que en la zona correspondiente a valores centrales o bajos del rasgo medido existe una mayor presencia de reactivos. En consecuencia, la función de información del banco alcanzará su valor máximo en esa zona, indicando que las pruebas que podamos construir a partir del banco permitirán una mayor exactitud en la estimación de capacidades para los sujetos que se sitúan en los niveles de capacidad aludidos.

Tabla 11. Parámetros *a*, *b* y *c* tras el proceso de equiparación (Estadística).

Item	a	b	c	Item	a	b	c
1	0.6452	-0.3222	0.2462	46	0.8397	0.7311	0.2106
2	0.8156	-2.0713	0.2423	47	0.7445	-1.6720	0.2432
3	0.8191	-0.5245	0.2318	48	0.7653	-2.4911	0.2452
4	0.7464	1.4716	0.2215	49	0.6908	-0.7088	0.2459
5	0.6626	0.1975	0.2496	50	0.7480	-0.9299	0.2384
6	0.6978	0.3181	0.2422	51	0.7115	0.2566	0.2337
7	0.6376	0.7552	0.2366	52	0.7307	0.2059	0.2424
8	0.6617	-2.2895	0.2461	53	0.8135	0.4964	0.2254
9	0.7553	1.1067	0.2279	54	0.6967	0.4992	0.2266
10	0.6534	-0.6068	0.2467	55	0.7980	0.8496	0.2314
11	0.5588	-1.0154	0.2537	56	0.7659	0.4653	0.2332
12	0.6534	-0.2582	0.2413	57	0.6812	-1.0148	0.2478
13	0.6248	0.1498	0.2515	58	0.6203	-1.6980	0.2463
14	0.6658	-1.2522	0.2427	59	0.7139	0.9240	0.2433
15	0.6056	-0.1793	0.2507	60	0.8017	-0.4397	0.2437
16	0.7021	-0.6468	0.2468	61	0.6077	0.5739	0.2530
17	0.6929	0.4392	0.2427	62	0.7592	-0.2232	0.2331
18	0.7280	-1.1566	0.2412	63	0.7513	2.1828	0.2026
19	0.6427	-2.1610	0.2480	64	0.7073	-0.2301	0.2385
20	0.6138	-1.3389	0.2455	65	0.8389	0.6336	0.2347
21	0.5816	-1.3341	0.2439	66	0.7108	-0.8172	0.2535
22	0.6753	-0.3337	0.2431	67	0.5627	0.3160	0.2633
23	0.7354	-0.0446	0.2458	68	0.7172	-1.3532	0.2533
24	0.5963	-0.9115	0.2524	69	0.6760	-0.2840	0.2563
25	0.5823	-1.5195	0.2530	70	0.5253	-0.4494	0.2637
26	0.6619	-0.4048	0.2466	71	0.8326	0.2142	0.2420
27	0.7979	0.2012	0.2410	72	0.5650	-1.9798	0.2555
28	0.7466	0.2530	0.2431	73	0.6266	0.8700	0.2562
29	0.6758	0.4238	0.2424	74	0.8238	0.7359	0.2363
30	0.6002	-0.0549	0.2522	75	0.5380	-0.1332	0.2602
31	0.6593	-1.5559	0.2420	76	0.6222	1.5317	0.2678
32	0.8325	-2.1650	0.2438	77	0.6349	-0.0399	0.2567
33	0.5756	-0.6404	0.2502	78	0.5911	-1.4626	0.2538
34	0.8118	0.4767	0.2304	79	0.6314	1.4667	0.2502
35	0.8106	0.4120	0.2416	80	0.6450	0.4728	0.2548
36	0.6807	0.6385	0.2416	81	0.5451	-1.3517	0.2577
37	0.5183	-1.1608	0.2542	82	0.7491	0.0969	0.2456
38	0.5878	0.2471	0.2582	83	0.6856	0.1186	0.2509
39	0.6582	0.5061	0.2530	84	0.6106	0.2007	0.2597
40	0.6792	0.0325	0.2438	85	0.5323	0.2138	0.2653
41	0.7534	0.7254	0.2189	86	0.6043	-1.3936	0.2567
42	0.5867	-0.7977	0.2486	87	0.7974	0.2178	0.2337
43	0.6580	-0.3423	0.2534	88	0.5402	-1.4741	0.2569
44	0.7454	0.2911	0.2427	89	0.8258	1.2790	0.2245
45	0.7997	-1.5153	0.2448	90	0.7413	1.3839	0.2514

Tabla 12. Medias y desviaciones típicas para los parámetros en el banco (Estadística).

