

ALFONSO MARTÍNEZ BATZÁN

Doctor en Filología Española por la Universidad de Granada, profesor de ELE en el CLM de la Universidad de Granada y miembro de ACTFL. Da cursos en el Máster de ELE de la Universidad de Granada y en los cursos de formación y acreditación para evaluadores del DELE del Instituto Cervantes. Ha trabajado en la Universidad de Utrecht y Willamette University (Oregon). Tiene diversas publicaciones sobre metodología, evaluación y español para inmigrantes.

Resumen:

Este artículo pretende abordar la problemática de la evaluación oral, tanto uno de los sistemas más efectivos para asegurarse del dominio de una lengua extranjera como uno de los más subjetivos, difíciles o complejos de realizar. Al mismo tiempo analiza uno de los aspectos que últimamente se han mostrado más importantes y que promete redefinir numerosos elementos de este proceso: el papel del entrevistador en la génesis del discurso y la concepción de la evaluación oral, y de la conversación, como resultado de la colaboración entre todas las partes que intervienen en ello.

THE ORAL LANGUAGE EVALUATION DIFFICULTIES AND PRECAUTIONS

The aim of this article is to look into the problematic system of the oral evaluation and its difficulties, as it is, on one hand, one of the most effective ways to evaluate for certain the dominion of the foreign language, and on the other hand, it is one of the most subjective, difficult and complex system to be undertaken.

At the same time, the oral evaluation analyses one of the aspects that lately has been proved to be the most important and that it would help to redefine numerous elements and stages of its own process: the interviewer's role within the genesis of the speech and the concept of the oral evaluation, and the conversation as a result of the collaboration between all parts involved.

PALABRAS CLAVE: evaluación, oral, entrevistador, interacción, conversación

KEYWORDS: evaluation, oral, interviewer, interaction, conversation.

Cuando pensamos en evaluación oral tenemos, sin duda, la imagen de un profesor sentado en una silla, hablando con un estudiante, y unas hojas con materiales como fotos, recortes de periódicos o cómics, para servir de estímulo en determinados temas. También se piensa, sin duda por propia experiencia, en el estrés que ello puede provocar en el estudiante, nervioso por tener que hablar ante un desconocido, en una lengua extranjera. Y, visto desde fuera, escuchado más bien, el asunto puede parecer tranquilo, interesante e incluso sencillo. Sin embargo, esto que vemos es sólo apariencia y, desde luego, es solamente la parte más externa de todo el proceso.

Evaluar la lengua oral es una de las cosas más difíciles de la evaluación de lenguas, puesto que los tests orales son los más difíciles de aplicar, los más costosos en medios humanos, los que precisan más formación para las personas encargadas de ponerlos en práctica, y los que más tiempo de realización llevan. Pueden parecer sencillos, ya que se trata simplemente de hablar con los candidatos, pero en realidad esta sencillez no es tal. Se hacen con el mismo tipo de planteamientos previos que el resto de pruebas de lengua, delimitando las características de los candidatos para los que el

modelo de examen pueda ser útil y realista y estableciendo claramente los protocolos de evaluación que se vayan a utilizar y los criterios de corrección y calificación de los resultados. Pero, además, requieren unos cuantos elementos específicos, tanto internos como externos; algunos de los cuales hacen de la evaluación oral algo específico y diferente a los demás tipos de tests, como la necesidad de dedicar gran cantidad de tiempo y trabajo a estos tres aspectos:

- La definición de la habilidad a evaluar, la competencia comunicativa o la habilidad lingüística.
- La formación de los evaluadores, ya que suelen condensar en una misma persona las figuras de entrevistador, corrector y calificador.
- La puesta en escena y administración de los tests, que necesita más tiempo y más evaluadores, puesto que a cada candidato hay que hacerle la entrevista personalmente.

Una de las principales características de un test oral, es que este debe ser capaz de generar en los candidatos el tipo de lengua que se desea medir. Un test oral trata de ser un mecanismo activo cuyo objetivo es conseguir que los candidatos produzcan un tipo de lengua con el que demuestren la existencia de determinadas habilidades lingüísticas. Para ello, el evaluador y las tareas que propone llevan el peso principal. Pero como un test por sí solo –el que sea- no puede provocar ni controlar todos los factores que intervienen en la actuación oral de una persona, en su diseño hay que eliminar o minimizar los efectos de los factores que no son deseables y no interesen en la evaluación, como por ejemplo la cantidad de conocimiento que tengan sobre un tema o la educación o personalidad, y maximizar los elementos de la habilidad que se deseen medir y hayan sido determinados previamente en los objetivos y especificaciones del test así como en sus escalas de evaluación y calificación.

Como dice Bachman “la capacidad de hablar una lengua extranjera es tanto la más importante como la más problemática de las habilidades que intentamos medir” (Bachman, 1988: 149).

Tendemos a creer que una nota obtenida en un test preparado para evaluar una habilidad es directamente una indicación de esa habilidad. Pero esto no es directamente cierto. Sólo será cierto si del test, del sistema determinado para obtener los datos (tareas, actividades, método, etc.) y de los sistemas de calificación determinados al efecto, se pueden deducir interpretaciones válidas sobre el uso externo de la lengua. Por lo tanto, hay que asegurarse de que la nota mide el uso real de la lengua y no una habilidad extra, la de hacer bien el test³⁶.

³⁶ Así, por un lado está la validación en sí del test y, por otro, la de las decisiones que se toman con sus resultados (Bachman, 2004: 257). Este planteamiento sobre la legitimidad del evaluador para tomar decisiones abre, según Leung / Lewkowicz (2000: 223) otras importantes cuestiones de ética profesional y uso democrático de test de las que no nos vamos a ocupar ahora.

La capacidad de un test para obtener y ofrecer datos fiables no es una consecuencia directa del test sino de su validez, fiabilidad y autenticidad³⁷. Es de suma importancia conseguir que todos los elementos que intervienen en su diseño y desarrollo e interfieren en sus resultados lo hagan de forma positiva, sirviendo para facilitar la actuación de los candidatos y obtener la mejor muestra de sus habilidades lingüísticas. No se puede hacer un test y suponer sin más que sus resultados muestran un conocimiento o predicen las habilidades necesarias para actuaciones futuras de los candidatos. Antes es necesario demostrar que el test sirve y funciona y que con sus resultados se pueden tomar las decisiones que se pretende. Y para ello hay que establecer una relación entre ambas cosas: el test y las interpretaciones de sus resultados (Bachman, 2004: 257).

