



Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Sistema de Información Científica

González-Such, José; Jornet Meliá, Jesús Miguel; Bakieva, Margarita  
Consideraciones metodológicas sobre la evaluación de la competencia oral en L2  
REDIE. Revista Electrónica de Investigación Educativa, vol. 3, núm. 15, 2013, pp. 1-20  
Universidad Autónoma de Baja California  
Ensenada, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=15529662001>



*REDIE. Revista Electrónica de Investigación Educativa,*  
ISSN (Versión electrónica): 1607-4041  
[mayer.cabrera@uabc.edu.mx](mailto:mayer.cabrera@uabc.edu.mx)  
Universidad Autónoma de Baja California  
México

¿Cómo citar?

Número completo

Más información del artículo

Página de la revista

[www.redalyc.org](http://www.redalyc.org)

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## Revista Electrónica de Investigación Educativa

Vol. 15, Núm. 3, 2013

### Consideraciones metodológicas sobre la evaluación de la competencia oral en L2

### Methodological Considerations Concerning the Assessment of Oral Competency in a Second Language (L2)

José González-Such\*

[jose.gonzalez@uv.es](mailto:jose.gonzalez@uv.es)

Jesús Miguel Jornet Meliá\*

[jesus.m.jornet@gmail.com](mailto:jesus.m.jornet@gmail.com)

Margarita Bakieva\*

[margaritabakieva@gmail.com](mailto:margaritabakieva@gmail.com)

\*Facultad de Filosofía y Ciencias de la Educación, Universitat de València

Av. Blasco Ibáñez, 30  
46021 Valencia, España

(Recibido: 12 de octubre de 2012; aceptado para su publicación: 25 de junio de 2013)

#### Resumen

En este artículo se realiza una revisión acerca de las consideraciones metodológicas para la evaluación de la competencia oral en el aprendizaje de una segunda lengua (L2). Se analizan las características del constructo y su definición. En este caso, se refieren al ámbito competencial, que incluye la competencia en comprensión oral y la de expresión oral. Asimismo, se revisan conceptos referidos al modo en que diseñar instrumentos y/o técnicas evaluativas para la evaluación de las competencias que la integran y cuestiones respecto a fiabilidad, validez y establecimiento de puntos de corte.

*Palabras clave:* Evaluación, Competencias, Aprendizaje de segunda lengua.

## **Abstract**

In this paper, methodological considerations are reviewed regarding the assessment of oral competence in the learning of a second language (L2). The characteristics of the construct and its definition are analyzed, including competence in listening comprehension and oral expression. Furthermore, concepts such as the design of instruments and/or evaluative techniques for the assessment of the competencies are examined as well as issues related to reliability, validity and cut off points.

*Keywords:* Evaluation, Proficiency assessment, Second language learning.

## **I. Introducción**

En este artículo abordamos la revisión de la competencia oral como constructo de evaluación. Nuestra revisión no es exhaustiva. Requeriría de mayor espacio para poder analizarla adecuadamente. En estas reflexiones hemos priorizado diversos objetivos. En primer lugar, la referencia acerca de la definición del constructo. Ciertamente la competencia oral es un constructo complejo y requiere un análisis cuidadoso para poder diseñar sus modos de evaluación. En un primer apartado, realizamos dicha revisión y expresamos nuestra opinión acerca de la conveniencia de abordarlo como dos subcompetencias: comprensión oral y expresión oral. El segundo objetivo del artículo es reflexionar acerca de los componentes que deberían considerarse para definir el sistema de evaluación y los instrumentos de medida. Para ello tomamos como referencia un esquema de trabajo que desarrollamos en un trabajo anterior y que, por la experiencia desarrollada, parece ser un marco de trabajo útil y abordable tanto por parte del profesorado como por los especialistas en los contenidos a evaluar. El último objetivo es considerar algunos aspectos relativos a la validez. La validez, como es bien conocido, es el criterio fundamental que deben cumplir todos los sistemas e instrumentos de evaluación. Podemos tener dificultades en otros criterios, pero la validez es la característica fundamental a cuidar en la elaboración de tecnologías de evaluación. En el apartado de conclusiones aportamos algunas consideraciones, y preocupaciones que, como especialistas en evaluación, nos suscita esta temática.