Parámetro	\bar{x}	S_x
a	0.6857	0.0864
b	-0.2352	0.9851
c	0.2448	0.0113

Tabla 13. Parámetros *a*, *b* y *c* tras el proceso de equiparación (Diagnóstico Pedagógico).

Items	a	b	c	Items	a	b	c
1	1.4493	0.0900	0.2450	44	0.5700	0.1489	0.2472
2	1.1757	-0.3913	0.2570	45	0.7545	1.1545	0.2280
3	1.2385	-0.5785	0.2604	46	0.5793	-0.8401	0.2522
4	1.3477	-0.4707	0.2588	47	0.5385	0.2120	0.2603
5	1.5738	0.4529	0.2269	48	0.5659	-1.5337	0.2534
6	1.3922	0.1395	0.2482	49	0.5597	0.1795	0.2551
7	1.4065	-0.8321	0.2562	50	0.5806	-1.0101	0.2502
8	1.3884	-0.3987	0.2517	51	0.7347	2.2328	0.2184
9	1.3723	0.0283	0.2527	52	0.7050	-0.5944	0.2454
10	1.3873	-0.0384	0.2522	53	0.5831	-0.9466	0.2513
11	1.4360	-0.0539	0.2472	54	0.6447	1.2382	0.2519
12	1.4836	-0.1266	0.2535	55	0.7554	-0.4720	0.2442
13	1.2359	-0.2285	0.2635	56	0.8521	1.0757	0.2137
14	1.4643	0.4029	0.2482	57	0.5296	0.7521	0.2581
15	1.1804	-0.8918	0.2616	58	0.6387	-0.6125	0.2474
16	1.2898	-0.6688	0.2601	59	0.5448	-1.1522	0.2527
17	1.4418	-1.2977	0.2582	60	0.6113	0.4794	0.2506
18	1.2720	-0.1394	0.2615	61	0.4795	-0.1918	0.2566
19	1.4955	-0.8072	0.2521	62	0.6974	0.0855	0.2382
20	1.2809	-0.5668	0.2567	63	0.5726	-0.6923	0.2469
21	1.5133	-0.8963	0.2532	64	0.7388	0.5432	0.2372
22	1.3130	-0.6636	0.2556	65	0.5016	-0.5447	0.2547
23	1.5501	0.4373	0.2347	66	0.6863	-0.4922	0.2457
24	1.2907	-0.6734	0.2565	67	0.6817	-0.2735	0.2432
25	1.3275	-1.1116	0.2581	68	0.6490	-0.7674	0.2491
26	1.4058	-1.0889	0.2587	69	0.6728	0.2976	0.2338
27	1.1361	-0.7089	0.2629	70	0.7177	0.8719	0.2378
28	1.2886	-0.3101	0.2571	71	0.9008	0.2535	0.2359
29	1.3034	-0.4523	0.2554	72	0.5661	0.8723	0.2557
30	1.3198	-0.3012	0.2539	73	0.5410	0.3724	0.2555
31	1.2739	0.0451	0.2620	74	0.7121	-0.0385	0.2396
32	1.2579	-0.4839	0.2632	75	0.5628	-1.2230	0.2535
33	1.4041	-0.4867	0.2550	76	0.7634	0.2666	0.2310
34	1.3188	-0.8690	0.2569	77	0.8139	-0.0099	0.2378
35	1.2455	-0.4583	0.2639	78	0.7468	-0.2501	0.2446
36	1.3411	-0.3204	0.2576	79	0.7039	-0.5616	0.2430
37	1.3352	0.1116	0.2583	80	0.6201	-0.5797	0.2467
38	0.5210	-1.5087	0.2522	81	0.6850	1.5029	0.2300
39	0.6836	1.2871	0.2403	82	0.7874	0.0441	0.2410
40	0.5932	0.5864	0.2414	83	0.5341	-0.6589	0.2518
41	0.5554	-1.3684	0.2530	84	0.6367	1.4676	0.2389
42	0.5949	0.8985	0.2583	85	0.7841	0.5051	0.2259
43	0.6471	-1.3536	0.2501	86	0.7662	0.5655	0.2374

Tabla 14. Medias y desviaciones típicas para los parámetros en el banco (Diagnóstico Pedagógico).

Parámetro	\bar{x}	S_x
a	0.9512	0.3604
b	-0.1673	0.7454
c	0.2491	0.0105

Tabla 15. Distribución de los ítems a lo largo de la capacidad medida (Estadística).

Capacidad	De -2.5 a -1.5	De -1.5 a -0.5	De -0.5 a 0.5	De 0.5 a 1.5	De 1.5 a 2.5
N1 ítems	11	21	40	16	2

Tabla 16. Distribución de los ítems a lo largo de la capacidad medida (Diagnóstico Pedagógico).