En este artículo vamos a tratar algunos de estos aspectos relacionados con la evaluación oral y sus dificultades.

1. LA NECESIDAD DE DEFINIR UN CONCEPTO DE LENGUA EVALUABLE

Para evaluar oralmente la lengua es necesario partir de un concepto de lengua, ya que esta –compleja y multidimensional- puede verse e interpretarse desde múltiples puntos de vista y parecer un objeto diferente desde cada uno de ellos. Esta variabilidad de la lengua no es exclusiva de los extranjeros sino que pertenece igualmente a los sistemas nativos. Así, no es lo mismo para un periodista que para un médico, para un abogado que para un profesor de instituto, para un psicoanalista o para un profesor de ELE.

Y para evaluarla no basta con formarse una opinión sobre cada candidato basada en el conocimiento directo o indirecto que se tenga de él, por un trato personal, del aula, ni a través de una muestra obtenida mediante una charla que no sabemos a dónde puede llevarnos, qué resultados nos puede ofrecer o cómo interpretarlos. Evaluar la lengua oralmente requiere de un examen específico que sea “una herramienta válida y fiable, que nos permita tomar decisiones acertadas (mediante) unos procedimientos de evaluación que sirvan para medir unos determinados conocimientos a través de las muestras que proporciona la actuación de los candidatos” (Bordón, 2006: 19). Y este examen específico “debe partir de una definición operativa del habla con la que se puedan desarrollar tests. Nunca el constructo (del test) va a captar la riqueza de la comunicación humana pero al menos el test sí debe ser capaz de medir todo lo que se haya definido en el constructo” (Fulcher, 2003: 19).

Todo esto exige que el test oral se defina con cuidado desde el comienzo (el momento del planteamiento de los objetivos) hasta el final (la manera de hacer público su resultado ponderando las calificaciones de que se componga y ofreciéndolas de forma inteligible y útil). Y antes de usar sus resultados debemos estar seguros de que la forma

³⁷ Estos conceptos de validez, fiabilidad y autenticidad son fundamentales en la evaluación de lenguas o en las ciencias sociales. Por ahora, para efectos de evaluación de lenguas los aclaramos diciendo que el primero (validez) es asegurarse de que un test mide lo que efectivamente pretende; el segundo (fiabilidad) es asegurarse de que lo mide sistemáticamente y de que si hay variaciones en sus resultados se deben a un aumento de las habilidades que mide y no a otros factores; y el tercero (autenticidad), que lo que mide es efectivamente la lengua real y no una lengua artificial producida por el test.

en que los interpretamos y usamos es válida, pues de lo contrario corremos el riesgo de hacer interpretaciones incorrectas o tomar decisiones que pueden dañar a nuestros candidatos en vez de ayudarles o beneficiarles (Bachman, 2004: 3)³⁸. La responsabilidad es grande y las consecuencias de una prueba (lo que se llama "impacto" -en el ámbito social- o "efecto de rebote" -en el individual o académico), aunque pueden pasar desapercibidas, pueden llegar a tener importantes efectos sociales e incluso legales. Y, desde luego, personales.

Debido a todas estas dificultades a lo largo del proceso de diseño y puesta en práctica de un test oral, sus prácticas más usuales adolecen de grandes carencias que abarcan desde el planteamiento de su diseño y llegan hasta la formación de evaluadores, pasando por un gran desconocimiento tanto de las técnicas necesarias para ponerlo en funcionamiento como de las limitaciones de cada formato. Estamos de acuerdo con Ortega cuando dice "¿Qué metas, contextos y situaciones elegir? La respuesta a esta pregunta no ha recibido aún forma definitiva" (Ortega, 1996: 32). Tal es así que estamos todavía lejos de poder afirmar que se evalúa oralmente la lengua con la objetividad que se merece y la efectividad que la sociedad requiere, o de poder hacerlo con total equidad y honestidad, igual que todavía estamos lejos de poder evaluar con suficiente fiabilidad otros aspectos como la pragmática (Walters, 2007: 176).

El problema de la definición de la habilidad lingüística ha sido tratado por diferentes teóricos de la lengua y la evaluación (Chalhoub-Deville, 1997) aportando diferentes modelos que intentan incluir todos los aspectos que intervienen en el uso de la lengua, la competencia lingüística, la estratégica y la pragmática o sociolingüística, pero que no ha dado resultados todo lo efectivos que se necesitan. La tensión entre la competencia lingüística y la habilidad comunicativa continúa y los sistemas desarrollados para medirlas -el último, el de Bachman / Palmer (1996)- siguen presentando problemas que necesitan más investigación.

Muchos han sido los intentos de delimitar la lengua a evaluar³⁹ y de establecer diferentes tipos de escalas de medición. El último, la descripción de la lengua que hace el *MCER*, que con sus 56 escalas pretende dar una completa visión de la lengua apta para todo tipo de usos y evaluaciones. Descripción, según dicen Alderson y Banerjee "del planteamiento del modelo de la lengua, uso y aprendizaje, a menudo basado explícitamente en el modelo de Bachman" (2002: 81) o "en las ideas del mundo de la enseñanza y lingüística aplicada angloamericana. La influencia de Wilkins (1976), Canale / Swain (1980) y Bachman (1990) es particularmente evidente" (Alderson, 2007: 660).

Todo esto tiene su propia historia. Según Fulcher (2003: 5), el primer test oral se hizo en Estados Unidos en los años 20, consecuencia de la legislación de las normas sobre inmigración de 1924. También se debe mirar a los Estados Unidos para encontrar una

³⁸ Alderson / Banerjee admiten que "el hecho de no poder atribuir con exactitud un significado a una nota es el agujero negro de la evaluación" (2002: 101).

³⁹ Como resumen (véase Bordón, 2006: 25-34 o Purpura, 1999: 19-20) la definición de la habilidad lingüística ha dado importantes avances descriptivos durante el siglo XX desde el estructuralismo norteamericano, pasando por las teorías chomskianas sobre el innatismo del lenguaje humano o las aportaciones comunicativas de Hymes, la pragmática y Canale, hasta la propuesta de Bachman. Y añadimos, pasando por el empuje funcional de los trabajos del Consejo de Europa desde los años 50.

de las escalas de evaluación más influyentes, la del FSI (*Foreign Service Institute*) de la que proceden muchos tests⁴⁰ y en especial el de ACTFL, así como para entender las causas de que las escalas de evaluación tengan el tipo de progresión aparentemente lineal que muestran, deudas de las escalas lineales descritas para el ejército en los 50 (Fulcher, 2003: XIII).