### **1.1 La competencia oral**

Probablemente la mayor dificultad que encontramos cuando intentamos abordar la competencia oral, como constructo a investigar y/o evaluar, es su definición. Así, podría señalarse que la evaluación de las destrezas orales ha tenido poca atención en comparación con otras destrezas del lenguaje, posiblemente debido a las dificultades derivadas de la identificación de los aspectos de lo que es en esencia una operación cognitiva invisible (Brindley, 1998), o cuanto menos parcialmente observable; de forma que aunque su estudio y evaluación tiene una larga historia, no es hasta la década de los ochenta cuando las pruebas de habilidad oral se hicieron comunes, debido al interés en la enseñanza por el lenguaje comunicativo (Alderson y Banerjee, 2002). En este sentido, la competencia oral es el foco central de la competencia en comunicación, e involucra diversas variables. Según el Marco Común Europeo de Referencia (en lo sucesivo MCER, o MCERL –Lenguas–) las estrategias de comunicación:

Son un medio que utiliza el usuario de la lengua para movilizar y equilibrar sus recursos, poner en funcionamiento destrezas y procedimientos con el fin de satisfacer las demandas de comunicación que hay en el contexto y completar con éxito la tarea en cuestión de la forma más completa o más económica posible, dependiendo de su finalidad concreta. (MECD, 2002, p. 59)

De este modo, además de las cuatro destrezas clásicas: comprensión auditiva (CA), expresión oral (EO), comprensión lectora (CL) y expresión escrita (EE), se incluyen la interacción y la mediación, contemplada en el MCER y derivada del reconocimiento del contexto tanto en los actos comunicativos de la vida real como en las aulas. Específicamente, respecto al plano oral, las estrategias de comunicación a considerar son las estrategias de expresión y las estrategias de interacción (Pinilla, 2007).

Desde el MCERL se considera que la competencia comunicativa tiene varios componentes: el lingüístico, el sociolingüístico y el pragmático, y cada uno de estos componentes comprende conocimientos, destrezas y habilidades.

- *Competencias lingüísticas.* Incluyen los conocimientos y las destrezas léxicas, fonológicas y sintácticas.
- *Competencias sociolingüísticas.* Se refieren a las condiciones socioculturales del uso de la lengua.
- *Competencias pragmáticas.* Tienen que ver con el uso funcional de los recursos lingüísticos (producciones de funciones de lengua, actos de habla) sobre la base de guiones o escenarios de intercambios comunicativos. También tiene que ver con el dominio del discurso, la cohesión y la coherencia, la identificación de tipos y formas de texto, la ironía y la parodia.

El estudio de la *habilidad pragmática* en L2 es un área que ha ido creciendo progresivamente. Leech (1983) distingue entre pragmalingüísticos, las herramientas lingüísticas necesarias para expresar y comprender las intenciones del habla, y sociopragmáticos, las reglas sociales que constriñen las elecciones lingüísticas del hablante y las posibles interpretaciones del oyente, estando ambos aspectos estrechamente relacionados. Crystal (1997) la caracteriza como el estudio del lenguaje desde el punto de vista de los usuarios, de las elecciones que realizan, las restricciones que encuentran al usar el lenguaje en la interacción social y los efectos que tiene su uso del lenguaje sobre otros participantes en la comunicación. Roever (2011) describe los componentes principales de esta competencia pragmática en la tabla I.

Tabla I. Componentes de la habilidad pragmática L2 con sub-constructos

<b>Monológica: Monólogo extendido</b>	<b>Dialógica: Participación en la interacción</b>	<b>Formulas de rutina</b>	<b>Implicación</b>
Producción y reconocimiento de: estilos de discurso, señales de contextualización, estructura de discurso	Producción y reconocimiento de: estilos de discurso, señales de contextualización, organización de secuencia: presecuencias, secuencias centrales, post-secuencias, aperturas y cierres, reparación Partes de respuesta al primer par Efectos sobre el interlocutor	Producción y reconocimiento de formulas de rutina	Comprensión de la implicación

Fuente: Roever, 2011, p. 473.

Desde una perspectiva de evaluación, sin embargo, es necesario considerar la competencia oral como un constructo macro-teórico, que incluye dos subconstructos diferenciados: la comprensión oral y la expresión oral, siendo las dimensiones comentadas anteriormente elementos o factores de las anteriores. Se trata de una definición operativa que nos puede facilitar la definición de estrategias evaluativas.

El proceso de *Comprensión Oral* se puede entender como una transición en la construcción del pensamiento desde un *input* auditivo, e incluye no sólo este tipo de variables de entrada, sino también conocimientos lingüísticos y no lingüísticos, como fonología, vocabulario y sintaxis, además de interpretación (Buck, 2001). Se da en tiempo real, de forma que contiene distintos conocimientos y variables específicas (por ejemplo, la atención), siendo un proceso individual, de carácter automático, y está influido –o condicionada– por otras variables, tales como el *input* acústico, la elisión, el manejo del estrés y la entonación. Las ideas no se expresan necesariamente de una forma gramatical y con frecuencia incluyen redundancia y vacilaciones (Alderson y Banerjee, 2002).

Por su parte, la *Expresión Oral* tiene a su vez una naturaleza compleja, en la que pueden identificarse diversos componentes (sintetizando la propuesta de Bordón, 2007a, p. 99):

- a) Como manifestación lingüística, sus elementos se pueden afectar por variedades dialectales y sociales:
  - Léxico: adecuación del vocabulario al tema y al contexto.
  - Morfosintáctico: uso de formas y estructuras necesarias en el discurso.
  - Fonológico: adecuación de la pronunciación y la entonación.
- b) El uso social está condicionado por el contexto y sus normas de adecuación, lo que afecta a: el discurso (configuración, extensión, etc.), selección del léxico y la morfosintaxis, y la presencia de vacilaciones, indecisiones, etc.
- c) En relación a la configuración de la expresión oral, podemos identificar tres factores:
  - Discurso interactivo: el texto se construye por la interacción entre los

participantes.