Capacidad	De -2.5 a -1.5	De -1.5 a -0.5	De -0.5 a 0.5	De 0.5 a 1.5	De 1.5 a 2.5
N1 ítems	3	21	33	21	7

2.5. DISEÑO INFORMATIZADO DE LOS T.A.I.S

Una vez que completamos la administración y calibración de los ítems procedimos a preparar su presentación en pantalla y a determinar los algoritmos de administración de ítems (nivel inicial de capacidad, criterio para escoger el ítem siguiente, asignación de puntuación a la respuesta, finalización de la sesión). En el caso de los dos bancos de ítems se siguió el mismo procedimiento.

La construcción de los dos tests, que denominamos LEX (en el caso del elaborado para la asignatura Diagnóstico Pedagógico) y ESTADES (para la asignatura Estadística), se realizó tomando como referencia las exigencias de formato establecidas desde el programa MicroCAT. Así, en archivos de formato texto se fueron introduciendo cada uno de los ítems, según un esquema en el que se contemplaron un identificador del ítem, el modelo de respuesta (v.g. tres parámetros: discriminación, dificultad y azar), número de alternativas de respuesta (v.g. cuatro), respuesta correcta, valores correspondientes a los tres parámetros estimados, modo de presentación en pantalla y texto del ítem, incluyendo el enunciado y opciones de respuesta.

Para construir y aplicar LEX y ESTADES, según un formato adaptativo computerizado, el programa MicroCAT dispone del lenguaje MCATL que permite la escritura de una diversidad de métodos para ejecutarlos. En estos tests la selección de los ítems está diseñada según un procedimiento de máxima información, de modo que se parte del ítem de dificultad media en cada banco, y desde ese ítem se selecciona y presenta un nuevo ítem para el cual la función de información alcanza su valor máximo precisamente en el valor de la capacidad estimada en ese momento para el alumno. Para estimar esa capacidad, el estudiante debe contestar a un número mínimo de 10 ítems y a un máximo de 30 (en la prueba de Estadística Descriptiva) a un mínimo 15 ítems hasta un máximo de 30 ítems (en el test de Diagnóstico Pedagógico) deteniéndose la administración de ambos tests en el momento en el que el error de estimación para la capacidad del estudiante se sitúa por debajo del valor 0.50. No obstante, para bajar ese error de estimación suele ser necesario que el alumno conteste a algunos ítems más de los señalados en ambos mínimos.

Una vez determinadas las especificaciones de cada test, el programa MicroCAT tradujo el fichero de especificaciones del lenguaje de autor a un lenguaje que utilizamos en ordenadores personales (es suficiente un procesador 386). En un solo disco se contó con los ficheros neces

sarios para ejecutar los tests en la forma dispuesta y con los ficheros de instrucciones preparados para cada uno de los tests. El siguiente paso fue instalar los tests en los ordenadores del aula de informática de la Facultad de Ciencias de la Educación, lugar donde se administraron a los estudiantes.

2.6. APLICACIÓN DE LOS T.A.I.S.

La administración de los tests se plantea como una etapa siguiente en la que los alumnos tienen la posibilidad de examinarse de los contenidos recogidos en los diferentes bancos. Nuestra intención no era otra que la de introducir paulatinamente estas aplicaciones informatizadas de modo que no supongan para el alumno una obligación sino una posibilidad. Así, en el caso del test ESTADES se hizo una administración experimental a dieciséis estudiantes sin repercusión alguna en la puntuación final en la asignatura Estadística, mientras en el test LEX se aplicó a 53 estudiantes que optaban voluntariamente a mejorar sus puntuaciones o a alcanzar el nivel de aprobado en la asignatura Diagnóstico Pedagógico después de haber obtenido un mínimo de cuatro puntos en el examen final.

La aplicación se realizó en días diferentes para cada uno de los tests y, según grupos de 15 a 20 alumnos. El tiempo empleado por los alumnos estuvo entre los 15 y los 25 minutos, incluyendo la introducción previa que los profesores hicieron sobre el carácter de las pruebas y sus rasgos más salientes.

Al terminar uno cualquiera de los tests, el estudiante obtenía en pantalla los resultados de la aplicación: la identificación de cada uno de los items, una indicación sobre el acierto o error para cada ítem, el porcentaje de items acertados y la puntuación en la prueba en una escala de 0 a 10 puntos.

Tras terminar la aplicación de cada una de las pruebas los estudiantes eran invitados a responder a un cuestionario para valorar distintos aspectos de la misma.

