Panorama histórico del desarrollo de algunas de las cuestiones fundamentales en la evaluación de segundas lenguas

TERESA BORDÓN

Resumen

La evaluación de la lengua, más específicamente de las segundas lenguas, es un campo que ha experimentado un notable desarrollo desde sus primeros pasos en los años sesenta del siglo veinte hasta el momento actual. En este artículo, se tratan, siguiendo un hilo cronológico, algunos de los asuntos que más han preocupado a los expertos en evaluación. De esta manera, se puede observar cómo algunos temas —como, por ejemplo, la definición de un modelo de lengua con el que contar para evaluar la competencia del aprendiz— han ido variando de acuerdo con los cambios experimentados en la lingüística aplicada. Por otra parte, la perspectiva cronológica permite observar cómo se ha ido complicando y enriqueciendo el campo de la evaluación de la lengua con la aparición de nuevas áreas de interés, la profundización en las ya existentes, así como con la utilización de las nuevas tecnologías.

Dada la complejidad y extensión del tema, no se puede hacer una revisión completa de todos los aspectos, por lo que se han seleccionado, bien asuntos fundamentales sobre los que se escribe desde uno u otro enfoque a lo largo de las décadas incluidas, bien asuntos que abren nuevos caminos a la investigación o a la práctica de la evaluación. Y, para concluir, se propone como reto la necesidad de una evaluación fundamentalmente ética.

1. Introducción

La evaluación de la lengua es un asunto complejo y poliédrico que cada vez recaba más atención y desde más puntos de vista. Con un lugar propio indiscutible dentro de la lingüística aplicada, desde ella se ha dado un evidente impulso en los últimos veinte años a nuevas líneas de investigación para profundizar en el conocimiento de la naturaleza de la habilidad lingüística o en los procesos de la adquisición de la segunda lengua, sólo por mencionar algunas de las áreas de interés. Igualmente, asuntos ineludibles, como la validez y la fiabilidad de las tareas de examen, siguen ocupando páginas de artículos y libros de evaluación. Más recientemente, la incorporación de las nuevas tecnologías ha proporcionado una nueva herramienta para la investigación teórica y la aplicación práctica.

Asimismo, en virtud del espacio que ocupa la evaluación dentro del currículo de enseñanza de una segunda lengua, se han abierto camino propuestas para un tipo de evaluación (conocida fundamentalmente como cualitativa, aunque dependiendo de su orientación puede recibir otras denominaciones como auténtica, pedagógica o educativa o alternativa) cuyo fin es obtener datos del proceso de enseñanza-aprendizaje en lugar de buscar información acerca del resultado de la instrucción.

Paralelamente, las instituciones que se dedican a diseñar y administrar exámenes a gran escala se esfuerzan en proporcionar unas herramientas validas y fiables que se diseñen y administren de manera adecuada por profesionales capacitados, para que resulten instrumentos de medición "justos". Y, aunque parezca paradójico, cuanto más se investiga y se sabe sobre el tema más evidente resulta su complejidad.

2. Los años sesenta y setenta: crisis del modelo estructuralista

La década de los sesenta está dominada, en el área de evaluación de la lengua, por un tipo de exámenes que utiliza procedimientos de corte estructuralista cuyo objetivo es la obtención de elementos discretos, que, en definitiva, no revelan más que el dominio del sistema lingüístico de la segunda

lengua por parte del candidato y difícilmente reflejan su habilidad para usar la lengua. La evaluación de la actuación del candidato, que, dado el tipo de prueba adoptado, se obtiene de manera indirecta y a través de pruebas de respuesta cerrada (con un predominio de las de selección múltiple), se realiza por medio de técnicas psicométricas, que permiten garantizar la fiabilidad de los resultados de los exámenes, al ser posible una calificación totalmente objetiva. En estos años sobresale la aportación de Robert Lado, figura destacada en el campo de la enseñanza del inglés como lengua extranjera y uno de los primeros[1] en preocuparse por dar un lugar propio a la evaluación de la lengua.

Respecto de lo que ocurre en la evaluación en este período, Bachman (2000) incide en lo señalado anteriormente: la existencia de exámenes encaminados a extraer unidades discretas y la preocupación por establecer la fiabilidad a través de procedimientos psicométricos utilizando una metodología apoyada en la investigación estadística. Y, aunque la investigación en el campo de la evaluación estaba dominada por la hipótesis de que la habilidad lingüística constituye un rasgo unitario, a partir de mediados de los años sesenta y durante todos los setenta, la evaluación de la lengua —como se refleja en exámenes de lengua institucionales a gran escala y en la mayoría de libros de exámenes de esa época— incorpora una visión de la habilidad lingüística como consistente en las cuatro destrezas (escuchar, hablar, leer y escribir) y en una serie de componentes (gramática, vocabulario, pronunciación).

Así las cosas, podemos observar que la evaluación de la lengua que se lleva a cabo en esos años refleja por una parte el tipo de enseñanza de segundas lenguas en auge en esa época y que corresponde a lo que en la pedagogía de segundas lenguas se conoce como etapa estructuralista. No obstante, por otra parte, ese modelo presenta claros síntomas de crisis en la segunda mitad de los años setenta, momento en el que ya existen propuestas didácticas[2] de evidente orientación comunicativa, así como exámenes que reflejan la influencia del nuevo enfoque. De hecho, Weir (2002)[3] señala que en la revisión que se hizo en 1975 del Certificate of Proficiency in English (CPE) se quiso reflejar el cambio de orientación hacia la enseñanza de corte comunicativo que se estaba produciendo desde principios de los años setenta. El examen en lugar de estar dividido en Escrito y Oral, pasó a constar de cinco secciones:

---

[1] Su libro, *Language Testing: The Construction and Use of Foreign Language Tests*, publicado en 1961, constituye una referencia obligada en el campo de la evaluación de lenguas extranjeras, a pesar de que las técnicas recogidas en él se consideran superadas hoy día.

[2] Por ejemplo, en 1975 se publicó el *Threshold Level* y en 1979, el *Nivel Umbral*, su versión española, cuya influencia fue decisiva para dar una orientación comunicativa a la enseñanza de segundas lenguas (SLAGTER, 1979).

[3] Me refiero al trabajo "Innovation and Continuity: CPE – past and present", publicado en el número 10 de *Research Notes* www.CambridgeESOL.org, basado en el capítulo 1 de WEIR y MILANOVIC (Eds.) (2002).

composición, comprensión de lectura, uso del inglés, comprensión auditiva y entrevista. Este nuevo formato muestra que la concepción de la lengua se aleja del modelo que la concibe como una competencia unitaria, y se inclina por la habilidad lingüística consistente en las cuatro destrezas.

En 1979, tiene lugar el primer Language Testing Research Colloquium, que supone el reconocimiento de la evaluación de la lengua como un área de estudio propia dentro del campo de la lingüística aplicada. En él se critica abiertamente el modelo de la habilidad lingüística concebida como un rasgo unitario, que, recordemos, había sido el principal tema de estudio y debate en la década que concluía. Bachman (2000) señala que en los ochenta va a triunfar el modelo de competencia comunicativa, entendida ésta en la línea que proponen los trabajos[4] de Widdowson (1978, 1979, 1983), Savignon (1972, 1983) y Canale y Swain (1981). Propuestas que, de una u otra manera, hacen hincapié en la enseñanza del "uso" de la lengua, en entender ésta como la creación de un discurso propio y la negociación de significado, y en tomar en consideración el contexto. Por lo tanto, si este es el modelo de lengua adoptado para su enseñanza-aprendizaje, la evaluación, parte de tal proceso, deberá asumir el mismo por una razón, en principio, de coherencia. Al mismo tiempo, el tipo de exámenes que no se inserta en un currículo de lengua, como los exámenes de nivel a gran escala, también necesita contar con un modelo de lengua del que partir. Así, el hecho de que la evaluación se decante por entender la lengua como uso (y no sólo como un sistema constituido por unidades discretas), repercutirá en el replanteamiento de las tareas de examen como "tareas auténticas" y en la preocupación por asegurar la validez de cada una de las pruebas en particular y del propio examen en general: aspectos necesarios para garantizar que se puedan hacer (a partir de los resultados del examen) inferencias sobre la habilidad para el uso de la lengua del candidato. Sin olvidar que los exámenes a gran escala influyen, a su vez, en el diseño de cursos para la enseñanza de las lenguas.

3. Los años ochenta: hacia la implantación del enfoque comunicativo

En el International Symposium on Language Testing celebrado en Hong Kong en 1983, se presentaron una serie de trabajos que se recogieron en el volumen *New Directions in Language Testing*[5]. Los trabajos seleccionados en el libro se agrupan en torno a cuatro apartados principales, que reflejan las áreas de mayor interés para investigadores y expertos en evaluación en ese momento.