Desde entonces, y en especial en los últimos 25 años, muchos han sido los temas sobre los que ha habido controversia y los problemas con los que ha habido que tratar.

Entre ellos (Alderson / Banerjee, 2002: 92-95) citan estos siete:

- La capacidad de un test oral para obtener muestras suficientes que permitan elaborar un juicio sobre las habilidades que evalúa⁴¹.
- La necesidad de continua investigación sobre cuál es la mejor forma de evaluar la lengua oral, cómo se define la competencia, habilidad y capacidad lingüística y cuáles son las mejores formas de elicitarlas.
- El insuperable obstáculo de que los datos se obtienen, incluso en una entrevista oral, siempre de forma indirecta.
- El análisis del efecto del evaluador en la conformación del discurso.
- La capacidad de las tareas para reflejar la habilidad del hablante.
- La posibilidad de que notas diferentes oculten capacidades similares y de que, al contrario, notas iguales oculten capacidades diferentes.
- La posibilidad de que la corrección gramatical tenga una poderosa influencia en la calificación final.

Alderson / Banerjee concluyen su larga lista diciendo que “los continuos trabajos que se están desarrollando en estas áreas indican que todavía estamos lejos de comprender bien el constructo lingüístico. (...) Cada vez hay más descubrimientos procedentes de áreas como la lingüística aplicada, el análisis del discurso y la adquisición de segundas lenguas” (2002: 95).

⁴⁰ Hasta que en Europa se comenzaron a diseñar escalas como las de ALTE o el proyecto suizo que daría paso al *MCER*, la mayoría de escalas de evaluación estaban directa o indirectamente relacionadas con las del Instituto de Asuntos Exteriores norteamericano (FSI en sus siglas inglesas) (Parrondo, 2004: 968).

⁴¹ Esta dificultad es para Widdowson insalvable. Para él “la competencia comunicativa no puede, como tal, ser medida. Quiero decir que los tests comunicativos son en principio imposibles, y eso es por lo que no sorprende que sean tan difíciles de diseñar. (...) No se puede evaluar la habilidad para comunicarse (...) Lo único que se puede enseñar, y evaluar, son aspectos de ella” (2001:18), “uno comienza a tener dudas sobre si *algún* modelo de la competencia comunicativa va a llegar a ser pedagógicamente útil como para servir de marco para la evaluación de la lengua”, (Widdowson, 2003: 170, la cursiva es del original).

A todos estos podríamos añadir las dificultades para validar y obtener fiabilidad en los tests orales que se basan en escalas, criterios y constructos referidos a situaciones y contextos de la vida real, o problemas con los criterios de calificación y con la formación de evaluadores. Con la consecuencia de que cada uno de estos aspectos tiene diferentes implicaciones en el desarrollo de nuevos tests.

Uno de los últimos intentos para resolver algunas de estas dificultades ha sido el *Marco común europeo de referencia para el aprendizaje, la enseñanza y la evaluación de lenguas* (2002). Entre sus objetivos trata de establecer unos criterios comunes para la evaluación de lenguas en diferentes ámbitos y en diversas lenguas, que sean reconocidos internacionalmente de forma que sus calificaciones sean homologables e inteligibles y sus resultados interpretables. En él se hace una descripción de numerosos aspectos del uso de la lengua a través de 56 escalas divididas en seis niveles que pretenden servir de base para definir el test y el constructo de lengua en cada modelo de examen que se relacione con él.

No obstante el MCER no es un test ni pretende serlo o decir cómo debe realizarse un test oral. Carece de protocolos para realización de entrevistas o de contenidos especificados para evaluar la lengua y presenta bastantes dificultades para una adecuada interpretación de los datos cuando se utiliza directamente en evaluación⁴². Bajo su paraguas se pueden desarrollar, debidamente diseñados, numerosos tests y constructos con concepciones y objetivos distintos e incluso dispares. Ello va a dificultar una adecuada y fiable evaluación y a obligar a los usuarios a llegar a acuerdos que no son fáciles de obtener⁴³.

2. DIVERSOS ENFOQUES EN LA EVALUACIÓN ORAL DE LA LENGUA

En la evaluación de la lengua oral hay varios enfoques que se han ido desarrollando desde el último tercio del siglo pasado. Estos orientan el hecho de manera distinta según el concepto que tengan de los diferentes elementos que intervienen en la evaluación: la lengua, las escalas como sistemas de medir y escalonar su dominio, los objetivos, las formas de calificar, etc. En realidad, no son enfoques claramente

⁴² Alderson señala por ejemplo la falta de claridad y definición de muchos de sus términos, junto a la opacidad, solapamientos y ambigüedad “cuando las escalas se examinan de cerca” (Alderson, 2007: 661); o la dificultad que supone el definir los niveles C1 y C2, pues “definir descriptores de acuerdo a la lengua nativa es dudoso dada la dificultad de describir los niveles C1 y C2 solamente mediante elementos lingüísticos y no cognitivos o educativos” (Alderson, 2007: 662).

⁴³ Como dice Figueras, tras la publicación del *MCER* se produjo una euforia y aceptación iniciales pero “una vez disipada la euforia inicial surgieron las reservas sobre su “infalibilidad”, incluso sobre la necesidad de su existencia, y empiezan las críticas a la complejidad del texto, a la ausencia de concreción y a algunas inconsistencias desde distintos niveles educativos y desde distintos campos” (Figueras, 2005: 6), pronto se resaltaron “las dificultades que presenta el documento a la hora de concretar diseños curriculares o especificaciones de pruebas de evaluación”, “las críticas más duras surgen de la comunidad de evaluadores, cuando se constata que precisamente a causa de la gran aceptación del *MCER* por la sociedad, las calificaciones, los certificados y los exámenes europeos están abocados a relacionarse con los niveles del *MCER*” (Figueras, 2005: 7). Todo ello llevó a que en 2003 se publicara un nuevo documento que ayudara a la evaluación, un manual titulado *Relating examinations to the CEF* (ver Consejo de Europa, 2003, o Martínez Baztán 2008: 240-243).

excluyentes (aunque puedan parecer a primera vista opuestos) ya que, aunque en las teorías se marquen distancias que aparenten ser insalvables, muchas de ellas se neutralizan en la práctica, a veces por falta de precisión en sus mecanismos de medición -lo que es debido a la dificultad de diseñar herramientas más precisas para evaluar un campo tan multidimensional-, otras veces porque la práctica pone al descubierto fallas en las teorías y limitaciones que las hacen entre sí menos diferentes de lo que parecía.