3. EVALUACIÓN DE LOS T.A.I.S POR PARTE DEL ALUMNADO

Tras la aplicación de los T.A.I.s., en el desarrollo del proyecto hemos pretendido conocer la opinión del alumnado sobre la experiencia. En concreto, nos interesa evaluar el grado de adecuación y el atractivo que tiene su administración por ordenador, así como el formato de presentación; y, por otro lado, su capacidad de adaptarse al nivel de conocimientos de la materia que presenta el/la examinado/a.

3.1. PROCEDIMIENTOS DE RECOGIDA DE INFORMACIÓN

A fin de recoger las opiniones de lo/as estudiantes sobre las cuestiones mencionadas, hemos elaborado una *Escala para la valoración de la Prueba Adaptativa Informatizada*, de acuerdo con el siguiente procedimiento:

Formulación de un conjunto previo de enunciados dirigidos a recabar la información deseada, de acuerdo con las normas habituales (Ander-Egg, 1980; Fox, 1981; Gronlund, 1985;

Morales, 1988, entre otros) para la construcción de este tipo de instrumentos (evitando preguntas que contengan un Año@ o referidas a más de un tema, procurando la relevancia de las preguntas y su independencia de posibles sugerencias, ...).

Lectura de los enunciados, por parte del profesorado participante y por estudiantes en formación -no sometido/as a los T.A.I.s-. Esto permitió una selección y depuración de los ítemes iniciales en función de criterios de claridad, brevedad y concisión.

Decisión sobre la gradación de las respuestas (de 1 a 5, siendo 1 la opinión que no es Anada@ de acuerdo con el enunciado ofrecido).

En la versión definitiva del instrumento (ver Anexo 1), se ha incluido una *introducción* inicial en la que, además de indicar el objetivo, se requieren los datos personales de lo/la encuestado/as relevantes para el estudio. Asimismo, la escala contiene 17 enunciados que responden a las dos dimensiones explicitadas en nuestro objetivo inicial, según se muestra en el Cuadro 1.

Cuadro 1: Distribución de los enunciados de acuerdo con las dimensiones objeto de evaluación.

DIMENSIÓN	ENUNCIADOS
Formato de la prueba: destinada a conocer la opinión del alumnado sobre las instrucciones ofrecidas, la presentación de la prueba, su administración por ordenador...	1, 2, 3, 4, 5, 6, 7, 8, 9 y 16
Nivel de adaptación: valoración que hacen del grado en que la prueba se ajusta al nivel de conocimientos de cada estudiante.	10, 11, 12, 13, 14, 15 y 17

La administración del instrumento tuvo lugar inmediatamente después de que cada estudiante finalizó la realización del correspondiente T.A.I., asegurando, como es obvio, el anonimato de su respuesta.

3.2. MUESTRA

El número final de estudiantes que realizaron los T.A.I.s, y que, por tanto, cumplimentaron la escala, es de 69, de lo/as cuales, 53 se evaluaron mediante el T.A.I. de ALEX@ (Diagnóstico Pedagógico) y 16 mediante el ESTADES de AEstadística Descriptiva@. En cuanto al sexo de lo/as encuestado/as, abunda la presencia femenina (un 92,8%), debido a que, en la titulación de Pedagogía, el alumnado se compone mayoritariamente por mujeres. Y, en la variable A experiencia previa con ordenadores@, la muestra está más equilibrada, entre el 45,5% que manifiesta tener dicha experiencia y el 54,5% que no la tiene.

3.3. ANÁLISIS DE LOS DATOS DE EVALUACIÓN

Se ha realizado un estudio descriptivo (porcentajes, medias y desviaciones típicas) de los distintos ítemes de los que se compone la escala a través del paquete estadístico SPSS. Asimismo, se ha aplicado la prueba t de comparación de medias para conocer si existen diferen

cias significativas en las medias de los 17 enunciados en función del T.A.I. (AESTADES@ o ALEX@) que se ha realizado. Se aplicó igualmente la prueba t para comprobar si existen diferencias en las valoraciones realizadas por las alumnas y los alumnos, así como por aquellos/as que tienen experiencia previa con ordenadores y los/as que no la tienen. En la realización de estos contrastes, se ha trabajado con un nivel de significación $=0.01$.

Con la finalidad de conocer las dimensiones subyacentes a esta escala, se ha realizado también un análisis factorial exploratorio mediante el método de componentes principales (rotación varimax).

3.5. PRESENTACIÓN DE LOS RESULTADOS

Los resultados del análisis factorial exploratorio (ver tabla 17) muestran que existe un primer factor en el que saturan la mayoría de los ítems incluidos dentro de la dimensión A nivel de adaptación de la prueba@, así como un segundo factor que sólo correlaciona con algunos de los enunciados de la dimensión Aformato@ (los ítems 1, 5, 6, 8, 9 y 16, sin embargo, no saturan en el factor II). El tipo de estructura dimensional apuntada, en la que la segunda dimensión no aparece claramente definida, puede deberse al escaso tamaño de la muestra y a la baja discriminación de algunos ítems².