- Elementos extralingüísticos: estrategias de comunicación, gestos, proximidad.
- Limitaciones psicolingüísticas: lagunas en la memoria, ansiedad, provocadas por la falta de tiempo para planificar el discurso.

Con todo, la investigación sobre este campo (Fulcher y Reiter, 2003; O'Loughlin, 2002; Kim, 2009; Roever, 2011) ha aumentado nuestro conocimiento acerca de la multitud de variables que pueden tener influencia en las pruebas orales de rendimiento. Sin embargo, el debate continúa sobre la naturaleza del constructo hablado y de los métodos válidos para evaluar la expresión oral.

## **II. La evaluación de la competencia oral**

Abordar la evaluación de esta competencia es complejo, sobre todo si se analiza como un constructo global, y no se consideran los diferentes componentes que deben tenerse en cuenta para el diseño de los sistemas de evaluación. Adicionalmente, considerar este constructo como un conocimiento o una competencia (enfoques alternativos) también ha supuesto estrategias diferenciales de evaluación.

En un trabajo anterior (Jornet *et al.*, 2010) revisamos la estrategia basada en una perspectiva criterial para desarrollar instrumentos y/o técnicas para la evaluación de competencias. Esta revisión estuvo basada en la complementariedad metodológica de los métodos para determinar estándares, partiendo de que el concepto de calidad es subjetivo y para lograr que sea objetivo necesitamos ampliar la perspectiva de la medida con procesos de tipo cualitativo, especialmente relacionados con la validez del diseño de pruebas y de la interpretación de puntuaciones. Los ámbitos del trabajo con jueces y los elementos cuantitativos sobre la convergencia y consistencia de juicios son clave en la investigación para desarrollar nuevos aspectos. El esquema allí presentado lo seguimos en este caso para reflexionar acerca de los componentes de la evaluación de la Competencia Oral. En la tabla II se sintetizan los componentes metodológicos allí analizados.

Tabla II. Componentes para el diseño de evaluación de competencias

Componentes del proceso	Finalidad
I. Determinación y especificación del Dominio Competencial del constructo a evaluar.	Definir las competencias, subcompetencias y dimensiones que la integran. Es especialmente importante analizar las variables a considerar.
II. Definición de criterios de evaluación	Determinar las características de calidad que se van a considerar en la evaluación. Deben orientar –y permitir– una evaluación de calidad (fiable, válida y objetiva).
III. Especificación previa de un referente o Estándar de Ejecución, para determinar los criterios de evaluación.	Operativizar y explicitar los criterios para evaluar la competencia y las subcompetencias. Deben estar formulados de forma clara y precisa, permitiendo una identificación de diversos niveles de competencia.
IV. Sistematización del proceso de recolección de evidencias.	Se trata de identificar: <ul style="list-style-type: none"> <li>• Las situaciones evaluativas, en las que se observará la competencia</li> <li>• Las tareas a realizar por parte del estudiante como expresión de su competencia.</li> <li>• El rol del evaluador/a.</li> <li>• Las situaciones y tareas deben ser realistas, representando, en la medida de lo posible, funciones habituales en las que se pueda observar la competencia a evaluar.</li> </ul>
V. Determinación del sistema de evaluación, instrumentos, recursos y técnicas de recogida de información: Indicadores	Se trata de definir qué instrumentos se van a utilizar para cada competencia y/o subcompetencia. Identificar indicadores fiables y válidos.
VI. Definir un sistema de comparación de evidencias con el Estándar.	Determinar un sistema que permita trasladar las valoraciones (cuantitativas o cualitativas) a puntuaciones. La graduación del desempeño en las competencias es un elemento clave a considerar en el sistema de evaluación. Por ello, la diferenciación entre niveles debe objetivarse de forma precisa: es necesaria una aproximación numérica que sintetice el grado de desempeño en una competencia.
VII. Determinación del proceso de formación de la decisión (por ejemplo, Competente vs. Aún no competente).	Establecer las puntuaciones de corte que permitan diferenciar entre diversos niveles o grados de competencia. Un aspecto clave del sistema de evaluación de competencias es definir los elementos que diferencian entre niveles de competencia. La calidad de las puntuaciones de corte entre niveles implican directamente a la validez de la evaluación.
VIII. Especificar los mecanismos de información a audiencias y/o retroalimentación para el estudiante.	Definir los procesos por los que se informará al estudiante acerca de su nivel de competencia y, en todo caso, de las evidencias que pongan de manifiesto sus lagunas en el dominio global de la competencia que se señala en el Estándar.
IX. Diseñar los mecanismos o criterios de calidad del propio sistema de evaluación.	Determinar las estrategias de evaluación para informar de la calidad del sistema de evaluación, y analizar sus posibilidades de mejora.