---

[4] Referencia tomada de BACHMAN (2000: 3).

[5] LEE, FOK, LORD, y LOW (1985).

El primero, evaluación de la lengua y el currículo, recoge artículos que tratan acerca de la evaluación de la lengua cuando ésta se lleva a cabo en el marco de la enseñanza-aprendizaje de la segunda lengua en el aula. Uno de los asuntos que se trata es si el modelo de lengua que se utiliza en los exámenes refleja la adopción de una determinada concepción de la lengua para su enseñanza, o si los exámenes son los que imponen el modelo hasta el punto de llegar a influir en el tipo de lengua que se debe enseñar en el aula. En este sentido, Davies (1985) señala que la mayoría de los exámenes siguen al líder —por líder entiéndase el programa de enseñanza-aprendizaje en el que se insertan tales exámenes—. En este caso, se está refiriendo a exámenes de aprovechamiento porque su objetivo es comprobar si el aprendiz ha cumplido los objetivos del curso. En contraposición a éstos, coloca los exámenes de nivel de dominio, que califica de "más creativos" puesto que sus contenidos y objetivos pueden influir en el "líder". Es decir, éstos últimos dan origen a programas y cursos destinados a preparar a los candidatos que van a realizar tales exámenes. De este modo, este autor señala que cierto tipo de exámenes —básicamente los de aprovechamiento— reflejan la concepción de la lengua adoptada para su enseñanza en el aula, pero otros —los de nivel de dominio a gran escala—, por su influencia, pueden ser pioneros en adoptar un modelo lingüístico que influirá en la enseñanza. Por ejemplo, la concepción de la lengua que se adopta en los DELE sin duda repercute en cómo se aborda la enseñanza de la lengua en el diseño de cursos y materiales destinados a candidatos que aspiran a obtener el diploma del nivel para el que se preparan.

Dentro de este apartado de las relaciones entre evaluación y currículo, merece la pena destacar la aportación de Rea (1985), que plantea varios puntos de interés. Comienza por afirmar que es innegable el hecho de que la evaluación es una parte fundamental del proceso de enseñanza y aprendizaje. A continuación, propone un breve recorrido por el estado de la cuestión en lo que se refiere a la evaluación de corte comunicativo y señala que uno de los asuntos principales en este ámbito es llegar a un consenso acerca de la naturaleza de la base teórica sobre la que se deben construir los exámenes de lengua, cuya necesidad está sobradamente justificada en virtud de su papel decisivo para garantizar la validez del constructo y del contenido. El artículo de Rea refleja varias de las preocupaciones presentes en esa época. Por ejemplo, reflexiona acerca de si es posible mantener la dicotomía chomskiana de competencia y actuación en la evaluación. Y afirma que para ella cualquier comportamiento lingüístico constituye una muestra de actuación, y que, desde esa perspectiva, la competencia *per se* no puede ser el asunto de una evaluación directa: sólo a través de la actuación se pueden inferir niveles de competencia. Rea también se refiere a un asunto en ese momento todavía vigente: el de los exámenes de unidades discretas frente a los integradores. Para ilustrar y aclarar su propuesta propone un esquema clasificatorio (1985: 28) de los rasgos clave para la

evaluación de la lengua dentro del marco de su enseñanza. También incluye una explicación acerca del contraste entre evaluación sumativa y formativa. Y, algo interesante: en su conclusión señala que la evaluación de la lengua de manera comunicativa es todavía problemática. Es decir, que aunque la enseñanza de la lengua desde el paradigma comunicativo está totalmente asumida, la evaluación todavía encuentra problemas en la adopción de este modelo. Y no es de extrañar: medir el uso de la lengua es una tarea mucho más complicada que hacerlo a través de elementos discretos.

En este simposium también se incluyó el tema de los exámenes a gran escala (recogido en el segundo apartado del libro). Éste es, sin duda, un aspecto importantísimo de la evaluación, dadas las complejidades que conlleva la elaboración, administración y calificación de este tipo de pruebas, así como la influencia que ejercen. Destaca, en este sentido, el trabajo de Merrill Swain (1985) en el que la autora explora los principios generales que en su opinión deben regir la elaboración de un examen de lengua a gran escala, identificando los siguientes:

1. Empezar por algún sitio: defiende que es fundamental tener un marco teórico del que partir, ya que el desarrollo de un examen se debe construir a partir de conocimientos y ejemplos ya existentes. Con marco teórico, Swain se refiere fundamentalmente a la necesidad de disponer de una concepción de la naturaleza de la lengua para poder elaborar un examen de lengua (presupuesto en el que coincide con lo anteriormente expuesto por Rea), aunque, como es de esperar tratándose de esta autora, tiene claro que el modelo idóneo será el de competencia comunicativa propuesto por ella y Canale (Canale y Swain, 1980).
2. Concentrarse en el contenido: este principio se refiere a tanto al contenido del material usado, como a la base para las actividades lingüísticas comunicativas y las tareas adoptadas para sacar a la luz actuaciones lingüísticas comunicativas. Para asegurar que los materiales y las tareas son capaces de generar lengua que incluya actuación gramatical, discursiva, sociolingüística y estratégica, Swain (1985: 38) considera que el contenido debe presentar cuatro características: ser motivador, sustancioso, integrador e interactivo por naturaleza. Estos puntos se refieren al punto de vista del candidato, porque un contenido motivador para éste puede no serlo para el examinador.
3. Sesgar con buen fin: para Swain, se debe hacer todo lo posible para que el candidato saque a la luz lo mejor de sí mismo y tenga la mejor actuación posible en el examen. En este sentido sugiere proporcionar tiempo más que suficiente para hacer el examen, dar información al candidato acerca de lo que se espera de él o, incluso, brindarles la oportunidad de utilizar recursos como diccionarios durante la prueba.

4. Trabajar más en el aspecto de la repercusión (*washback*)[6] del examen: esto es, indagar en el efecto que éste tiene en las prácticas de enseñanza. Y considera muy importante involucrar a los profesores en el proceso de diseñar, administrar y calificar exámenes, porque de esta manera, no sólo cambiarán aspectos de lo que enseñan, sino que también sugerirán estrategias alternativas de enseñanza-aprendizaje.

Otro asunto vigente en ese momento (y que todavía se mantiene) es el relativo a la evaluación de la habilidad oral. En cualquier examen actual a gran escala se incluyen pruebas de expresión oral que supongan la utilización de lengua hablada con fines comunicativos, y si no se hace, pongamos por caso, en algunos exámenes de aprovechamiento, es porque la expresión oral se evalúa de manera continua a lo largo del curso. Lo que es impensable hoy día es que la expresión oral se evalúe de manera indirecta como defendía Lado (1961). En este sentido, resulta cuanto menos curioso comprobar que en el momento del Simposium de Hong Kong (1983) varios de los trabajos[7] presentados en esta área se esfuerzan en demostrar la idoneidad de una prueba oral directa y la necesidad de incluirla en los exámenes. También encontramos que preocupan asuntos como la dificultad de evaluar —y calificar— las pruebas orales de la manera más "objetiva" posible mediante la utilización de criterios y la formación de evaluadores (calificadores); establecer el tipo de relación entre el examinador y el candidato; valorar el coste económico de la prueba; o la posibilidad de utilizar apoyo tecnológico para facilitar la evaluación. Preocupaciones que, dicho sea de paso, no están superadas todavía, aunque la experiencia acumulada desde entonces y el perfeccionamiento de la tecnología hayan podido influir favorablemente en el diseño de mejores tareas de examen y criterios de evaluación más operativos.

La validación de los exámenes de lengua: una cuestión fundamental, ya que si un examen no es válido no sirve para nada, fue el motivo de varias de las ponencias presentadas en el simposium, recogiéndose algunas de ellas en el cuarto apartado de *New Directions in Language Testing*. En primer lugar, conviene aclarar el propio concepto de validez, algo difícil de precisar, ya que en el campo de la evaluación se han identificado distintos tipos[8]. En este

---

[6] Repercusión o efecto de rebote (*washback*): se denomina así a la influencia o al impacto que pueden ejercer los exámenes en la enseñanza, ya que, por ejemplo, si los estudiantes tienen que superar un examen con unas ciertas características habrá que orientar la enseñanza a prepararlos para hacerlo con éxito. De esta manera el tipo de examen está definiendo el tipo de enseñanza.