Tabla 17. Resultados del análisis factorial (las celdas sombreadas representan resultados que coinciden con nuestra propuesta de distribución de los ítems por dimensión).

	FACTOR I	FACTOR II
ITEM 1	1771	4885
ITEM 2	3005	27590
ITEM 3	-12121	64538
ITEM 4	11833	85669
ITEM 5	3788	2299
ITEM 6	12810	-12261
ITEM 7	4368	31424
ITEM 8	61083	-4679
ITEM 9	31711	21302
ITEM 10	68064	12463
ITEM 11	46711	15140
ITEM 12	83874	15237
ITEM 13	33009	73413
ITEM 14	77505	480
ITEM 15	75766	30081
ITEM 16	-11441	18591
ITEM 17	45764	18591

² Como puede apreciarse en las tablas 19 y 21, algunos ítems presentan desviaciones típicas con valores muy bajos, existiendo un alto grado de acuerdo en las opiniones manifestadas por los/as sujetos.

A pesar de que la interpretación del análisis factorial arroja algunas dudas, hemos decidido utilizar las dos dimensiones subyacentes que en él se dibujan como base para presentar los resultados del estudio descriptivo, diferenciando así la información sobre el Aformato@ y Anivel de adaptación@ de los T.A.I.s.

Por otro lado y puesto que los resultados de la aplicación de la prueba t, muestran que, con $\alpha=0.01$, no existen diferencias significativas en las valoraciones de los dos T.A.I.s administradas (salvo para los ítemes 6, 12 y 14), vamos a presentar los valores alcanzados por la totalidad del alumnado, sin distinguir si la prueba que éstos/as valoran es la de Estadística o de Diagnóstico, a excepción de los enunciados 6, 12 y 14, en los que la significatividad de las diferencias encontradas, aconseja un estudio específico de la valoración en cada T.A.I.

a) Valoración del formato de la pruebas.

En la tabla 18 se presentan los resultados del estudio descriptivo de los ítemes que componen esta dimensión. En general, la valoración que hace el alumnado es bastante positiva para todos los aspectos considerados: las preguntas son legibles, los resultados son inmediatos, fáciles de comprender, no es necesario poseer conocimientos de informática, la prueba no resulta pesada, etc. Como puede observarse, las medias alcanzadas en todos los enunciados superan ampliamente el valor 3, a excepción del ítem 5, en el cual una media de 1,26 puede considerarse una buena valoración, al tener este ítem una formulación negativa.

Por otra parte, la mayoría de los ítemes presentan valores muy bajos en sus desviaciones típicas, lo cual es un indicador de la homogeneidad de las respuestas, existiendo un alto grado de acuerdo en las opiniones manifestadas.

Como se comentó anteriormente, en el ítem 6 se han encontrado diferencias significativas en la opinión manifestada por lo/as estudiantes que realizaron el T.A.I. de Diagnóstico y lo/a que realizaron la de Estadística, estando esta última prueba mejor valorada por lo/as estudiantes, como se observa en la tabla 19.

En lo que respecta a las diferencias en las valoraciones en función del sexo, los alumnos tienden a valorar más alto casi todos los ítemes que componen esta dimensión, si bien sólo existen diferencias significativas en el ítem 2 -@los ítemes en pantalla se leen con facilidad@-, en el cual, la media de las mujeres (4,5) supera considerablemente a la de los hombres (3,4). No obstante, el escaso número de varones ($n=5$) en nuestra muestra nos obliga a interpretar con cautela este resultado.

También existen diferencias significativas en algunos ítemes en función de tener o no experiencia previa con ordenadores, de tal manera que lo/as que tienen dicha experiencia valoran más positivamente el modo en que comienza la prueba (ítem 1: $t=2.9$; $p=0.005$), e que ésta se administre por ordenador (ítem 3: $t=2.94$; $p=0.005$), la mayor facilidad de la prueba por estar informatizada (ítem 4: $t=3.87$; $p=0.006$) y el haber sabido en todo momento cómo trabajar con el ordenador (ítem 7: $t=3.29$; $p=0.002$).

Tabla 18. Resultados del estudio descriptivo de los enunciados que componen la dimensión A *Valoración del formato de las pruebas*@.