Fuente: adaptado de Jornet *et al.*, 2010.

Siguiendo este análisis de componentes para el diseño de evaluaciones de competencias, el primer elemento a determinar es *¿qué se va a evaluar? ¿cómo definimos la competencia oral?*

En los últimos años, en Estados Unidos, se ha producido una controversia importante entre dos enfoques. Por una parte, la identificación de la competencia en una segunda

lengua con una serie de dificultades para evaluarla a nivel formal y, por tanto, de calificar de forma estándar; por otra, el deseo de tener exámenes ordenados para medir el conocimiento adquirido por los alumnos en los cursos, enfatizando las preguntas con una única respuesta correcta, y con un ahorro en su aplicación frente a la entrevista o la exposición oral. Esta dicotomía de evaluar conocimiento frente a evaluar competencia ha ido cambiando en función de la corriente pedagógica dominante (Liskin-Gasparro, 2007).

En la actualidad el enfoque se centra en desarrollar formas basadas en la habilidad práctica de utilizar el idioma, es decir, “formas de evaluación cuyos resultados indiquen lo que el individuo puede realizar en el mundo real con su habilidad lingüística”, y como resultado el proyecto del American Council on the Teaching of Foreign Languages<sup>1</sup>, en el que se realizó la escala de ACTFL, basada en las normas generales de la Asociación y que a su vez derivan de las descripciones de la habilidad lingüística desarrolladas por el gobierno de Estados Unidos, que “definen lo que una persona puede o no puede hacer con una lengua, sin considerar dónde, cuándo o cómo la haya aprendido” (Liskin-Gasparro, 2007, p. 22).

En definitiva, el énfasis se sitúa en el aspecto comunicativo, funcional del lenguaje, de manera que ambas subcompetencias, comprensión y expresión, pueden identificarse como las competencias clave implicadas. En cualquier caso, es obvio que ambas subcompetencias tienen problemas de definición y de evaluación claramente diferentes y de niveles de complejidad en sus soluciones también diferenciados.

El segundo componente a abordar es el de la determinación de criterios de evaluación: *¿cuáles son los indicios de calidad que debemos considerar para poder afirmar que se tiene adquirida la competencia oral?* La determinación de los criterios es más compleja, y se requiere diferenciar, desde nuestro punto de vista, entre ambas subcompetencias (comprensión y expresión oral).

Probablemente la comprensión oral es más sencilla de evaluar, dado que pueden extenderse –a nivel de evaluación– los criterios habituales que suelen utilizarse en la evaluación de la comprensión lectora, adaptando en su caso, a las características propias del fenómeno: la emisión se realiza de forma oral, y habrá que considerar elementos sociolingüísticos y contextuales.

Respecto a la evaluación de la expresión oral el MCER (MECD, 2002) considera las siguientes categorías cualitativas:

Estrategias de turnos de palabra, estrategias de colaboración, petición de aclaración, fluidez, flexibilidad, coherencia, desarrollo temático, precisión, competencia sociolingüística, alcance general, riqueza de vocabulario, corrección gramatical, control de vocabulario, y control fonológico. (p.189).

En el mismo MCER (MECD, 2002) se señala la dificultad de aplicar los 14 criterios, por lo que se recomienda su selección y reformulación en criterios más simples. No obstante,

---

<sup>1</sup> Consejo Estadounidense para la Enseñanza de Lenguas Extranjeras



aunque el MCER ha supuesto una importante contribución a la evaluación de la lengua, presenta una serie de limitaciones, como no ser suficientemente extenso, coherente o transparente para un uso no crítico en las pruebas de lenguaje: las escalas de descriptores son insuficientes para tomar en cuenta la variación en términos de parámetros contextuales y en los parámetros de la validez basada en la teoría (por ejemplo, no definir de forma comprensiva el constructo a medir (Weir, 2005).

El tercer componente es la definición de un referente o Estándar de Ejecución, que permita establecer niveles diferenciales de desempeño: *¿Qué nivel de competencia ha alcanzado una determinada persona?* La especificación de estándares es un coadyuvante para la orientación de la definición del sistema de evaluación, y la elaboración de instrumentos y/o técnicas. En el MCER (MECD, 2002) se proponen aspectos técnicos de la descripción de niveles lingüísticos alcanzados: la formulación de descriptores (formulación positiva, precisión, claridad, brevedad e independencia) y metodologías para el desarrollo de escalas.

Desde nuestro punto de vista, la complejidad de la evaluación de la competencia oral puede verse minimizada si se establecen niveles de desempeño bien descritos, y posiblemente en los que se involucren características relativas a la complejidad del discurso (en diferentes contextos), en el caso de la comprensión oral; y a los elementos de descripción mencionados en el MCER para el caso de la expresión oral, especificando asimismo diversos contextos y usos sociolingüísticos.