[7] JONES, R., "Some Basic Considerations in Testing Oral Proficiency", pp. 77-84; MORRISON, D. y LEE, N., "Simulating an Academic Tutorial: A Test validation Study", pp. 85-92; BERKOFF, N., "Testing Oral Proficiency: A New Approach", pp. 93-99; PENDERGAST TH. JR. "OLAF N.73: A Computerized Oral Language Analyser and Feedback System", pp. 101-107. Todos en LEE, Y. P. y otros (Eds.) (1985).

[8] Una caracterización sucinta pero operativa se encuentra en el *Multilingual glossary of language testing items*, n.º 6 de *Studies in Language Testing*, Cambridge, UCLES/CUP, pp. 206-208. Y también véase BACHMAN (1990: 285-289).

sentido, Stevenson (1985) propone, en primer lugar, revisar algunas opiniones acerca de la validez que no proceden de profesionales y, sin embargo, están ampliamente extendidas; sobre todo, en lo que respecta a la validez aparente (*face validity*)[9]. Destaca que una creencia muy enraizada pero, por otra parte, ingenua entre los expertos en evaluación es que una prueba de examen en la que el candidato proporciona una actuación directa, es decir, una prueba que incluye tareas de examen que simulan una situación de la realidad, es esa situación. Sin embargo, un experto en medición, tendrá claro que es sólo un ejemplo, una simulación. Las situaciones simuladas deben seleccionarse siguiendo algún principio teórico y la calificación de esas pruebas requiere las mismas bases teóricas si queremos hacer una inferencia válida.

De las aportaciones recogidas en el Simposium de Hong Kong, podemos observar que a principios de los ochenta existe una clara voluntad de establecer definitivamente el modelo de lengua como competencia comunicativa. Y, a partir de ahí, tratar las cuestiones derivadas de esta concepción de la lengua. En los aspectos prácticos de la evaluación, destaca el interés en torno a dos áreas que van a seguir haciendo correr tinta en etapas posteriores: la evaluación de la expresión oral y el complejo asunto de la validez de los exámenes.

Para Bachman (2000), está claro que a finales de los años ochenta la evaluación de la lengua ha conquistado definitivamente su espacio en el territorio de la lingüística aplicada, al haber ampliado su visión de la habilidad lingüística y haber definido una línea de trabajo orientada a asuntos que conciernen tanto al análisis del discurso como a la adquisición de segundas lenguas.

4. La década de los noventa

En los años noventa el gran reto son los exámenes de tipo comunicativo, con todo lo que supone la adopción de una concepción comunicativa de la lengua para su evaluación. En este sentido va a ser muy influyente la propuesta de Bachman (1990), cuya base se encuentra en el modelo de competencia comunicativa de Canale y Swain (1980) y Canale (1983)[10]: mantiene prácticamente los mismos componentes, pero organizados y jerarquizados de otra manera. Bachman (1990)[11] no incluye la competencia estratégica entre los constituyentes de la competencia lingüística, aunque la considera parte de la habili-

---

[9] La validez aparente (*face validity*) es la que se establece en función de la aceptación de una prueba, en cuanto que mide lo que debe medir, por parte de los administradores de la misma o los candidatos que la realizan.

[10] La competencia comunicativa está constituida por cuatro subcompetencias: lingüística, discursiva, sociocultural y estratégica.

[11] BACHMAN (1990). El modelo completo de competencia lingüística se explica en las páginas 87-98. Y la competencia estratégica, en las páginas 98-104.

dad lingüística comunicativa y, además, amplía su alcance respecto del modelo de Canale y Swain. De esta manera, para Bachman, la competencia estratégica no consiste en una serie de estrategias compensatorias que sirven para remediar fallos en la comunicación o para reforzarla, sino que es descrita como la responsable de una serie de funciones en la planificación, valoración y ejecución, que proporcionan los medios para explicar cómo los diversos componentes de la competencia lingüística (gramática, textual, pragmática y sociolingüística) interaccionan entre ellos y con rasgos de situaciones de uso de la lengua. El papel de la competencia estratégica en el marco de este modelo se amplia posteriormente (Bachman y Palmer 1996)[12] y se reelabora como una serie de componentes o estrategias metacognitivas que incluyen establecer las metas —decidir qué se va a hacer—, valorar —establecer qué se necesita, con qué se ha de trabajar y cómo se ha hecho— y planificar —decidir cómo usar lo que se tiene—. Esta propuesta ha resultado muy influyente como base teórica para indagar e investigar en todos y cada uno de los componentes de la habilidad lingüística comunicativa, e, igualmente, ha servido para proporcionar el modelo de lengua en la elaboración de numerosos exámenes.

4.1. El inicio de la década: triunfo del modelo comunicativo

En noviembre de 1989, la IATEFL (*International Association of Teachers of English as a Foreign Language*) organizó en Bournemouth (Reino Unido) un Simposium sobre evaluación de la lengua, titulado *La evaluación de la lengua en los años noventa: el legado comunicativo*. En él se presentaron tanto el resultado de investigaciones acerca de cuestiones teóricas y prácticas en el área de evaluación de la lengua, como modelos concretos de evaluación, procedimientos innovadores para la evaluación de la lengua o proyectos para el desarrollo de nuevos exámenes: se trataba de establecer los asuntos relevantes para la investigación en la década que estaba a punto de comenzar. Lo más destacado de este simposium se recogió en el volumen *Language Testing in the 1990*, que editaron Charles Alderson y Brian North[13].

El volumen, dividido en dos partes, recoge en la primera las conferencias de los investigadores, y, en la segunda, la de aquellos que "no necesariamente se consideran investigadores de evaluación en el ámbito académico" [14] y cuyas aportaciones reflejan las preocupaciones de tipo práctico y los logros

[12] BACHMAN. y PALMER (1996: 70-75).
[13] ALDERSON y NORTH (Eds.) (1991).
[14] ALDERSON y NORTH (Eds.) (1991), Introducción, p. XII.

de proyectos en diversas áreas de la evaluación de la lengua. Los artículos seleccionados para la publicación constituyen una buena muestra de las cuestiones y áreas de interés que venían ocupando a expertos de la teoría y práctica de la evaluación cuando los años ochenta tocaban a su fin y se planteaba el reto de comenzar una nueva década.

Es curioso notar que algunos autores (Carroll, Skehan) se refieren en sus artículos a la falta de permeabilidad para las innovaciones que detectan en el campo de la evaluación. Carroll (1991), basándose en un su propia experiencia, identifica cinco colectivos que se resisten a aceptar cambios. En primer lugar, cita a los profesores, un grupo cuya contribución es fundamental para el desarrollo de los exámenes, pero que en ocasiones ven con suspicacia las innovaciones, ya que no perciben con claridad las razones para llevarlas a cabo. A continuación, menciona a los que se ocupan de administrarlos, para los que la puesta en marcha de un nuevo tipo de examen —y se refiere fundamentalmente a los que se administran a grandes grupos— supone gastos que no siempre consideran justificados, de manera que se muestran reacios a las innovaciones. La tercera causa la atribuye a las diferencias culturales que a veces se esgrimen para ocultar otras resistencias a los cambios en un examen. Otro impedimento está en "la gente": en todos los ambientes hay gente que se opone a los cambios porque siempre le parece que son para mal. Al final de la lista coloca al medio académico, para Carroll la resistencia más insidiosa a los cambios, ya que se da el hecho de que profesores o investigadores de miras estrechas critican todo lo que suena a novedad.

Skehan (1991), por su parte, identifica la no aceptación de innovaciones y cambios, por un lado, con causas que tienen que ver con la teoría acerca de la naturaleza de la lengua. Se refiere a que existen numerosas investigaciones sobre la naturaleza de la lengua realizadas por lingüistas y que ofrecen diversas perspectivas que se van superponiendo las unas a las otras. Por ejemplo, a la gramática estructural, le sigue la gramática transformacional, después los enfoques basados en los actos de habla, a continuación la lengua como discurso y luego, otra vez, Chomsky con la Gramática Universal. Este cambio de orientación, que le parece muy productivo para la investigación lingüística, no lo es tanto, sin embargo, para quien se dedica a la evaluación y los exámenes, que busca un modelo concreto sobre la naturaleza de la lengua con el que poder trabajar. Señala, también, que la oposición a los cambios puede encontrarse en colectivos para los que la puesta en marcha de nuevos exámenes genera un sinnúmero de problemas. No olvidemos que los exámenes a gran escala precisan una rigurosa elaboración y posterior experimentación para garantizar su validez y fiabilidad y, si se realizan cambios en un examen de este tipo, habrá que probar el nuevo instrumento para asegurar su idoneidad, lo cual supone una gran inversión de medios humanos y económicos que no siempre se está dispuesto a asumir. Otra fuente de oposición se

halla en la diversidad de intereses de las personas implicadas en la evaluación: investigadores, profesores, expertos en evaluación, gerentes, etc. Efectuar cambios en un examen no afecta igualmente a quien los defiende por considerarlos resultado de la aplicación de una teoría que los va a mejorar, como a quien tiene que administrarlos directamente, para el cual las modificaciones realizadas suponen adoptar otra logística, gastos adicionales, etc. Aunque expuestas de distinta manera, son evidentes las coincidencias entre los factores que impiden o ralentizan los cambios en los exámenes de lengua en la versión de Carroll y de Skehan. Este último, en su artículo, también incluye su propuesta acerca de las áreas en las que piensa que debe incidir la evaluación para progresar, identificando como relevantes el impulso de la autoevaluación, la necesidad de perfeccionar la evaluación con referencia a un criterio, y, sobre todo, el desarrollo de los exámenes de aprovechamiento.