Entre estos enfoques presentamos cinco, de forma somera, con la intención de ofrecer una visión general de algunas de las diferencias en los modos de evaluación oral. Creemos que ello es necesario para entender alguno de los problemas, a veces de interpretación, otras de uso práctico, en que se puede caer por mezclar o confundir conceptos y métodos.

Estos enfoques varían en función de la diferente forma de entender:

1. El concepto de lengua oral y sus escalas de medición
2. El objeto de la evaluación oral
3. El uso dado a la evaluación oral
4. Los medios y mecanismos para hacer las entrevistas orales
5. Los objetivos de corrección de los exámenes

2. 1. El concepto de lengua y sus escalas de medición

Observando los conceptos de lengua a evaluar y las diferentes formas de definirla o de delimitar su dominio, podemos trazar dos posiciones o dos modos de evaluación.

a. Una considera la lengua como un campo vasto cuyo uso se debe delimitar en áreas de dominio predefinidas, pues de lo contrario no es posible abarcarla, medirla y calificarla con objetividad. Esta compartimenta la lengua mediante puntos discretos ya sean desde un punto de vista estructural (gramática, vocabulario, pronunciación) o desde un punto de vista más pragmático (habilidad para iniciar intervenciones, tomar turnos de palabra, hacer uso de la interacción o compensar situaciones problemáticas);

b. Otra, holística, no usa puntos discretos para compartimentar la lengua sino que intenta dar una calificación general del uso de la lengua sin predefinir sus componentes, porque estima que la existencia de estos no refleja el uso real de la lengua en ningún momento concreto ni ayuda a entender cómo va a ser la lengua que se use en la vida real. Normalmente la describe y delimita mediante definiciones de su uso concreto y contextualizado.

Liskin-Gasparro (1987: 33) manifiesta con un ejemplo práctico estas dos variedades cuando se pregunta cómo se califica en un test oral a los estudiantes a los que se presenta un dibujo de un chico hablando por teléfono en una cabina. El primer candidato -recordando lo estudiado en clase y los ejercicios hechos respecto al uso de

los verbos en presente- dice: *El chico habla por teléfono*. Mientras que el segundo candidato dice: *Juan hablo en teléfono con su amigo Jorge que es en el hospital porque es enfermo*.

Si los calificáramos en una escala de puntos discretos el primero obtendría mejores notas que el segundo por el uso correcto del léxico, de las formas verbales y de la corrección sintáctica. El otro quizás saliera perjudicado por no haber hecho bien su intervención, a pesar del uso correcto del léxico. La primera escala olvida que el lenguaje se aprende para satisfacer determinadas necesidades funcionales, en este caso explicar un asunto, y que probablemente el segundo estudiante debería ser premiado por un uso creativo de la lengua para expresar sus propios pensamientos en vez de repetir ejemplos de lo aprendido en los ejercicios de clase. Lógicamente si los calificáramos según una escala holística, el segundo candidato saldría premiado.

2.2. El objeto de la evaluación

Este segundo concepto también produce modelos distintos de tests: los que se interesan por averiguar y predecir el conocimiento que el hablante tenga de la lengua entendida esta como sistema (en la definición clásica de la lingüística estructuralista o como competencia de uso de la lengua⁴⁴); y los que se interesan por demostrar el uso concreto de la lengua en contextos sociales o personales considerados útiles, bien para trabajar, para desenvolverse o para interactuar, independientemente de cual sea la competencia que se posea. Cada uno de ellos elabora diferentes mecanismos para evaluar la lengua.

a. El primero pretende elaborar, mediante los datos obtenidos en el test, predicciones sobre la capacidad de usar la lengua en situaciones reales basándose en que el test es prueba de que esa competencia se posee y que se podrá poner en funcionamiento fuera del test, en la vida real. Trata, por lo tanto, de inferir el conocimiento de la lengua de forma indirecta a través del uso de la lengua que el candidato muestra durante el test. Y, para elicitarse el discurso que sea prueba de ese otro discurso ajeno al test, elabora unos mecanismos que demuestren que la lengua usada en el test es producida efectivamente por las habilidades que el candidato posee, que es lo que en última instancia se le quiere evaluar.

Este modelo, bien establecido en Bachman (1990) se centra en la autenticidad de las tareas y de las inferencias que con ella se realizan; es decir, en la capacidad que muestren los sistemas de calificación para relacionar las pruebas obtenidas mediante el test, con el uso de la lengua en la vida real.

b. El segundo se conforma con probar que el candidato posee unas habilidades en contextos concretos y que el uso directo de la lengua durante el test es prueba de ellas. Entonces, este modelo se interesa en buscar mecanismos que prueben que esa lengua es muestra de la mejor lengua posible de los candidatos, que es lo que se quiere evaluar en realidad.

⁴⁴ Bachman aclara (1988: 161) que él se refiere a *competencia* en el sentido de Hymes, no limitado a la competencia lingüística definida por Chomsky, sino relacionada con todo lo que el hablante tiene en la mente para hacer uso de la lengua.

Para este último modelo la prueba del funcionamiento en la vida real se obtendría mediante el test mismo, considerando (sin entrar a definir las competencias existentes, las cuales también se acepta que existen), también de forma predictiva, que lo usado en el test se podrá volver a usar en otros momentos y situaciones similares. Estos tests también buscan una muestra de lengua apta para elaborar predicciones sobre su uso fuera del test, solo que utilizan mecanismos diferentes⁴⁵.

Cabría plantearse si estas dos concepciones delimitan tests que son realmente excluyentes y en qué situaciones se solapan o hablan en realidad de lo mismo. Porque muchos de los tests existentes hoy en día basados en puntos discretos que intentan medir el conocimiento de elementos como el dominio del vocabulario, de la gramática, etc. (que son parte de ese sistema de la lengua al que pretenden acceder indirectamente), incluyen otros elementos pragmáticos y de compensación y adecuación sociolingüística que relativizan el conocimiento del sistema de la lengua y se fijan en el uso que de este se hace ante situaciones concretas. Ya que en las teorías pragmáticas actuales cada vez se tiene más en cuenta a los elementos contextuales como propios del sistema de la lengua y las condiciones de significación y sus elementos constitutivos pasan a ser estudiadas como parte de ese sistema general que es la lengua.