Es una tarea compleja, que debe ser realizada por comités compuestos por especialistas en filología y profesorado; es recomendable, además, el concurso en dichos comités de antropólogos culturales, sociólogos, psicólogos y/o pedagogos, que coadyuven a establecer estándares objetivos, bien adaptados a los usos culturales del lenguaje, que sirvan para prevenir posibles sesgos.

Los componentes clave (el cuarto y el quinto de los señalados en la tabla II), la sistematización del proceso de recolección de evidencias y la determinación de los instrumentos y/o técnicas evaluativas a utilizar, los abordamos en esta reflexión de forma conjunta: *¿Qué situación o situaciones evaluativas vamos a definir para poder observar la competencia oral? ¿Qué instrumentos y/o técnicas podemos utilizar para su evaluación?* En la tabla III mostramos posibles formatos de situaciones evaluativas.

Tabla III. Modalidades de formatos de situaciones evaluativas

		Modo en que se presenta la tarea al sujeto		
		Oral	Escrito	Audiovisual
Modo en que debe ofrecer la respuesta el sujeto	Oral	A	B	C
	Escrito	D	E	F
	Ejecución Manual	G	H	I
	Acción	J	K	L

Fuente: Jornet y González-Such, 2010

*Las situaciones evaluativas adecuadas para la evaluación de la Comprensión Oral* podrían ser cualesquiera de las descritas en la tabla III. Obviamente, la presentación de tareas será más adecuada en la medida en que esté bien estandarizada la presentación de tareas, de modo que si pudiéramos optar, las formas más deseables serían la audiovisual (por su cercanía a la realidad y equidad) y la escrita (por su equidad). En la presentación oral podrían concurrir variables que produjeran sesgo o diferencias en las presentaciones entre sujetos, debidas al emisor o la situación. En cuanto a la forma de ofrecer respuesta, es claro que depende del tipo de tarea que se presente al sujeto, por lo que todas las opciones serán válidas siempre y cuando se ajusten de manera coherente con la forma en que se presentan las tareas.

En cualquier caso, criterios de operatividad/viabilidad de la evaluación pueden, en este caso, ser decisivos para optar por una u otra solución, dado que en todas ellas pueden encontrarse garantías de calidad para la realización de evaluaciones con instrumentos de calidad (técnicamente fiables y válidos).

Por su parte, en cuanto a las *situaciones evaluativas para evaluar la Expresión Oral*, sí que existen diferencias clave. Y se pueden identificar en la investigación consideraciones que conviene tener en cuenta. *El principio fundamental será que la situación evaluativa replique en la medida de lo posible la realidad*; es decir, no se trata de crear una situación artificial, basada en entrevistas estructuradas en las que los roles de evaluador/evaluado queden preestablecidas de forma estricta, sino que se oriente la evaluación a partir de un diálogo –dirigido, pero no directivo.

Como *elementos subsidiarios*, pero también importantes, para asegurar la calidad de la evaluación, deberán considerarse aspectos tales como:

- a) La igualación de estímulos (todos los sujetos deberían estar expuestos al mismo tipo de estímulos como tareas; de forma que serían preferibles formatos en los que la presentación fuera audiovisual, o estuviera establecida con un guión muy bien sistematizado... en ambos casos nos referimos a situaciones bien estandarizadas).
- b) La diversificación de situaciones sociolingüísticas (basadas en diversidad de temáticas de diálogo, con variación de emisores).
- c) Estandarización (objetivación) de la forma de valorar las respuestas de los

sujetos (basadas en rúbricas bien definidas, operativas, representativas de los diversos elementos que se contemplen en los estándares de evaluación).

Todos los elementos mencionados están presididos por un intento de acercamiento a la evaluación como una función de calidad técnica y ética, en la que se tenga presente la equidad como un principio implícito en la selección de situaciones evaluativas, instrumentos y técnicas de evaluación.

Permítasenos, en este punto, realizar algunas reflexiones respecto a formatos habituales de prueba en este contexto. Podemos definir los distintos tipos de interacción que se dan en la ejecución de las pruebas en el entorno de aprendizaje de las lenguas: si consideramos el tipo de canal que se utiliza, el evaluador puede utilizar el canal oral y la respuesta del candidato puede ser oral. Este tipo de pruebas son útiles para evaluar aspectos en los que se desea evaluar la calidad de la interpretación que el alumno realiza de una situación, las decisiones que tomaría ante un problema determinado y los motivos que las avalan y la capacidad del sujeto para dar respuesta a una cuestión en la que se requiere atención y rapidez de ejecución. Este tipo de formato favorece en nuestro contexto la identificación de aspectos fonológicos, sintácticos y gramaticales. Este formato es similar a la entrevista.