Un asunto de gran relevancia es el que ocupa a Alderson (1991): la utilización de escalas para asignar una puntuación en distintos niveles de actuación. Las escalas surgen de la necesidad de tener que otorgar un puntuación lo más "objetiva" posible a muestras de actuación que no se pueden medir por procedimientos objetivos, tal como es el caso de la expresión oral y escrita en exámenes comunicativos en los que el candidato da muestras directas de lengua. Esta manera de evaluar la lengua resulta, a todas luces, más válida que medir la expresión oral de manera indirecta, utilizando pruebas orientadas a extraer elementos discretos. No obstante, si bien la validez se incrementa, la fiabilidad puede sufrir un fuerte menoscabo en cuanto a la pérdida de estabilidad en la calificación de los candidatos, ya que no se podrá evaluar de manera totalmente objetiva, como permiten las pruebas de respuesta cerrada. Las escalas son un instrumento fundamental para los evaluadores que tienen que valorar las destrezas de producción, ya que constituyen una manera de paliar la subjetividad inherente a la calificación de muestras de lengua por personas, al proporcionar un estándar común para interpretar los resultados de pruebas de respuesta abierta. Mediante ellas, además, se garantiza la fiabilidad de las puntuaciones. Alderson destaca, asimismo, que las escalas son útiles para los usuarios del examen ya que le permiten entender lo que significa su puntuación: si un candidato en una escala de cinco niveles es calificado como 3, no tiene más que leer la descripción de este nivel para comprobar qué significa en cuanto a su habilidad para usar la lengua.

El trabajo de Rea (1991) se plantea responder a dos preguntas: 1. ¿Qué es la gramática comunicativa?; 2. ¿La gramática comunicativa se puede medir comunicativamente? La autora es algo pesimista al respecto en el sentido de que piensa que, mientras en otras áreas de la evaluación de la lengua se ha evolucionado hacia el modelo comunicativo, en la evaluación de la gramática no se reflejan las innovaciones. Y eso, a pesar de que el modelo de gramática actual entiende el conocimiento de la lengua no sólo como el dominio de

un código y un sistema, sino también como la manera en que se procesa la lengua y cómo se usa para fines comunicativos. Concluye afirmando que, a su parecer, la evaluación comunicativa de la gramática comunicativa sólo puede ocurrir cuando el candidato está involucrado en un acto comunicativo en contextos que le permiten la creación de significado.

Davies (1991)[15] trata acerca de los principios en los que se fundamenta su visión de la evaluación, pasa revista a los trabajos presentados en el coloquio y se extiende acerca de la relación de la evaluación con la lingüística aplicada. Para él, el objeto de la lingüística aplicada es el uso de la lengua en un sentido muy amplio. Sin embargo, las pruebas destinadas a evaluar el uso de la lengua están sujetas a una serie de restricciones impuestas por las características propias del examen. Concluye su artículo aventurando en qué sentido se va a desarrollar la evaluación de la lengua en los años noventa. Cree que, en el plano teórico, se debe profundizar en las relaciones entre evaluación de la lengua y adquisición de segundas lenguas, mientras que en la práctica piensa que es importante que la enseñanza de la lengua y su evaluación estén coordinadas (en el sentido de que sigan el mismo modelo de concepción de la lengua). En este sentido, Davies sostiene que lo relevante es el efecto que tienen los exámenes en la enseñanza para prepararlos, y afirma que los exámenes comunicativos se han introducido no como consecuencia de que se enseñara la lengua comunicativa, sino porque se impuso que la lengua se debe enseñar de esa manera. También defiende que en la década que comienza es más importante concentrarse en cuestiones de validez más que de fiabilidad. Y añade que la teoría y la práctica se necesitan mutuamente y que sólo se tendrán mejores exámenes con mejor teoría.

4.2. Áreas de mayor impulso durante la década

Valgan las referencias anteriores de muestra de algunos de los asuntos que interesaban en el campo de la evaluación al iniciarse los años noventa. Bachman (2000) destaca como lo más relevante de la década, cuando ésta ya está tocando a su fin, la ampliación de los estudios que profundizan en las relaciones entre evaluación de la lengua y otras áreas de la lingüística aplicada, y el desarrollo de puntos de contacto entre la evaluación de la lengua y la medición en el contexto general de la educación. Concretamente, señala que la evaluación ha profundizado en cinco asuntos: la metodología para la investigación; los avances prácticos; los factores que afectan la actuación en los exámenes de lengua; la calificación de la actuación; y la preocupación respecto de la ética de los exámenes, junto con la profesionalización del campo de la evaluación.

---

[15] DAVIES, A., "Language Testing in the 1990s", en CH. ALDERSON y B. NORTH (Eds.) (1991), pp. 136-149.

Respecto de la metodología en la investigación, destaca los avances en varias técnicas de evaluación cuantitativas. Recoge, por ejemplo, las innovaciones en la medición de la actuación realizada con referencia a un criterio, según la cual los resultados del examen se interpretan en función de un dominio de la habilidad lingüística, frente a la medición con relación a la actuación de los otros candidatos (referencia a la norma). Alude, asimismo, a los estudios que profundizan en aspectos de la evaluación cuantitativa más elaborados y complicados como la teoría de la generalizabilidad (*Generalizability theory*) y la teoría de respuesta al ítem (*Item response theory*). Esta última, señala, se utiliza, con mucha frecuencia, en alguna de sus variedades en la medición educativa, mientras que en los exámenes de lengua se sigue adoptando preferentemente alguna de las formas del modelo de Rasch[16]. Igualmente, Bachman se refiere al aumento de la investigación que indaga en enfoques para la evaluación cualitativa y al desarrollo de procedimientos orientados a llevar a cabo este tipo valoración.

Como es bien sabido, la evaluación de la lengua se puede realizar mediante la utilización de pruebas de examen destinadas a medir el resultado de un proceso de aprendizaje o a determinar la posesión de un cierto nivel de dominio, o por medio de procedimientos enfocados a obtener información del proceso de aprendizaje. La evaluación de la lengua que utiliza exámenes se llama también cuantitativa, mientras que el segundo tipo se conoce, en general, como evaluación cualitativa, y puede adoptar una variedad de procedimientos: análisis del discurso y de la conversaciones, informes del profesor o de los evaluadores, observaciones de clase, diarios del aprendiz, cuestionarios o entrevistas y los portafolios. Esta última también ha recibido otras denominaciones como evaluación auténtica, educativa, pedagógica y alternativa.

En este ámbito destacan las investigaciones de Brown y Hudson (1998) que prefieren hablar de alternativas en la evaluación de la lengua en vez de *evaluación alternativa* ya que esta caracterización la encuentran cargada de connotaciones negativas en el sentido de que parece que sea una manera completamente nueva de evaluar, totalmente separada y diferente de lo que se ha hecho hasta el momento y que, además, no se debe someter a los requisitos que implica la construcción de exámenes y la toma de decisiones a partir de sus resultados. Estos autores incluyen como alternativas todo tipo de pruebas de exámenes, así como los procedimientos cualitativos.