2.3. El uso dado a la evaluación

El tercer concepto se diferencia por el campo de actuación para el que se realiza la evaluación y atañe también al uso que se haga de sus resultados. En este sentido podemos diferenciar cuando se hace con fines académicos o cuando se hace con otros objetivos sociales.

a. Cuando se hace con fines académicos es para medir el estado de la lengua de un candidato (sea el que sea y desde el concepto de lengua que sea) con el fin de describir las actuaciones futuras que se puedan llevar a cabo para mejorarla. O con el de elaborar mecanismos que ayuden a mejorar la enseñanza, la calidad del aprendizaje o el aprendizaje en sí (tests de diagnóstico, tests de clasificación, tests de rendimiento, etc.).

b. La evaluación con fines no académicos se desinteresa del contexto de aprendizaje, del diagnóstico de la lengua y del conocimiento pasivo que el candidato tenga. Se centra en medir la lengua para fines personales, sociales o laborales y en definir su uso en situaciones determinadas con el fin de demostrar que el candidato sabrá realizar fuera del test el mismo tipo de actividades que haya mostrado que sabe hacer durante el test.

⁴⁵ El principal examen que responde a este modelo es la OPI (Oral Proficiency Interview) de ACTFL (American Council on the Teaching of Foreign Languages, ver www.actfl.org), tan criticado durante los años 90 por el propio Bachman, entre otros, por creer que el sistema le restaba fiabilidad y autenticidad. No obstante Bachman ha terminado por aceptar las explicaciones de la organización y otros investigadores y situarlo entre los test basados en un criterio (Bachman, 2004: 301).

Aunque estos dos modelos tampoco sean excluyentes⁴⁶, suelen emplear conceptos o enfoques diversos sobre la lengua, que llevan a resultados también diferentes.

2.4. Los medios y mecanismos para hacer las entrevistas

El cuarto elemento basa sus diferencias en la concepción material del test y sus mecanismos o herramientas y en los modos de operar para obtener la información; es decir, en los métodos de elicitar y obtener las muestras de lengua y en el papel que juega el evaluador como motivador del discurso.

Aquí encontramos la opción de no intervenir aun cuando esto pueda penalizar la actuación de un candidato; y la que propugna perseguir los objetivos de la evaluación mediante un protocolo de intervención que logre extraer de cada candidato las mejores muestras de su habilidad. Estas diferencias inciden directamente en la concepción de la entrevista oral como un diálogo natural, o no, y ofrecen técnicas diversas y excluyentes.

a. El primer modelo de entrevista utiliza técnicas que pretenden acercarse a lo que es una conversación lo más natural posible, creyendo que ese es el mejor o más fiable camino para demostrar la existencia de las habilidades. Este enfoque persigue la naturalidad absoluta del discurso y critica las entrevistas que no contemplan en sus objetivos la variedad de usos y aspectos que se han delimitado en los últimos años en un buen número de estudios de pragmática como son la toma de turnos de palabra, el análisis de la interacción, el reconocimiento de intenciones, la negociación de significados o la diferencia entre conversación y actividades conversacionales.

b. La segunda utiliza técnicas más planificadas y se aleja de la naturalidad de un diálogo real pretendiendo ganar -con esta pérdida de pragmática- eficiencia en sus resultados y sistemas de calificar. No presta tanta atención a la variedad de elementos sino que se centra en probar que lo que se logra elicitar realmente es prueba del mejor hacer de los candidatos. Entre estas hay modelos que también pretenden acercarse a una conversación natural, pero partiendo de estos otros mecanismos.

2.5. Los objetivos de corrección de los exámenes

Y el quinto elemento marca diferencias en los objetivos de las mediciones de la lengua, en si estas pretenden averiguar si lo que el hablante sabe es parte del conocimiento explícito o implícito y qué cantidad de gramática conoce de forma implícita. La distinción la hacen Shohamy (1988: 170) o Ellis (2001) al hablar de estas dos formas de acceder a la gramática y al conocimiento de la lengua.

El conocimiento implícito es el que se produce en el discurso no planeado y el explícito el que surge en condiciones sin presión ante las que se puede recurrir a otras formas de conocimiento y reglas de gramática aprendidas.

⁴⁶ No son siempre excluyentes porque se puede evaluar conjuntamente con fines académicos y laborales (para establecer un programa de formación de técnicos, de profesores, etc.) e, igualmente, una evaluación con fines sociales puede tener usos académicos, sirviendo como base para tomar decisiones de selección para cursos, o como dato homologable para cualquier otro fin.

El problema para Ellis está en saber qué tipo de acceso permiten los tests y qué conocimiento están midiendo (Ellis, 2001: 253). Para él, generalmente, si no siempre, se preocupan por medir el conocimiento explícito, ya que es difícil diseñar otros tests que logren inferir el conocimiento implícito. Y ello supone un problema de medición, ya que los estudios de adquisición de las lenguas manifiestan que se puede mostrar un determinado grado de corrección y, sin embargo, estar en un estadio inferior de adquisición (Ellis, 2001: 255-56).

La corrección no muestra el grado de conocimiento. Medir el conocimiento de la lengua por el grado de corrección al usar el conocimiento explícito puede llevar a errores de medición y de predicción sobre el verdadero dominio de la lengua. Así se puede penalizar a candidatos con un mayor desarrollo de sus interlenguas frente a otros que sólo por una mayor corrección obtuvieran mejores notas (Martínez Baztán, 1995: 366).

2.6. Otros enfoques

A estos cinco enfoques sobre el concepto de la evaluación general, podemos añadir otros tres con diferentes perspectivas sobre el modo de construcción de los tests (Alderson / Banerjee, 2002; Fulcher, 2003).

Hasta ahora los principales modos de evaluar la lengua hablada han sido conductistas y se han basado en la premisa de autenticidad de que el modo de evaluar y el resultado obtenido se asemejen a la lengua usada en la vida real. Así son muchos de los que se utilizan desde los años 50, junto a los constructos lingüísticos derivados del modelo de lengua de Bachman (1990: 84) y Bachman / Palmer (1996: 68). Estos se caracterizan por describir una serie de tareas o ítems que provoquen un estímulo lingüístico, una actividad o comportamiento observable del cual se deduzca la existencia de una habilidad. Las tareas o ítems se diseñan con el objetivo de provocar una respuesta determinada y se valoran en la medida en que son capaces de provocarla de acuerdo con el constructo que se desee evaluar (Alderson / Banerjee, 2002: 99).

Pero esta no es la única perspectiva. Podemos ver tres más: el nuevo conductismo, la teoría de los rasgos lingüísticos, y el interaccionismo.

2.6.a. El nuevo conductismo estima que el constructo lingüístico viene definido principalmente por tareas (Fulcher, 2003: 62). Igual que el conductismo lingüístico, que cree que las tareas y sus contextos están indisolublemente ligadas a la habilidad que provocan.