ENUNCIADO	PORCENTAJE DE RESPUESTA					\bar{x}	S_x
	Nada	Poco	Algo	Bastante	Mucho		
(1) <i>El modo en que ha comenzado la prueba me ha gustado.</i>	0	11,6	23,2	50,7	14,5	3,681	0,866
(2) <i>Los ítemes en pantalla se leen con facilidad.</i>	1,4	2,9	4,3	34,8	56,5	4,42	0,83
(3) <i>Me gusta trabajar en una prueba como esta cuando se presenta en un ordenador.</i>	8,8	11,8	30,9	26,5	22,1	3,412	1,212
(4) <i>Contestar a estas preguntas por ordenador es más sencillo que hacerlo con lápiz y papel.</i>	13	20,3	24,6	33,3	8,7	3,043	1,194
(5) <i>Trabajar con la prueba requiere tener conocimientos de informática.</i>	75	23,5	1,5	0	0	1,265	0,477
(7) <i>En todo momento he sabido cómo trabajar con el ordenador.</i>	0	1,4	10,1	52,2	36,2	4,232	0,689
(8) <i>Los resultados que se obtienen en la prueba son fáciles de comprender.</i>	2,9	8,7	42	29	17,4	3,493	0,98
(9) <i>Me gusta obtener los resultados de una prueba nada más terminar.</i>	1,5	1,5	10,3	20,6	66,2	4,485	0,855
(16) <i>Mientras contestaba a las preguntas se me ha pasado el tiempo rápido.</i>	7,2	4,3	8,7	40,6	39,1	4	1,15
VALORES PROMEDIO PARA LA DIMENSIÓN	12,2	9,56	17,2 9	31,97	2897	3,559	0,917

Tabla 19. Resultados del ítem 6 para cada una de los dos T.A.I.s

(6) <i>Las instrucciones son claras y fáciles de seguir.</i>	Porcentajes					\bar{x}	S_x	Valor de t	Grado de Significación
	Nada	Poco	Algo	Bastante	Mucho				
PRUEBA DE ESTADÍSTICA DESCRIPTIVA	0	0	0	18,8	81,3	4,81	0,4	2,839	0,007
PRUEBA DE LEGISLACIÓN (DIAGNÓSTICO)	0	0	7,5	41,5	50,9	4,43	0,64		

b) Valoración del grado en que las pruebas se adaptan al nivel de conocimientos de la materia

La tabla 19 recoge los resultados del análisis descriptivo de los enunciados incluidos en esta dimensión. El alumnado ha valorado muy positivamente todas las cuestiones: capacidad del T.A.I. para ajustarse progresivamente a su nivel de conocimientos, diferencias en las pruebas según el nivel de la persona que las realiza, nivel de dificultad de las preguntas. Todas las medias - a excepción del enunciado 15, que recibe una valoración algo más baja - han superado de nuevo el valor 3, estando la mayoría de las opiniones entre las opciones A y B. Abastante@.

Tabla 20. Resultados del estudio descriptivo de los enunciados que componen la dimensión "Valoración del grado en que las pruebas se adaptan al nivel de conocimientos de la materia".

ENUNCIADO	PORCENTAJE DE RESPUESTA					x	S _x
	Nada	Poco	Algo	Bastante	Mucho		
(10) A medida que iba contestando a las preguntas, he notado que el nivel de dificultad de éstas se ajustaba progresivamente a mis conocimientos.	4,4	11,8	35,3	38,2	10,3	3,382	0,978
(11) Me parece bien que cada estudiante haga una prueba diferente según la capacidad que demuestre.	2,9	5,8	18,8	33,3	39,1	4	1,043
(13) Las preguntas me resultan fáciles.	4,5	16,4	53,7	23,9	1,5	3,015	0,807
(15) Contestando a la prueba he clarificado algunos conceptos.	11,6	26,1	34,8	17,4	10,1	2,884	1,145
(17) El número de preguntas que he contestado es suficiente para saber si conozco la materia.	8,7	5,8	29	40,6	15,9	3,493	1,106
VALORES PROMEDIO PARA LA DIMENSIÓN	6,42	13,18	34,32	30,68	15,38	3,355	1,016

Las desviaciones típicas alcanzan valores algo más altos que los presentados en la dimensión A-formato de las pruebas@, siendo 1,016 el valor promedio de la desviación para toda la dimensión. Existe, por tanto, menor cohesión en las respuestas, aunque el nivel de acuerdo entre lo/as sujetos sigue siendo alto.

En los resultados de las valoraciones de los ítems 12 y 14, se han encontrado diferencias significativas - con un =0.01- entre lo/as estudiantes que realizaron el T.A.I. de Diagnóstico : lo/as que realizaron la de Estadística. Esta última prueba resulta mejor valorada en lo que respecta a su capacidad para informar al alumno/a sobre su nivel de conocimientos y de ayudarle a plantear su estudio, como se observa en la tabla 21.