En el punto de los avances prácticos, Bachman (2000) menciona los logros alcanzados en la evaluación de la pragmática, el desarrollo de la evaluación de la lengua con fines específicos, el renacido interés por las pruebas de vocabula-

[16] El modelo de Rasch, basado en el enfoque estadístico propuesto por el matemático danés G. RASCH (1906-1980), consiste en una fórmula matemática (que se puede aplicar a cualquier disciplina) para la relación entre la probabilidad de éxito y la diferencia entre la habilidad de un individuo y la dificultad de un ítem.

rio y destaca de manera especial la aplicación de la informática a la evaluación. Respecto de este último, cabe decir que es evidente que la década de los noventa ha conocido una gran expansión de la tecnología y de las técnicas informáticas aplicadas a la enseñanza y también a los exámenes. Circunstancia favorecida por el perfeccionamiento de técnicas de medición cuantitativas (como la de respuesta al ítem), la accesibilidad a ordenadores personales y la aparición de programas informáticos de fácil manejo para los usuarios. No obstante, conviene recordar que ya en la década anterior se reflexionaba acerca del papel de la informática en la enseñanza y la evaluación de segundas lenguas. Alderson (1986) publicó un artículo en el que recoge el lado positivo de los procedimientos informáticos, siendo, a su modo de ver, lo más interesante de ellos que permitirán aumentar la interacción entre la actuación del candidato y: a) las posibles evaluaciones de la actuación, b) el contenido de los ítems y del examen. Así se podrán obtener datos para la investigación acerca de cómo se procesa y aprende la información o modificar los ítems del examen. De esto se deriva la posibilidad de obtener una retroalimentación inmediata que conducirá a la desaparición de la distinción entre ejercicio y examen, a la posibilidad de un diagnóstico más acertado, a una intervención pedagógica a través de las claves y guía tras un examen, o a que los candidatos dispongan de la oportunidad de revisar sus respuestas antes de que se guarden definitivamente en la memoria del ordenador. En definitiva, piensa que la utilización de un programa informático aplicado a la evaluación puede contribuir a desarrollar las teorías acerca de lo que debe ser un examen de diagnóstico y quizá proporcionar una mejor comprensión de la naturaleza del aprendizaje de la lengua. Sin embargo, también advierte de los posibles peligros, como, por ejemplo, el que una aplicación perversa del ordenador puede limitar, por ejemplo, el desarrollo de procedimientos para extraer información del candidato que no sean ítems de selección múltiple.

Por otra parte, la incorporación de las nuevas tecnologías ha aportado, y, sin duda, seguirá haciéndolo, grandes beneficios a la evaluación, especialmente para la administración y corrección o calificación de exámenes a gran escala. No obstante, como se mencionaba anteriormente, puede suponer ciertas limitaciones y conllevar, asimismo, el replanteamiento de algunas cuestiones básicas como ¿es lo mismo una prueba de comprensión de lectura en la que el candidato lee un texto impreso en papel que la misma prueba en soporte informático?

Bachman (2000) también se refiere a los factores que afectan la actuación del candidato al realizar exámenes de lengua. Señala que la investigación se ha centrado especialmente en tres aspectos:

1. Las características del procedimiento de examen —algo que ha estudiado el propio Bachman[17]—, y de los calificadores, con trabajos (Chalhoub-Deville, 1996 y McNamara, 1996) que han encontrado que se dan des-

[17] En BACHMAN y PALMER (1996).

ajustes en las calificaciones en función de la lengua materna del evaluador, así como de su formación previa o experiencia.
2. Los procesos y estrategias adoptados por los candidatos al responder a las tareas de examen, para cuyo estudio se han utilizado métodos introspectivos (Wu, 1998) o cuestionarios (Purpura, 1999).
3. Las características de los propios candidatos, con estudios (Kunnan, 1995) que han indagado acerca de cómo la formación académica, la lengua materna, la cultura, o el sexo repercuten en la actuación del candidato en un examen[18].

4.3. Perfeccionamiento de técnicas de evaluación existentes y aparición de nuevos procedimientos

Cuando hablamos de la evaluación de la lengua a través de la actuación, nos referimos, fundamentalmente, a la evaluación de muestras de lengua oral y escrita obtenidas a través de pruebas directas; es decir, hablando y escribiendo[19]. Tratar en profundidad de estos asuntos requeriría un tiempo considerable y un espacio que no es el de este artículo. En este momento, vamos a ceñirnos a mencionar dos tipos de procedimientos: los portafolios, para la evaluación de la producción escrita y las entrevistas orales, para la evaluación de la expresión oral. El primero de ellos está relacionado con la evaluación cualitativa, y constituye una excelente herramienta para la obtención de información del proceso de aprendizaje de la segunda lengua. La entrevista oral, por su parte, se utiliza con frecuencia en todo tipo de exámenes, especialmente en los destinados a determinar niveles de habilidad lingüística.

La carpeta de trabajos o portafolio (*Portfolio*), cada vez más utilizado en el ámbito de la enseñanza de segundas lenguas, constituye una técnica que permite, en primer lugar, averiguar cómo el aprendiz adquiere la lengua, así como extraer información de los errores e indagar acerca de las estrategias empleadas. Contribuye, asimismo, a afianzar la autonomía del aprendiz y le facilita llevar a cabo su propia autoevaluación. Un ejemplo es el Portafolio de la lengua europeo (*European Language Portfolio)* que consta de tres componentes obligatorios:

1. un pasaporte lingüístico, en el que se debe hacer constar una visión general de la identidad lingüística de su dueño, de su experiencia de aprendizaje de la lengua y de su aprovechamiento, así como una autoevaluación de su habilidad lingüística en la segunda lengua.

---

[18] Las referencias recogidas en este párrafo están tomadas del artículo de BACHMAN (2000).

[19] Recordemos, no obstante, que para REA (1985) cualquier muestra de lengua obtenida de un aprendiz es actuación.

2. una biografía lingüística, que incluirá información del proceso de aprendizaje del poseedor del pasaporte, y le permitirá definir sus objetivos de aprendizaje, revisar su progreso, registrar el aprendizaje significativo y las experiencias interculturales, así como reflexionar sobre diversos aspectos del aprendizaje y el uso de la lengua
3. un dossier, que consiste en reunir ejemplos del trabajo del aprendiz en la segunda lengua.

Según David Little (2002), las diferentes versiones del Portafolio de la lengua europeo (PLE) tienen dos cosas en común: todas siguen los principios y guías del Consejo de Europa, y todas tienen en cuenta para la evaluación los niveles[20] de referencia del *Marco Común Europeo*. Afirma que el PLE, además cumplir una función pedagógica, que se deriva de la transparencia con que aparece ante el aprendiz su proceso de aprendizaje, permitiéndole desarrollar la capacidad de reflexión y de autoevaluación, también sirve para proporcionar información sobre la experiencia de aprender y usar una lengua, al tiempo que da evidencia concreta de los logros en el aprendizaje de la lengua. Sostiene Little (2002: 5) que "la autoevaluación es fundamental para el uso efectivo de la PLE, así como para los procesos de planificación, monitorización y evaluación del aprendizaje". Señala que la autoevaluación se debe basar en la capacidad del aprendiz para reflexionar sobre su propio aprendizaje, sus habilidades y sus logros: la reflexión es imprescindible si se pretende que el aprendiz llegue a alcanzar una capacidad de autoevaluación fiable. No obstante, la autoevaluación no debe excluir la evaluación externa realizada por otros: de hecho, se deben apoyar la una a la otra y uno de los objetivos de la PLE es encontrar el equilibrio entre ambas.

Como se recogía anteriormente, la entrevista es uno de los procedimientos más extendidos para la evaluación de la expresión oral. El origen de esta técnica se remonta a los años setenta, cuando se adoptaron definitivamente pruebas directas para la lengua hablada.

Oller (1979: 320-326) ya se refiere a la *Foreign Service Institute Oral Interview*: posiblemente uno de los primeros intentos de aplicar esta técnica a exámenes de gran escala y que ha servido de inspiración a entrevistas posteriores. En la mecánica de esta entrevista el candidato se enfrenta a dos evaluadores y todo el proceso, que dura unos treinta minutos —excepto cuando el candidato demuestra poca habilidad desde el principio, caso en el que se reduce a quince—, se graba para contar con un documento permanente que sirva tanto de referencia en el futuro, como para su utilización en estudios de validez u otros

[20] El *Marco de referencia común europeo para las lenguas* (CONSEJO DE EUROPA, 2001) propone los siguientes niveles: A1 Acceso y A2 Plataforma, que corresponden al usuario básico; B1 Umbral y B2 Avanzado, que corresponden al usuario independiente; C1 Dominio operativo eficaz y C2 Maestría, que corresponden al usuario competente.

propósitos como la formación de evaluadores. El nivel de dominio de lengua hablada de los candidatos se determina de acuerdo con una escala de cinco puntos. Oller señala que esta escala refleja la influencia de la teoría de elementos discretos (*discrete point theory*), porque la habilidad de los candidatos se califica separadamente en escalas que se refieren al acento, la gramática, el vocabulario, la fluidez y la comprensión —como vemos, componentes relacionados sólo con la competencia lingüística—. Cada una de ellas se puntúa de manera diferente: la que tiene más peso es la gramática, después el vocabulario y siguen la comprensión, la fluidez y el acento que es a lo que se le da menor puntuación. La orientación hacia el modelo de lengua comunicativo supuso el replanteamiento de la entrevista para la evaluación de la expresión oral, sobre todo en lo que se refiere a su calificación por medio de escalas. Éstas ya no se definen en términos de elementos discretos, sino de habilidad lingüística, habilidad que, dicho sea de paso, se puede calificar de manera analítica —se definen niveles de actuación en términos de competencia gramatical, pragmática, etc.—, u holística —se especifican niveles de actuación de forma global—. Todos los exámenes a gran escala que se adoptan hoy día incluyen una variedad de pruebas orales, revistiendo una de ellas, generalmente, la forma de entrevista, entendida ésta como un intercambio cara a cara examinador-examinando. No obstante, pueden diferir en cuestiones de procedimiento: duración de la prueba, número de entrevistadores, grabación o no de la conversación, y otros pormenores.