2.6.b. La teoría de los rasgos lingüísticos estima que los tipos de tareas utilizados son irrelevantes para el constructo a medir y que los rasgos lingüísticos son independientes de los contextos en que se generan. Así, las notas de un test se atribuirían fácilmente a las características del candidato, a su conocimiento y a procesos internos antes que a características del contexto o del lugar.

Alderson / Banerjee explican esta perspectiva como una de las que más tradición tienen en el ámbito de la evaluación educativa, aunque la consideran simplista pues se sabe, por la lingüística aplicada y los estudios de adquisición de segundas lenguas, que el uso

de la lengua no es independiente del contexto en que surge (Alderson / Banerjee, 2002: 100).

2.6.c. El interaccionismo, por otra parte, cree que ambas perspectivas son erróneas y que hay que tomar elementos de todas partes. Algunas características de la habilidad lingüística proceden de los rasgos, otras del contexto y otras de la interacción “aunque lo difícil consiste en averiguar la medida en que se presentan los rasgos, cómo interactúan con el contexto y bajo qué circunstancias, para combinarlos en el diseño de un test” (Alderson / Banerjee, 2002: 100)⁴⁷.

3. EL DISEÑO DE UN TEST ORAL

Al diseñar un test se deben especificar adecuadamente y de la mejor forma posible, siempre dependiendo de las necesidades del momento y sus objetivos de evaluación, todos los elementos que se aspire a calificar; ya que dejarlos indeterminados, sea por no definir bien las habilidades a medir, por no aclarar cómo se interpretan las actuaciones de los candidatos en relación a la existencia de las habilidades, o por no controlar los efectos de factores que no se desean evaluar, hace que el resultado reste validez al test.

No obstante, y sin olvidar este problema de falta de definición, se debe tener en cuenta que los tests no son absolutos y siempre van a ser mejorables, bien porque se hallen nuevos datos en investigaciones sobre desarrollo o adquisición de la lengua, bien por el hecho de que su diseño complejo y costoso puede ser mejorado tomando en cuenta elementos que quizás no hayan sido bien considerados o se hayan minimizado en su diseño.

Este problema de la definición supone al mismo tiempo una limitación de los resultados, que hay que tener clara y saber manejar para evitar hacer un mal uso de los datos obtenidos extrayendo de ellos conclusiones inadecuadas, exageradas, inexistentes o falsas.

Las limitaciones más importantes de las que hay que ser consciente son (Bachman / Savignon, 1986: 380; Bachman, 1990: 30-40; Bachman, 2004: 14):

1. La falta de especificación de los contenidos del test, entre los que destacan las escalas de evaluación y gradación del dominio.
2. Las derivadas de una inadecuada forma de realización de entrevistas.

⁴⁷ Fulcher (2003: 62), quien diferencia en la práctica los elementos (rasgos o tareas) y los tests en los que entran a formar parte, aclara que hay tests en los que las tareas son parte del constructo, como los tests para fines específicos, pero que también hay elementos de las tareas que no influyen en las posibilidades de generalizar las habilidades, esto es, que no impiden que se puedan generalizar a otras situaciones más allá de las directamente evaluadas en el test; por lo que también se pueden hacer tests en los que la influencia de la tareas quede minimizada.

3- La dificultad de observación de las habilidades y -derivado de ello- la dificultad de interpretación de los datos obtenidos.

El primero, la falta de especificación de los contenidos, incluye también la importancia de diseñar tests con buenas tareas, que sean –lógicamente- facilitadoras de las habilidades a evaluar; así como de los sistemas de definir la habilidad lingüística mediante el uso de las escalas de medición y criterios de calificación de la lengua. Al respecto, Bachman (2004: 28) dice que, además de la propia habilidad lingüística, es tal la variedad de factores que afectan a la actuación lingüística, como el estilo cognitivo, la ansiedad, la afectividad o el conocimiento del tema, que es imposible tenerlos todos en cuenta al diseñar el test o al interpretar los resultados, lo cual, de alguna manera, indetermina las interpretaciones.

Del segundo, las limitaciones derivadas de una inadecuada forma de realización de entrevistas, se hablará al final de este artículo (apartado 4) al tratar del nuevo papel que el entrevistador⁴⁸ tiene en la realización de las entrevistas como cooperador necesario en la génesis del discurso y su influencia en todo el proceso.

Respecto al tercer foco de limitaciones, la dificultad de observación de las habilidades, aunque se relaciona con las demás y a veces es difícil de diferenciar, podemos decir con Bachman⁴⁹ que procede, en principio, de que las habilidades lingüísticas no se pueden observar de forma directa y ni siquiera la observación de la actuación viva de los hablantes es clara ya que está sometida a variaciones derivadas del contexto, de las tareas que se estén realizando en el test, de la alerta mental y otras características cognitivas y personales o de elementos más impredecibles, como la hora del día⁵⁰.

Algunas limitaciones se pueden disminuir mediante el diseño de los mecanismos del test que contribuyen a dar objetividad. Por eso es importante delimitar los numerosos aspectos que influyen en su realización y, en última instancia, ser conscientes de qué tipo de herramientas, de entre todas las que se pueden utilizar, estamos manejando, qué consecuencias conllevan y cuáles son sus puntos fuertes y sus limitaciones⁵¹. Sin

⁴⁸ Véase también Lazaraton, 1996 y Lazaraton, 2002.

⁴⁹ También el *MCER* dice claramente que “nunca se pueden evaluar las competencias directamente. Hay que basarse siempre en una serie de actuaciones a partir de las cuales se intenta generalizar una idea del dominio lingüístico (...) Todas las pruebas evalúan solo la actuación aunque partiendo de esa evidencia se puede intentar extraer inferencias respecto a las competencias que subyacen” (Consejo de Europa, 2002: 188).

⁵⁰ Bachman cita cinco elementos en relación a estas limitaciones (Bachman, 1990: 32-39; Bachman, 2004: 28-29): 1. Dependencia de un método de observación indirecto, 2. Deficiencias de la actuación y obtención de datos, 3. Imprecisión en la interpretación de los datos, 4. Subjetividad general de los tests y sus mecanismos y 5. Relatividad de las habilidades a evaluar (Esta relatividad de definir el objeto se hace más compleja al tratar de definir la lengua de las comunidades de residentes extranjeros. Esto también lo señala Winford argumentando que entre las comunidades de no nativos e inmigrantes las variedades se adaptan y adquieren su propia versión de la lengua extranjera -Winford, 2003: 237-).