Un segundo tipo se da cuando el evaluador realiza las preguntas de forma escrita y la respuesta del evaluado se produce de manera oral. Es útil para analizar la capacidad del estudiante en el análisis de casos, las decisiones que tomaría ante una situación determinada y los motivos para estas decisiones y la capacidad del alumno para dar respuesta a una situación en la que debe dar una respuesta rápida a partir de la información suministrada. Este formato es similar al de una entrevista estructurada o un cuestionario.

Otra opción puede ser que las cuestiones se planteen de forma audiovisual y las respuestas las ofrezca el alumno oralmente. Este formato incluye generalmente preguntas acerca de una grabación de video o audio que el alumno debe responder de acuerdo con la comprensión del material expuesto.

El siguiente formato se basa en que el evaluador realice sus cuestiones (preguntas, texto, etc.) de forma oral y la respuesta del alumno sea también oral. Es útil para analizar la comprensión oral del candidato, implicando su atención, para analizar la capacidad del alumno para identificar los elementos clave de la situación planteada y para observar la forma en que el sujeto plantea la respuesta a las cuestiones.

Se han analizado distintas técnicas para la evaluación de L2, aunque todas se pueden integrar en la clasificación anterior, cada una de ellas tiene sus propias ventajas y desventajas: exposiciones orales, entrevista/conversación, preguntas y respuestas, debate, *role plays*, dramatizaciones, dictado, dictado graduado, dictado basado en el ordenador, pruebas de traducción de resúmenes como el Listening Translation Exam<sup>2</sup> (LSTE), grupos orales, etc., cada una presenta ventajas e inconvenientes (Kaga, 1991;

---

<sup>2</sup> Examen de traducción (Traducción de los autores).

Coniam, 1998; Scott *et al.*, 1996; Bonk y Ockey, 2003, Martín y Ramírez, 2007; Bordón, 2007b).

La elección y el diseño de las tareas de evaluación afectan a las fases de generalización y extrapolación de la validación (Roever, 2011). Así, se ha producido un aumento en el uso de tests orales para la evaluación de L2 en todas formas y tamaños, debido al aumento de la interpretabilidad de las puntuaciones de los tests, la validez potencial de las puntuaciones cuando se utilizan criterios vinculados al mundo real y los efectos positivos del washback<sup>3</sup> de estos instrumentos (Bonk y Ockey, 2003).

En cualquier caso, un aspecto fundamental es que debemos conseguir que las pruebas repliquen la realidad, es decir, que la situación se convierta en un diálogo entre el evaluador y el evaluado, y que ello no se convierta en un mero interrogatorio. En este sentido, cabe recordar que la entrevista se define como una relación interpersonal asimétrica, en la que el entrevistador tiene un rol que es percibido por el entrevistado de forma diferente a un diálogo entre iguales. Martín y Ramírez (2007) señalan algunos aspectos a considerar, como conseguir que la evaluación de la expresión oral siga un formato de diálogo va a favorecer que el alumnado utilice sus propias expresiones. Así, el profesor/evaluador se convierte en un facilitador, dejando de ser un simple interlocutor. Otros aspectos a considerar son la utilización de distintos tipos de actividades y de técnicas para atender los estilos de aprendizaje individuales; minimizar el efecto del contexto en el que se desarrolla la evaluación, que pueden hacer que el candidato se distraiga por distintos estímulos contextuales; utilizar pruebas útiles, justas y transparentes, informando al candidato de los distintos aspectos de la evaluación que se realiza; aplicar los criterios de viabilidad, fiabilidad y validez en el proceso de evaluación y utilizar unos criterios de evaluación y calificación. Las tareas para evaluar la expresión oral y la interacción deberían responder a tareas de uso real de la lengua, especialmente en pruebas de nivel de dominio, para lo que se pueden utilizar materiales gráficos, descripciones, etc. (Bordón, 2007b).

Los componentes sexto y séptimo (Definir un sistema de comparación de evidencias con el Estándar, y Determinación del proceso de formación de la decisión): *¿Cómo puntuamos el desempeño en cada tarea? ¿Cómo puntuamos el desempeño global en la competencia?* Son componentes esencialmente técnicos, en los que se juega la fiabilidad de los instrumentos y/o técnicas de evaluación, así como afectan a su validez. Los sistemas de interpretación de puntuaciones deben estar enmarcados en procesos bien establecidos y sólidamente constituidos desde análisis psicométricos sólidos. Incluyen desde las rúbricas de evaluación (elementos observacionales que se van a considerar para valorar el desempeño de los sujetos en cada tarea), hasta los estándares de interpretación y definición de los puntos de corte para asignar a cada persona evaluada su nivel de competencia de acuerdo con los estándares de calidad predefinidos. No vamos a extendernos en este punto, pues se trata de un ámbito esencialmente técnico-métrico y que ha sido tratado in extenso en diversos trabajos de índole general, por lo que remitimos a los lectores a dichas propuestas (Jornet y González-Such, 2009; Jornet, González-Such y Suárez, 2010).