Hoy día, una de las entrevistas más ampliamente usadas es la *Oral Proficiency Interview* (OPI) desarrollada por el *American Council for the Teaching of Foreign Languages* (ACTFL). El objetivo de esta entrevista es estimular una conversación de manera eficaz, de manera que se obtenga una muestra de la lengua del candidato a la que se le pueda asignar una calificación.

Según recoge el manual para la formación de examinadores (Swender, 1999), la OPI es un procedimiento estandarizado para la evaluación global de la destreza de producción oral, que mide la producción lingüística de manera holística indicando los fallos y los éxitos. Es decir, a través de la OPI se establece el nivel de dominio para la expresión oral de una persona, en lo que respecta a la habilidad para usar la lengua de manera efectiva y apropiada en situaciones de vida real, en relación con una serie de tareas lingüísticas determinadas, y no en comparación con la actuación de otra persona. La OPI es un tipo de examen de nivel de dominio, esto es: no tiene como objetivo averiguar cómo el candidato ha adquirido la segunda lengua, ni determinar el aprovechamiento del aprendiz dentro de un currículo de enseñanza de la misma. Su función es averiguar el nivel de habilidad oral del candidato en función de lo que puede y sabe hacer con ella. En la escala de niveles de ACTFL para la OPI, se especifican cuatro macroniveles que de menor habilidad a mayor son: 1 novato, 2 intermedio, 3 avanzado, 4 superior, cada uno de ellos se define en función de cuatro categorías:

1. Tareas globales o funciones que realizan con la lengua: preguntar y contestar preguntas simples, narrar, describir, etcétera.
2. Contextos o áreas de contenido: la serie de circunstancias, lingüísticas o situacionales en las que se desarrollan las tareas y los temas que se relacionan con ellas. El contexto puede ser "en la calle de una ciudad española" y el contenido "preguntar una dirección".
3. El rigor con el que se realiza la tarea: uso del vocabulario, corrección formal, fluidez, pronunciación, adecuación sociolingüística, aceptabilidad, utilización de estrategias idóneas para el manejo del discurso.
4. El tipo de texto oral que resulta de la actuación de las tareas: palabras sueltas, frases, oraciones, párrafos o discurso extenso.

Así las cosas, el nivel novato se define de esta manera: "El candidato puede comunicarse mínimamente con emisiones formulaicas y aprendidas de memoria, o con listas y frases". El nivel intermedio está descrito como: " El candidato puede crear con la lengua, hacer y contestar preguntas sencillas sobre temas corrientes y manejar una situación o una transacción sencilla". El avanzado se caracteriza así: "El candidato puede narrar y describir en todos los tiempos verbales y manejar una situación con complicaciones". Y el nivel superior especifica: "El candidato puede fundamentar una opinión, hacer hipótesis, hablar sobre una variedad de temas de manera concreta y abstracta y manejar una situación lingüística poco corriente".

Para conseguir una muestra efectiva de lengua, también habrá que tener en cuenta el tipo de pregunta, el comportamiento y las estrategias del examinador, los temas elegidos y su desarrollo, así como la situación ficticia seleccionada.

Esta entrevista, ampliamente utilizada en EE UU y que existe en varias lenguas, entre ellas el español, ha sido criticada en lo que respecta a su validez y fiabilidad[21]. No obstante, es un instrumento muy eficaz para asignar niveles de habilidad a grupos numerosos y, a pesar de que se cuestione su fiabilidad, proporciona resultados muy estables. Esto es así, porque los evaluadores que califican la OPI son sometidos a un riguroso entrenamiento para la obtención del diploma que los certifica como capaces de llevar a cabo la entrevista de manera adecuada y calificarla con justicia. Y no sólo eso, el certificado tiene una duración limitada, por lo que si el evaluador desea poder seguir ejerciendo como tal, debe someterse periódicamente a un proceso de reciclaje para asegurar que sigue manteniendo el estándar que evita que se emitan calificaciones "subjetivas".

---

[21] Ver el artículo de SALABERRY (2000) en el que revisa la nueva versión de la OPI (SWENDER, 1999) y propone algunos cambios estructurales que sin afectar la operatividad de la entrevista contribuirían a mejorar la validez y la fiabilidad del procedimiento.

En la actualidad, todos los exámenes a gran escala, sirvan de ejemplo los DELE, constan de una o más pruebas para determinar el nivel de lengua hablada de los candidatos. Y, en general, suele incluirse algún tipo de entrevista, entendiendo ésta como un intercambio cara a cara entre el examinador o examinadores y el candidato. En cualquier caso, es frecuente la utilización de otros procedimientos[22], además de la entrevista, para evaluar la habilidad de la expresión oral.

4.4. El reto de la ética

Otra área que, en opinión de Bachman (2000), ha conocido un notable impulso en los años noventa es la que aglutina asuntos relacionados con la ética, las consecuencias del uso de los exámenes y la profesionalización del campo de la evaluación. Muestra del interés suscitado por los aspectos éticos de la evaluación en estos años, es que el *Language Test Reseach Colloquium*, celebrado en 1997, tuvo como lema *Fairness in Language Testing* (Equidad en los exámenes de lengua) y la prestigiosa publicación *Language Testing* dedicó ese mismo año un número especial[23] a la ética en los exámenes de lengua. No obstante, la reflexión sobre este tema no es totalmente nueva, ya que, en la década anterior, Spolsky (1981) había manifestado su preocupación por las cuestiones éticas en el campo de la evaluación. En los años noventa, este autor vuelve a tratar sobre ello (Spolsky, 1997), así como otras figuras destacadas (Shohamy, 1993 y 1997; Lynch, 1997), que abordan el asunto de la ética en los exámenes de lengua desde la metodología de la teoría de la crítica social.

Sin embargo, Bachman piensa que dicha teoría, tal y como ha sido aplicada, no proporciona soluciones. En su opinión, esta cuestión conviene abordarla desde el campo de la evaluación y afirma (Bachman, 1998: 17) que el mal uso que a veces se hace de los exámenes de lengua podría evitarse promoviendo la profesionalización de los evaluadores y contando con un código deontológico y unos estándares en la línea de lo que Stansfield (1993), Davies (1997) y otros han defendido.

Algunas de estas propuestas —y en ciertos casos, todas— son una realidad para muchos exámenes. De hecho, la *Association of Language Testers of Europe* (ALTE)[24] decidió en 1994 adoptar un código de práctica que proporcionara los estándares que, tanto los miembros actuales de la asociación co-

---

[22] Bordón (2000).

[23] *Language Testing*, n.º 14, 3, 1997. Editado por Alan Davies.

[24] Se puede consultar la dirección: www.alte.org para información más detallada en inglés sobre esta organización.

mo los que se incorporaran en el futuro, aceptaran asumir para redactar sus exámenes y que al mismo tiempo, también informara a los posibles usuarios de dichos exámenes de lo que se esperaba de ellos. El *Educational Testing Service*[25], el Instituto Cervantes[26] y otras instituciones que diseñan y administran exámenes a gran escala publican información en la que el posible usuario del examen puede encontrar datos acerca de las características del mismo —nivel que define, partes, tipo de pruebas, duración, etc. y consultar ejemplos—.