⁵¹ Hay muchos manuales y autores que tratan de forma profesional pero sencilla aspectos y técnicas importantes que, fundamentalmente desde una perspectiva cualitativa y otra estadística, permiten llevar a los tests la objetividad necesaria para que estas limitaciones -si no se pueden reducir al máximo- tengan, al menos, poca influencia en los resultados. Ver por ejemplo Alderson et al., 1998, Fulcher, 2003 o Bachman, 2004.

que se puedan hacer generalizaciones ni comparaciones simples entre diferentes métodos.

Todas estas precauciones que venimos señalando no deben apartarnos del intento de realizar una buena evaluación oral, tanto si se hace con fines académicos como con funciones laborales, sociales o jurídicas. Pues la evaluación oral es un tipo superior de evaluación que recoge a todas las demás, la única que tiene la capacidad de decir, si se hace adecuadamente, cómo es efectivamente la capacidad para comunicarse en una lengua extranjera desde cualquier punto de vista, la expresión y la comprensión en general o con fines específicos, la interacción y la toma de turnos de palabra, la compensación, la competencia gramatical y léxica, etc. Aunque para ello sea necesario diseñar tests orales adecuados a cada situación y a los candidatos, ya que –como hemos apuntado al comienzo- una conversación natural por sí misma, por muy satisfactoria que parezca, no es ni buena ni fiable (Bordón, 2006: 62; Bachman, 1990: 21, 34).

4. EL PAPEL DEL ENTREVISTADOR EN LOS EXÁMENES ORALES

Es bien sabido que un elemento que resta validez y fiabilidad a un test es que un mismo candidato pudiera recibir diferentes calificaciones en la misma entrevista. Otro de los elementos que influye en dotar de fiabilidad a los tests orales, uno de los últimos analizados, desde perspectivas de análisis del discurso, y que promete replantear muchos de los elementos y conceptos de la evaluación oral, es el referente a si las técnicas de elicitación del discurso y el comportamiento del evaluador generan un discurso especial en el hablante o le influyen subordinando el discurso del candidato al comportamiento del evaluador, lo que puede comprometer el resultado tanto de la entrevista como de su calificación⁵².

Desde que se analizan los test orales y sus muestras grabadas con técnicas de análisis del discurso, se ha llegado a entender que los modos de los entrevistadores pueden variar mucho, haciendo que cada entrevista sea diferente a otra e impidiendo a los candidatos tener las mismas oportunidades (Lazaraton, 1996: 166), imposibilitando que sean evaluados o calificados de la misma manera y –en peor caso- que puedan demostrar sus habilidades en igualdad de condiciones. Para esta autora, el resultado de una evaluación oral, como el de una conversación, es el resultado de la colaboración entre todas las partes. Si esto es así⁵³, habrá que delimitar los elementos para

⁵² La bibliografía es ya larga y el tema se presta a ser tratado por sí solo en otro momento aclarando y explicando las diferentes posiciones existentes al respecto, los datos que se poseen sobre la interferencia de elementos, y la redefinición que ello puede suponer para los elementos tradicionales de la evaluación oral. Aquí no hacemos sino una breve introducción a este problema. Véase Young, 1995, He / Young, 1998: 5, Ross, 1998, Lazaraton, 2002: 44, Brown, 2003: 20, entre otros. Para estos autores la habilidad no preexiste a la entrevista sino que se construye durante ella, ellos (los entrevistadores) producen o fabrican las habilidades que se supone que se miden (He / Young, 1998: 8). En esta línea está el trabajo de Ross (1998) para quien en las entrevistas orales se crea un contexto específico que influye en el tipo de respuestas que los candidatos dan (1998: 336).

⁵³ Los resultados de los análisis e investigaciones no aportan datos suficientes o suficientemente concluyentes. El problema está en dilucidar el tipo de interacción que se produce en las entrevistas y, lo que es de más calado, el papel del entrevistador en todo ello, en la realización del proceso mediante el

establecer los correctivos necesarios para garantizar que los resultados sean verdaderos, ni ficticios ni sesgados; y que no premien a determinados tipos de candidatos.

Ya no son solamente el test en sí y sus tareas los catalizadores de las diferencias de habilidad existente entre diferentes candidatos sino que el evaluador, conductor de la máquina, aparece ahora como el responsable de su uso y de los resultados que con ella se obtengan. La pregunta es ahora ¿en qué medida influye el evaluador en la obtención y calificación de las habilidades del candidato? O más crudamente, ¿en qué medida es causante de las diferencias de nivel que él mismo califica?

Estos hechos dan al entrevistador un papel mucho más importante del que hasta ahora se había considerado y parece ser que ello puede tener consecuencias en las calificaciones, si no se le presta la debida atención. A partir de ahora será necesario analizar cómo actúa el candidato, cómo interpreta las preguntas en el transcurso de la reorganización del contexto y sentido generado dentro de la entrevista y, asimismo, cómo interpreta el entrevistador las respuestas, sobre la marcha.

En este sentido, el entrevistador ha dejado de ser un mero actor secundario que ponía en práctica, con mayor o menor celo, todo el mecanismo de actividades, tareas y preguntas que un equipo de expertos había diseñado cuidadosa y estadísticamente. El evaluador ha pasado a considerarse como *cooperador necesario* en la construcción del discurso del candidato, mediante la selección de temas a tratar, mediante el uso de la entrevista y de las técnicas elicitoras, adaptándose al nivel y necesidades del candidato, interpretando su discurso o interpretando las escalas de calificación.

El tema es importante porque compromete no solo al constructo del test y a su contenido sino los juicios que de ellos resulten. Y porque ello implica no sólo al evaluador y al candidato, sino a la propia estructura y configuración del test elegido.

Desde esta nueva perspectiva, diseñar un buen test oral y una adecuada formación de evaluadores que permita elicitar e interpretar los discursos de forma consistente requiere ser consciente del papel tan importante que el entrevistador tiene en la génesis del discurso, junto con otros como la dificultad de la tarea, la situación y el contexto producido en la interacción a lo largo y dentro del test. Por ello nos parece de suma importancia entender el proceso de creación de sentido que se produce en una entrevista, proceso que deben conocer el entrevistador y el calificador, y al que están sujetos todos los participantes, junto al candidato, consciente o inconscientemente.