---

<sup>3</sup> Retroalimentación, influencia (Traducción de los autores).

No obstante, realizamos algunas consideraciones al respecto. El concepto de estándares tiene matices relacionados con el tipo de instrumento al que se refieren (Jornet y González-Such, 2009; Jornet, González-Such y Suárez, 2010). Es frecuente su uso en las pruebas estandarizadas para la interpretación de las puntuaciones, aunque también se ha utilizado en aproximaciones cualitativas, en las que actúan como un argumento de referencia para poder explicitar los criterios de evaluación que faciliten la comparación de la ejecución que muestra una persona con un referente que operativiza la evaluación de la calidad del aprendizaje observado. Por estándar nos referimos a la especificación de los niveles de calidad que deben mostrarse como dominio de una competencia (Jornet *et al*, 2011).

Alderson y Banerjee (2001) señalan tres conceptos asociados a la utilización de la palabra estándares, derivados de la *Task Force on Language Testing Standards*<sup>4</sup>. El primero se refiere a procedimientos para asegurar la calidad, como “códigos de la práctica”. El segundo significado son “niveles de habilidad”, mientras que el tercer significado está relacionado con las pruebas estandarizadas, es decir, en este caso los estándares son equivalentes a “normas”.

En este sentido, el MCER es fundamental para el establecimiento de una escala común de referencia y comparación. No se trata de un conjunto de escalas, sino más bien de un conjunto en el que se indica qué se conoce sobre el aprendizaje, uso y habilidades lingüísticas, además de una guía para elaborar materiales docentes, especificaciones de las pruebas y criterios de puntuación, y para la formación de los profesores (Alderson y Banerjee, 2001). Estos autores ya auguraban: “We are confident that we will hear much more about the Common European Framework in the coming years, and it will increasingly become a point of reference for language examinations across Europe and beyond<sup>5</sup>” (Alderson y Bernerjee, 2001, p. 220), como así ha sido, el MCER se está convirtiendo en un referente surgido de la cooperación para el aprendizaje, uso y conocimiento de los aspectos lingüísticos, aunque no ha sido un camino fácil y sin dificultades (Papageorgiou, 2010).

Figueras (2008) analiza el impacto del MCER, indicando que los motivos de su éxito radica en tres aspectos: el primero es la solvencia de la institución que lo publica; el segundo es de naturaleza estratégica, relacionada con la naturaleza del MCER, que combina política educativa y didáctica, destinado a reconocer los esfuerzos que se producen en los distintos países en el campo de la docencia, el aprendizaje y la evaluación de las lenguas. El tercer aspecto es académico-pedagógico, por la pertinencia y la necesidad de un enfoque orientado a la acción, en el que la lengua se considera como un instrumento de comunicación en sentido amplio. A estos tres aspectos añade la oportunidad en el tiempo, en un momento de unión de la Europa multilingüe.

---

<sup>4</sup> Grupo de Trabajo sobre Estándares de Evaluación de la Lengua (Traducción de los autores).

<sup>5</sup> Estamos seguros de que vamos a escuchar muchas más sobre el Marco Común Europeo en los próximos años, y que su uso se incrementará como un punto de referencia para los exámenes de idiomas en toda Europa y más allá (Traducción de los autores).

La implementación de una puntuación de corte para la evaluación de estudiantes en el sistema universitario ha sido propuesta recientemente, con un aumento de la investigación sobre pruebas orales de grupo. Van Moere (2006) aplica un test oral de grupo para investigar el adecuado uso de puntuaciones para evaluar el logro de los examinados. Las decisiones sobre los puntos de corte basadas en una puntuación simple de test deberían tener en cuenta un amplio margen para el error, proporcionando al candidato el beneficio de la duda para ser evaluado en un grupo que puede tener restringido su desempeño, o que las percepciones de los evaluadores de su habilidad puedan estar afectadas por otros factores. El beneficio de poder observar a los candidatos en una interacción natural entre ellos, permitiéndoles dirigir y controlar su propio discurso, puede ser un riesgo para la fiabilidad de las puntuaciones. No obstante, las ventajas potenciales para las evaluaciones de grupo, y el hecho de que los candidatos tengan reacciones positivas hacia esta forma de evaluación, hacen que sea necesario profundizar en la investigación de este tipo de aplicaciones (Van Moere, 2006).

Estándares y puntuaciones de corte constituyen en sí mismos un componente clave, pues afectan directamente a la interpretación del logro en la competencia evaluada y, por ello, debe ser considerado como un ámbito de especial atención para investigadores y evaluadores en esta área; máxime si se trata de pruebas, como las Pruebas de Acceso a la Universidad (PAU), a partir de las cuales se determina la selección de estudiantes para su acceso a las titulaciones universitarias.