McNamara (2000) ofrece una buena perspectiva de la cuestión ética en relación con la evaluación de la lengua. Distingue entre los que consideran que la evaluación de la lengua es simplemente una actividad científica y técnica y los que defienden una postura de responsabilidad social. Esta segunda postura considera que la práctica de la evaluación ética supone que los autores de exámenes asuman su responsabilidad en los efectos de los mismos. Se destacan tres aspectos principales que deben tenerse en cuenta. En primer lugar, la necesidad de poder rendir cuentas (*accountability*): es decir los responsables de los exámenes tienen una obligación (una responsabilidad) hacia la gente que se verá afectada por el examen: de entrada, los usuarios del examen en cuestión, pero también quienes usen la información extraída de los resultados del examen. De ahí la necesidad de proporcionar información acerca del examen a los posibles candidatos mediante la publicación de manuales, folletos informativos o páginas en Internet. El segundo aspecto que se debe considerar es la repercusión (*washback effect*) del examen: entendida ésta como la influencia que el examen tiene en la enseñanza-aprendizaje. De esta manera, una práctica ética deberá asegurar que la repercusión del examen sea positiva. En tercer lugar, se habla del impacto[27] del examen, entendiéndose por ello el efecto de los exámenes en la comunidad. Por ejemplo, los exámenes que llevan a la obtención de un DELE tienen un mayor "impacto" en la sociedad que un examen de aprovechamiento realizado en una aula a un pequeño grupo de alumnos, después de un período limitado de instrucción. Una evaluación ética debe tener en cuenta este aspecto.

Por otra parte está la postura de quienes consideran que la responsabilidad social de los examinadores debe limitarse a cuestiones relacionadas con la práctica ética de su profesión, de tal manera que el tratamiento ético de la práctica de la evaluación de la lengua deberá ser el mismo que el de cual-

---

[25] En la dirección www.ets.org se puede encontrar información a cerca de ETS, así como consultar la publicación de sus estándares, y ver ejemplos de exámenes.

[26] En la página del Instituto Cervantes: www.cervantes.es, se incluyen vínculos que dan acceso a las características de los distintos diplomas DELE y se pueden obtener ejemplos de exámenes antiguos.

[27] Wall (1997) afirma que una ampliación de la noción de repercusión (*washback*) es la de impacto, cuyo objetivo es profundizar en la relación entre el uso del examen y la sociedad en la que se usa.

quier práctica profesional, como la medicina o la abogacía. Desde esta perspectiva, se hará hincapié en aspectos como la profesionalización de los evaluadores y la publicación de un código deontológico para el ejercicio de la profesión.

De hecho, Stansfield (1993:190) une profesionalización a ética, y se refiere a ésta como "la conducta moral de los examinadores de lengua como gente que ejerce su profesión". Este autor defiende que un código de ética profesional, junto con estándares para desarrollar exámenes de lengua y operar programas de evaluación de la lengua, proporcionarán una guía para que los examinadores de lengua puedan enfrentarse con éxito a la cuestión ética. Davies (1997) incide en que la ética en la evaluación se basa en la formación de profesionales y también en la existencia y adopción de un código ético. Para él, la ética es parte del proceso de convertir la evaluación en una profesión. La profesionalización tiene que ver con códigos, contratos, entrenamiento y estándares de práctica.

La década de los noventa ha supuesto, por lo tanto, la consolidación total del área de la evaluación de la lengua, y a lo largo de ella se ha experimentado un notable desarrollo tanto en las investigaciones teóricas como en el tratamiento de las cuestiones prácticas. Temas de siempre, como la reflexión acerca de la naturaleza de la lengua, la validez y la fiabilidad, se han estudiado desde nuevas perspectivas. Las nuevas tecnologías han revelado su utilidad tanto para la investigación como para la administración y corrección de pruebas por ordenador. Se ha conseguido que haya más y mejores expertos en el área, como demuestra el aumento de las publicaciones de todo tipo dedicadas a aspectos generales y concretos de la evaluación de la lengua, el aumento de los congresos y las conferencias o el incremento de las asociaciones profesionales. En la mayoría de los países existen instituciones públicas o privadas (o ambas) que se dedican a la evaluación de la lengua, elaborando y administrando exámenes a gran escala, realizando formación de evaluadores, redactando estándares y definiendo códigos de ética.

5. El siglo XXI, ¿qué nos espera?

En el año 2000, Bachman enuncia los asuntos que considera que van a preocupar y a estudiarse en el campo de la evaluación de la lengua en el siglo XXI. En primer lugar, señala que la evaluación de la lengua como tal se debe convertir en una profesión, lo cual implica la formación de evaluadores y la definición de estándares de práctica. Señala este autor la importancia de establecer el alcance de la validez y la equidad (*fairness*), ya que ambas se encuentran en el centro de lo que hace que nos definamos como profesionales no sólo como evaluadores de la lengua, sino también como lingüistas aplicados.

A su modo de ver, la manera en la que se desarrolle la investigación de estos asuntos y las respuestas que se den definirán el paradigma para la evaluación de la lengua en los próximos veinte años. También dice que a consecuencia de la mayor sensibilidad hacia cuestiones de tipo ético y la diversidad y complejidad de los procedimientos para la investigación en este siglo que estamos iniciando, se puede trabajar con algo más que los resultados de los exámenes, pudiendo ir más allá de la especulación acerca de su significado y de su uso.

6. Conclusión

De todo lo anterior, se deduce que la evaluación de la lengua es un campo que se ha desarrollado notablemente en los últimos treinta años, ocupando en la actualidad un lugar destacado en el ámbito de la lingüística aplicada. Tratando de sintetizar al máximo las áreas de interés en esta área, podemos identificar unos asuntos fundamentales —y limitados— que encontramos de manera recurrente, ya que son inherentes a la evaluación y a lengua, pero que se pueden tratar desde muchos ángulos y que constituyen los ejes en torno a los cual se articulan todos los demás temas y preocupaciones. Uno de ellos es la necesidad de determinar la naturaleza de la lengua y contar con un modelo para su evaluación, y los otros, los requisitos básicos, garantía de cualquier instrumento evaluador: la validez y la fiabilidad.

Concretando un poco más, observamos que los asuntos que parece que han requerido mayor atención por parte de los investigadores y de otros expertos en evaluación se centran en torno a la búsqueda de un modelo acerca de la naturaleza de la lengua, la evaluación de la expresión oral, la búsqueda de alternativas a los exámenes para obtener información acerca de los procesos de aprendizaje de la segunda lengua, la aplicación del desarrollo de la tecnología informática al diseño y calificación de exámenes, el diseño de escalas de nivel de dominio y la profesionalización de los evaluadores. Y la preocupación constante por la fiabilidad y la validez.

Cualquier reflexión teórica, preocupación o práctica relacionada con la evaluación se engarzará en torno a uno de estos ejes que a su vez se conectan entre sí. El modelo de lengua es la sustancia necesaria para que existan la validez y la fiabilidad. Y como una tela de araña que cubre todo ello, pero que a su vez, nace de todo ello, tenemos la ética. La ética garantiza la justicia, la equidad del instrumento evaluador. Un examen de lengua no es ético si no hay una concepción de la lengua, un examen no es ético si no mide lo que se supone que pretende medir, un examen no es ético si sus resultados no son consistentes. De ahí que tal vez el reto que la evaluación de la lengua tiene por delante sea el de elaborar exámenes, o tal vez otros procedimientos alternativos, cuyo objetivo sea el de ser justos.

7. Bibliografía

ALDERSON, CH. (1986), "Computers in language testing", en G. LEECH y CH. CANDLIN (Eds.), *Computers in English Language Learning and Research*, Harlow, Longman. pp. 99-110.

ALDERSON, CH. (1991), "Bands and Scores", en CH. ALDERSON y B. NORTH (Eds.), *Language Testing in the 1990s,* pp. 71-86.

ALDERSON, CH. y NORTH, B. (Eds.) (1991), *Language Testing in the 1990s*, Londres, Modern English Publications y The British Council.

BACHMAN, L. (1990), *Fundamental Considerations in Language Testing*, Oxford, Oxford University Press.

BACHMAN, L. y PALMER, A. (1996), *Language Testing in Practice*, Oxford, Oxford University Press.

BACHMAN, L. F. (1998), "Language Testing at the Turn of the Century", *American Association for Applied Linguistics Letter*, 20, 2.

BACHMAN, L. F. (2000), "Modern language testing at the turn of the century: assuring that what count counts", *Language Testing*, 17, 1, pp. 1-42.

BORDÓN, T. (2000), "La evaluación de la expresión oral en el aula de E/LE", *Carabela*, 47, Madrid, pp. 151-175.

BROWN, J. D. y HUDSON, T. (1998), "The Alternatives in Language Assessment", *TESOL Quarterly*, 32, 4, pp. 653-675.

CANALE, M. (1983), "From communicative competence to communicative language pedagogy", en J. RICHARDS y R. SCHMIDT (Comps.), *Language and Communication*, Londres, Longman, pp. 2-27.

CANALE, M. y SWAIN, M. (1980), "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing", *Applied Linguistics*, 1, 1, pp. 1-47.

CARROLL, B.(1991), "Resistance to change", en CH. ALDERSON y B. NORTH (Eds.), *Language Testing in the 1990s,* pp. 22-27.