Tabla 21. Resultados de los ítemes 12 y 14 para cada una de los dos T.A.I.s

<i>(12) La prueba me informa sobre mi nivel de conocimientos.</i>	Porcentajes					\bar{x}	S_x	Valor de t	Grado de Significación
	Nada	Poco	Algo	Bastante	Mucho				
PRUEBA DE ESTADÍSTICA DESCRIPTIVA	0	0	12,5	43,8	43,8	4,31	0,7	3,532	0,001
PRUEBA DE LEGISLACIÓN (DIAGNÓSTICO)	5,8	11,5	40,4	28,8	13,5	3,33	1,04		
<i>(14) Esta prueba me ha ayudado a saber cómo debo estudiar.</i>	Porcentajes					\bar{x}	S_x	Valor de t	Grado de Significación
	Nada	Poco	Algo	Bastante	Mucho				
PRUEBA DE ESTADÍSTICA DESCRIPTIVA	0	12,5	50,0	6,3	31,3	3,56	1,09	3,409	0,001
PRUEBA DE LEGISLACIÓN (DIAGNÓSTICO)	15,4	36,5	30,8	13,5	3,8	2,54	1,04		

En relación con las diferencias en función del sexo del alumnado y al contrario de lo que ocurría en la dimensión Aformato de las pruebas@, son las alumnas las que tienden a valorar más alto casi todos los ítemes que componen esta dimensión, si bien las diferencias no son en ningún caso significativas. No se encuentran tampoco diferencias significativas -con $\alpha=0.01$ - en función de tener o no experiencias anteriores en el uso de ordenadores, en la valoración que se hace de los distintos ítemes de la dimensión. Es decir, la opinión del alumnado sobre el nivel de adaptación de la prueba es similar, tanto si se ha manejado previamente el ordenador como si no.

4. CONCLUSIONES

Una primera conclusión que cabría extraer de este proceso de innovación habla claramente de la posibilidad de realizar cambios que afecten a algún aspecto de la enseñanza más allá del binomio un profesor/una asignatura, a pesar de las dificultades de coordinación existentes entre los profesores de una o varias asignaturas.

Otra conclusión de esta experiencia es que pueden elaborarse tests adaptativos informatizados para evaluar el rendimiento de los estudiantes en la universidad, en relación con el contenido de asignaturas incluidas en los currícula oficiales. En este sentido, los TAI.s puede plantearse como un procedimiento de evaluación susceptible de ser incluido dentro de los que ordinariamente se utilizan en las universidades para valorar los aprendizajes.

Respecto a la evaluación realizada sobre los dos TAI.s construidos, un primer dato que puede llamar la atención es la escasa participación de los estudiantes en ambas pruebas. No obstante, debe tenerse en cuenta que las aplicaciones de LEX y ESTADES tuvieron lugar a

final de curso -tercera semana de junio-, período en el que muchos estudiantes se hayan inmersos en exámenes finales de diferentes asignaturas.

A tenor de los resultados expuestos, podemos concluir que los T.A.I.s elaborados y aplicados han recibido una buena valoración por parte de nuestro alumnado. En este sentido, el hecho de que sean presentados en un entorno informático, no sólo no plantea ningún problema para su aplicación, sino que los hace más sencillos que los de lápiz y papel, para una gran mayoría de lo/as estudiantes. Asimismo, a nivel de su formato, presentan otras ventajas tales como un mayor atractivo o la inmediatez de los resultados de la evaluación.

Por otra parte, lo/as estudiantes han valorado también de forma muy positiva el grado de adaptación de los T.A.I.s a su nivel de conocimientos de la materia. A este respecto, no parecen muy significativas las respuestas al ítem 12 -@la prueba me informa sobre mi nivel de conocimientos@-, que sugieren que las pruebas aplicadas se han mostrado eficientes para medir sus conocimientos (recuérdese que las medias obtenidas fueron de 4,31 para el T.A. de Estadística y de 3,33 para la de Diagnóstico).

En nuestra opinión, estos resultados apuntan a que las pruebas adaptativas administradas por ordenador podrían constituirse en estrategias adecuadas para la evaluación del alumnado universitario, con las consiguientes ventajas que pueden aportar, tanto desde el punto de vista del profesorado como del alumnado. Por tanto, es hora ya que dejen de aplicarse experimentalmente, como iniciativa de profesore/as aislado/as, y adquieran una mayor presencia protagonismo en la evaluación que se realiza en la universidad.