El estatuto e importancia del trabajo del entrevistador de evaluación oral ha cambiado desde que se ha estudiado su influencia en la producción del discurso. Y ello, sin duda, va a introducir otros cambios en la evaluación oral y en su concepción: por ejemplo en el papel de las escalas de evaluación y de calificación; en los métodos de elicitación del discurso y tipología de preguntas utilizados; en el comportamiento de los evaluadores;

cual se toman las decisiones finales sobre la habilidad y en la forma de tomar esas decisiones (Ross, 1992: 173). Lo cual exige más profundidad de análisis de la conversación y de la interacción, para analizar en qué medida todo ello influye en el discurso, en la estructura de la entrevista y en las calificaciones (Leung / Leukowicz, 2000: 217).

el método y estructura de la entrevista oral, con la necesidad de la grabación de las entrevistas; en el análisis de las pretensiones e intenciones que tenga el candidato ante las tareas del test; y en la interpretación que el evaluador haga del discurso obtenido, tanto durante la entrevista como durante la fase de calificación.

Estos asuntos no sólo deben ser introducidos en el diseño del test y en la formación de evaluadores sino que obligan a replantear la necesidad de definir otros elementos como el marco en que tenga lugar la entrevista, su estructura, los objetivos de cada fase, el sistema de interacción que se va a producir, o las formas de elicitar el discurso e interpretarlo. Elementos que se suman a los ya tratados a lo largo de este artículo y que convierten a la evaluación oral de las lenguas en una materia necesitada de investigación y formación.

BIBLIOGRAFÍA

Alderson J. Ch. / Clapham, C. / Wall, D. (1998): *Exámenes de idiomas. Elaboración y evaluación*, Madrid: Cambridge University Press. [Título original: *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press (1995)].

Alderson, J. Ch. (2007): "The CEFR and the Need for More Research", *Modern Language Journal*, 91 (4): 659- 663.

Alderson, J. Ch. / Banerjee, J. (2002): "Language Testing and Assessment (Part 2)", *Language Testing*, 35: 79-113.

Bachman, L. F. (1988): "Problems in Examining the Validity of the ACTFL Oral Proficiency Interview", *Studies in Second Language Acquisition*, 10 (2): 149- 163.

Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

Bachman, L. F. (2004): *Statistical Analyses for Language Assessment*, Cambridge: Cambridge University Press.

Bachman, L. F. / Savignon, S. J. (1986): "The Evaluation of Communicative Language Proficiency: A Critique of the ACTFL Oral Interview", *The Modern Language Journal*, 70 (4): 380-390.

Bachman, L. / Palmer, A. (1996): *Language Testing in Practice*, Oxford: Oxford University Press.

Bordón, T. (2006): *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*, Madrid: Arco/ Libros, S. L.

Brown, A. (2003): "Interviewer variation and the co-construction of speaking proficiency", *Language Testing*, 20 (1): 1-25.

Canale, M. / Swain, M. (1980): "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing", *Applied Linguistics*, 1:1- 47.

Chalhoub-Deville, M. (1997): "Theoretical models, assessment frameworks and test construction", *Language Testing*, 14 (1): 3-22.

Consejo de Europa (2002): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Ministerio de Educación, Cultura y Deporte; Instituto Cervantes. [Título original: *Common European Framework of Reference for Languages Learning, teaching, assessment*, (2001): Strasbourg: Council of Europe.]

Consejo de Europa (2003): *Relating Language Examinations to the CEF, Preliminary Pilot Version*, [pdf], Strasbourg: Council of Europe, [consulta: 8 de agosto de 2005], Disponible en la web: <http://www.coe.int/lang>

Ellis, R. (2001): "Some thoughts on testing grammar: An SLA perspective", en Elder, C. / Brown, A. / Grove, E. / Hill, K. / Iwashita, N. / Lumley, T. / Mcnamara, T. / O'Loughlin, K. (eds): *Experimenting with uncertainty. Essays in honour of Alan Davies*, Cambridge: Cambridge University Press: 251- 263

Figueras, N. (2005): "El Marco común europeo de referencia para las lenguas: de la teoría a la práctica", *Carabela*, 57: 5-23.

Fulcher, G. (2003): *Testing Second Language Speaking*, London: Pearson / Longman.

He, A. W. / Young, R. (1998): "Language Proficiency Interviews: A Discourse Approach", en Young, R. / He, A. W. (eds.) (1988): *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam / Philadelphia: John Benjamin: 1-23.

Lazaraton, A. (1996): "Interlocutor support in oral proficiency interviews: the case of CASE", *Language Testing*, 13 (2): 151-72.

Lazaraton, A. (2002): *A Qualitative Approach to the Validation of Oral language Tests*, Cambridge: UCLES, Cambridge University Press.

Leung, C. / Lewkowicz, J. (2000): "Expanding Horizons and Unresolved Conundrums: Language Testing and Assessment", *TESOL Quarterly*, 40: 211-240.

Liskin-Gasparro, J. (1987): "The ACTFL Proficiency Guidelines: A Historical Perspective", en Higgs, T. V. (ed.) (1987): *Teaching for proficiency, the Organizing Principle*, Lincolnwood, IL: National Textbook Co.: 11-42.

Martínez Baztán, A. (1995): "¿Les enseñamos, aprenden o aprenden a pesar de nosotros?", en *Aspectos de la enseñanza de lenguas extranjeras*, Granada: Método ediciones: 357-69.

Martínez Baztán, A. (2008): *La evaluación oral: una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER*. Tesis doctoral presentada en la Universidad de Granada.

Ortega Olivares, J. (1996): "Evaluación en el aula de español/LE: algunas consideraciones generales", en Rueda, M. / Prado, E. / Lemen, J. / Grande, F. J. (eds.): *Tendencias actuales en la enseñanza del español lengua extranjera II: actas del VI Congreso Internacional de ASELE*, León: Universidad de León: 19-33.

Parrondo, J. R. (2004): "Modelos, tipos y escalas de evaluación", en Sánchez Lobato, J. / Santos Gargallo, I. (dirs.): *Vademécum para la formación de profesores*, Madrid: SGEL Ross, 1998

Ross, S. (1992): "Accommodative questions in oral proficiency interviews", *Language Testing*, 9: 173- 186.

Shohamy, E. (1988): "A proposed framework for testing the oral language of second foreign language learners", *Second language acquisition*, 10: 165-179.

Walters, F. S. (2007): "A conversation- analytic hermeneutic rating protocol to assess L'oral pragmatic competent", *Language Testing*, 2: 155-183.

Widowson, H. (2003): *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.

Wilkins, D. A. (1976): *Notional Syllabuses*, Oxford: Oxford University Press.

Winford, D. (2003): *An introduction to contact linguistics*, London: Blacwell.

Young, R. (1995): "Conversation Styles in Language Proficiency Interviews", *Language Learning*, 45: 3-42.