CONSEJO DE EUROPA (2001), *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Traducción en español realizada por la Dirección Académica del Instituto Cervantes y publicada de acuerdo con el Consejo de Europa (2002), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid, Ministerio de Educación, Cultura y Deporte, Subdirección General de Cooperación Internacional, y Grupo Anaya.

CHALHOUB-DEVILLE, M. (1996), "Performance assessment and the components of the oral construct across different tests and rater groups", en M. MILANOVIC y N. SAVILLE (Eds.), *Performance testing, cognition and assessment*, Cambridge, University of Cambridge, Local Examinations Syndicate y Cambridge University Press, pp. 55-73.

DAVIES, A. (1985), "Follow my leader: Is That What Language Tests Do?", en Y. P. LEE y otros (Eds.), *New Directions in Language Testing*, Oxford, Pergamon, pp. 3-13

DAVIES, A. (1991), "Language Testing in the 1990s", en CH. ALDERSON y B. NORTH (Eds.), *Language Testing in the 1990s,* pp. 136-149.

DAVIES, A. (1997), "Demands on being professional in language testing", *Language Testing,* 14, 3, pp. 328-339.

KUNNAN, A. J. (1995), *Test taker characteristics and test performance: a structural modelling approach*, Cambridge, University of Cambridge Local Examinations Syndicate y Cambridge University Press.

LADO, R. (1961), *Language Testing: The Construction and Use of Foreign Language Tests*, Londres, Longman.

LEE, Y. P., FOK, A. C. Y. Y., LORD, R. y LOW, G. (1985), *New Directions in Language Testing*, Oxford, Pergamon.

LITTLE, D. (2002), "The European Language Portfolio and learner autonomy", *Málfridur,* 18, 2, pp. 4-7.

LYNCH, B. K. (1997), "In search of the ethical test", *Language Testing,* 14, pp. 315-327.

MCNAMARA, T. (1996), *Measuring second language performance*, Londres, Longman.

MCNAMARA, T. (2000), *Language Testing,* Oxford, Oxford University Press.

OLLER, J. JR. (1979), *Language Tests at School*, Harlow, Longman.

PURPURA, J. E. (1999), *Modeling the relationship between test takers´reported cognitive and metacognitive strategy use and performance on language tests*, Cambridge, University of Cambridge Local Examinations Syndicate y Cambridge University Press.

REA, P. (1985), "Language Testing and the Communicative Language Teaching Curriculum", en Y. P. LEE y otros (Eds.), *New Directions in Language Testing*, Oxford, Pergamon, pp. 15-32.

REA, P. (1991), "What Makes a Grammar Test Communicative?", en CH. ALDERSON y B. NORTH (Eds.), *Language Testing in the 1990s,* pp. 112-131.

SALABERRY, R. (2000), "Revising the revised format of the ACTFL Oral Proficiency Interview", *Language testing*, 17, 3, pp. 289-310.

SHOHAMY, E. (1997), "Testing methods, testing consequences: are they ethical?", *Language Testing,* 14, pp. 340-349.

SHOHAMY, E. (1993), "The power of tests: the impact of language tests on teaching and learning", *NFLC Occasional Papers*, Washington D.C., National Foreign Language Center.

SKEHAN, P. (1991), "Progress in Language Testing: the 1990s", en CH. ALDERSON y B. NORTH (Eds.), *Language Testing in the 1990s,* pp. 3-6.

SLAGTER, P. (1979), *Nivel Umbral*, Estrasburgo, Consejo de Europa.

SPOLSKY, B. (1981), "Some ethical questions about language testing", en C. KLEIN-BRALEY y D. K. STEVENSON (Eds.), *Practice and problems in language testing*, Frankfurt, Peter Lang, pp. 5-21.

SPOLSKY, B. (1997), "The ethics of gatekeeping tests: what have we learned in a hundred years?", *Language Testing,* 14, pp. 242-247.

STANSFIELD, C. W. (1993), "Ethics, standards and professionalism in language testing", *Issues in Applied Linguistics,* 4, 2, pp. 15-30.

STEVENSON, D. (1985), "Pop Validity and Performance Testing", en Y. P. LEE y otros (Eds.), *New directions in Language Testing*, Oxford, Pergamon, pp. 111-118.

SWAIN, M. (1985), "Large-scale Communicative Language testing: A Case Study", en Y. P. LEE y otros (Eds.), *New Directions in Language Testing*, Oxford, Pergamon, pp. 35-46

SWENDER, E. (Ed.) (1999), ACTFL *Oral Proficiency Interview Tester Training Manual*, Yonkers (NY), The American Council for the Teaching of Foreign Languages.

WALL, D. (1997), "Impact and washback in language testing", en C. CLAPHAM y D. CORSON (Eds.), *Encyclopedia of language and education*, 7, *Language Testing and Assessment*, Dordrecht, Kluwer Academic.

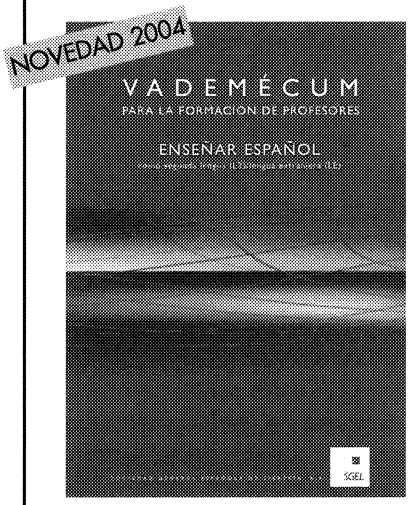
WEIR, C. y MILANOVIC, M. (Eds) (2002), *Innovation and Continuity: Revising the Cambridge Proficiency Examination,* Cambridge, Cambridge University Press.

WU, Y. (1998), "What do tests of listening comprehension test? A retrospection study of EFL test-takers performing a multiple-choice task", *Language Testing,* 15, 1, pp. 21-44.

V A D E M É C U M

PARA LA FORMACIÓN DE PROFESORES

ENSEÑAR ESPAÑOL COMO SEGUNDA LENGUA/LENGUA EXTRANJERA



SGEL presenta en un solo libro el conocimiento más actual sobre la enseñanza-aprendizaje de español como segunda lengua/lengua extranjera.

SGEL, por vez primera, reúne a los mayores especialistas de la enseñanza del español en una obra revolucionaria en su género, en un manual de referencia imprescindible en esta materia, en un libro que nace con vocación de clásico.

Obra colectiva y plural, rigurosa y alejada de planteamientos dogmáticos.

61 artículos • 1.312 páginas • 58 autores

Abello Contesse, Christian
Acquaroni Muñoz, Rosana
Aguirre Beltrán, Blanca
Alonso Belmonte, Isabel
Baralo Ottonello, Marta
Bernárdez Sanchís, Enrique
Bordón Martínez, Teresa
Briz Gómez, Antonio
Cassany i Comas, Daniel
Cenoz Iragui, Jasone
Cestero Mancera, Ana María
Díaz-Corralejo Conde, Joaquín
Eguiluz Pacheco, Juan
Eguiluz Pacheco, Ángel
Escandell Vidal, M.ª Victoria
Fernández López, M.ª del Carmen
Fernández López, Sonsoles
García Parejo, Isabel
Gil-Toresano Berges, Manuela
Gómez del Estal Villarino, Mario
Gómez Molina, José Ramón
González Landa, Carmen
Guillén Díaz, Carmen
Gutiérrez Ordóñez, Salvador
Gutiérrez Rivilla, Rebeca
Hernández García, M.ª Teresa
Higueras García, Marta
Juan Lázaro, Olga
López García, Ángel
López Morales, Humberto
Lorenzo Bergillos, Francisco José
Marcos Marín, Francisco Adolfo
Martín Martín, José Miguel
Martín Peris, Ernesto
Matte Bon, Francisco
Mayor Sánchez, Juan
Melero Abadía, Pilar
Miquel López, Lourdes
Morales Pastor, Jorge
Moreno Fernández, Francisco
Moreno García, Concha
Muñoz López, Belén
Olmo Pintado del, Margarita
Parrondo Rodríguez, José Ramón
Perales Ugarte, Josu
Pinilla Gómez, Raquel
Pisonero del Amo, Isidoro
Poch Olivé, Dolors
Rojo, Guillermo
Ruipérez García, Germán
Sánchez Lobato, Jesús
Sánchez Pérez, Aquilino
Santos Gargallo, Isabel
Tomás Puch de, José María
Tuts, Martina
Vázquez, Graciela E.
Vez Jeremías, José Manuel
Villalba Martínez, Félix

SGEL SOCIEDAD GENERAL ESPAÑOLA DE LIBRERÍA, S.A.