BIBLIOGRAFÍA

- ANDER-EGG, E. (1980). *Técnicas de investigación social*. Buenos Aires. El Cid.
- ASSESSMENT SYSTEMS CORPORATION (1995). *User's manual for the XCALIBUR Marginal Maximum-Likelihood Estimation Program (Version 1.0)*. St. Paul, Minnesota.
- BOCK, R.D. y LIEBERMAN, M. (1970). Fitting a response model to a dichotomously scored item. *Psychometrika*, 35, 179-197.
- FOX, D.J. (1981). *El proceso de investigación en educación*. Pamplona. Eunsa.
- GIL, J.; GARCÍA, E. Y RODRÍGUEZ, G. (1995). Estadísticos de la teoría clásica de los test y ajuste a modelos de la teoría de respuesta a los ítems, en AIDIPE (Comp.) *Estudios de investigación educativa en intervención psicopedagógica*, 183-186. Valencia. AIDIPE.
- GRONLUND, N.E. (1985). *Measurement and evaluation in teaching*. Nueva York. MacMillan.
- HAMBLETON, R.K.; ZAAL, J.N. Y PIETERS, J. (1991). Computerized adaptive testing theory, applications, and standards, en R.K. Hambleton y J.N. Zaal (Eds.), *Advances in Educational and Psychological Testing*. Londres. Kluwer Academic Press.
- LINN, R. (1988). Medición educativa: Algunos problemas y tendencias actuales, en I. DENDALUCE (Coord.). *Aspectos metodológicos de la investigación educativa*, 148-163, Madrid. Narcea.

- MORALES, P. (1988). *Medición de actitudes en psicología y educación*. San Sebastián: Tarttalo.
- MUÑIZ, J. (1990). *Teoría de respuesta a los items*. Madrid. Pirámide.
- RENOM, J. (1993). *Tests adaptativos computerizados. Fundamentos y aplicaciones*. Barcelona. PPU
- WAINER, H. (1990) (Ed). *Computerized Adaptive Testing: A primer*. Hillsdale. Nueva Jersey. Lawrence Erlbaum Associates Pub.
- WEISS, D.J. (1990). Adaptive testing, en J.P. KEEVES (Ed.). *Educational research, methodology, and measurement: An international handbook*, 372-376. Nueva York. Pergamon Press.

ANEXO 1

ESCALA PARA LA VALORACIÓN DE L TEST ADAPTATIVO INFORMATIZADO (TAI)

Prueba que acabas de realizar: "ESTADÍSTICA DESCRIPTIVA"
 "LEGISLACIÓN (DIAGNÓSTICO)"

Edad: ----- Sexo : MUJER Experiencia previa SI
 HOMBRE con ordenadores: NO

Tras terminar tu trabajo con la Test Adaptativo Informatizado (TAI), nos gustaría que expresaras tu opinión sobre distintos aspectos de la misma. Se trata de rodear con un círculo la opción que consideres correcta en cada una de las afirmaciones que aparecen a continuación.

+ + + + +	Mucho	Bastante	Algo	Poco	Nada
1. El modo en que ha comenzado la prueba me ha gustado.	5	4	3	2	1
2. Los ítems en pantalla se leen con facilidad.	5	4	3	2	1
3. Me gusta trabajar en una prueba como ésta cuando se presenta en un ordenador.	5	4	3	2	1
4. Contestar a estas preguntas por ordenador es más sencillo que hacerlo con lápiz y papel.	5	4	3	2	1
5. Trabajar con la prueba requiere tener conocimientos de informática.	5	4	3	2	1
6. Las instrucciones son claras y fáciles de seguir.	5	4	3	2	1
7. En todo momento he sabido cómo trabajar con el ordenador.	5	4	3	2	1
8. Los resultados que se obtienen en la prueba son fáciles de comprender.	5	4	3	2	1
9. Me gusta obtener los resultados de una prueba nada más terminar.	5	4	3	2	1
10. A medida que iba contestando a las preguntas, he notado que el nivel de dificultad de éstas se ajustaba progresivamente a mis conocimientos.	5	4	3	2	1
11. Me parece bien que cada estudiante haga una prueba diferente según la capacidad que demuestre.	5	4	3	2	1
12. La prueba me informa sobre mi nivel de conocimientos.	5	4	3	2	1
13. Las preguntas me resultan fáciles.	5	4	3	2	1
14. Esta prueba me ha ayudado a saber cómo debo estudiar.	5	4	3	2	1
15. Contestando a la prueba he clarificado algunos conceptos.	5	4	3	2	1
16. Mientras contestaba a las preguntas se me ha pasado el tiempo rápido.	5	4	3	2	1
17. El número de preguntas que he contestado es suficiente para saber si conozco la materia.	5	4	3	2	1