El componente octavo de la definición de diseños de evaluación de competencias (Especificar los mecanismos de información a audiencias y/o retroalimentación para el estudiante) es también de vital importancia. Se trata de poder informar al alumnado acerca de su nivel de competencia y, asimismo, es deseable –por no decir imprescindible– que se les informe acerca de sus lagunas de aprendizaje. Este aspecto es frecuentemente desatendido. Los usos de la evaluación se circunscriben a utilidades sumativas a partir de las cuales sólo se les indica su puntuación final o si pasa o no-pasa. Cualquier evaluación, aunque sea para la selección, debería atender cuidadosamente estos aspectos de información para el alumnado. Ello debería incorporarse como una práctica habitual. Para este cometido, es fundamental referirse a los estándares, en los que se deberán haber definido las características de los sujetos que, según su nivel de desempeño, son identificables en cada nivel.

Por último, en cuanto a los componentes de evaluación, el noveno se refiere al control de calidad del propio proceso de evaluación. ¿La evaluación que realizamos es de calidad? ¿Contiene error? ¿Hasta qué punto es justa? ¿Podemos asumir una puntuación de corte de manera indudable o existen incertidumbres? Este tipo de preguntas difícilmente las respondemos con las prácticas habituales de evaluación que se dan en España. Incluso en procesos de selección, como son las PAU, éstas no se estudian. Es el profesorado quien fija sus criterios y se convierte en el único garante de calidad de la decisión que se toma. Ello no es una práctica justa ni equitativa, ni mucho menos deseable en un sistema educativo de un país democrático que pretende alinearse con los más avanzados.

En este sentido, cabe reseñar que la utilización de soportes físicos, como grabación en video o audio, otorgaría mayor credibilidad a este tipo de evaluaciones, dado que cabría la posibilidad de revisar la sesión por evaluadores externos. El estudio acerca de la calidad de los sistemas de evaluación es fundamental, no sólo como una característica vinculada a sus garantías técnicas, sino como expresión de la ética con que se realiza la evaluación. Es una asignatura pendiente en nuestro país que habrá que abordar de forma decidida. Sin embargo, ello sólo es posible si se profesionaliza la evaluación, reconociendo el rol fundamental que tiene la metodología en esta disciplina.

### **III. Validez de la evaluación de la competencia oral**

La validez es la característica esencial de calidad de cualquier instrumento o técnica de medida/evaluación. Se refiere a si el instrumento o la técnica mide realmente lo que pretende medir, o bien, los resultados observados dependen de otras variables (Nunnally, 1978; Magnusson, 1990). Junto a la fiabilidad, se constituye en la garantía técnica (y ética) de uso de las evaluaciones. En este sentido, hablamos de un proceso continuo de validación, nunca de una cuestión definitiva, que puede ser modificada por distintas evidencias (Messick, 1989; 1994; 1996). En esencia, no validamos el instrumento, sino si los resultados son útiles para el fin para el que ha sido construido el instrumento. La validación se presenta, pues, como un proceso de recogida de evidencias y de información relevante, un proceso permanente que puede incluir diversos enfoques tradicionales e innovadores de investigación evaluativa (Kunnan, 1998; Chapelle, 1999; Teasdale y Leung, 2000; Alderson y Banerjee, 2002)

Un aspecto central de los estudios de validación es que la validez de cualquier instrumento puede verse afectada por la influencia de diversas variables, entre ellas los evaluadores. La formación especializada en el campo de la lengua que se quiere evaluar es fundamental. En este sentido existen diversas evidencias relativas al desempeño diferencial en las pruebas orales en función diversas variables. En tabla IV se recogen algunas de las variables que se han estudiado a estos efectos, así como algunos de los estudios en los que se han analizado.

Tabla IV. Variables asociadas al desempeño observado en pruebas orales, y estudios en los que se han analizado

Variable	Resultados	Referencias
Personalidad	Timidez. Los sujetos no tímidos obtuvieron una puntuación algo mejor que los tímidos.	Bonk y Van Moere, 2004
Severidad del profesor que evalúa	Se han encontrado diferencias significativas entre jueces.	Lynch y McNamara, 1998; Weigle, 1998; Bonk y Ockey, 2003
Entrenamiento de los evaluadores y <i>feedback</i>	Aumento de la fiabilidad inter e intra.	Weigle, 1998; Bonk y Ockey, 2003
Características de los interlocutores y las dinámicas de interacción dentro del grupo	La ejecución de los candidatos, o cómo los jueces perciben la habilidad individual de los candidatos, puede estar afectada de forma importante.	Van Moere, 2006
Familiaridad del evaluador con la pronunciación del candidato o la localización del centro de examen	Una proporción significativa de examinadores puntuaron mejor la pronunciación cuando habían estado expuestos de forma prolongada al interlenguaje de los candidatos, y peor cuando no era así. La localización del centro del test también tuvo un efecto significativo sobre las puntuaciones de pronunciación, independientemente de la variable familiaridad. Recomienda que la familiaridad con la fonología de interlenguaje sea considerada en el diseño de tests hablados.	Carey, Mannell y Dunn , 2011
Género	No hay diferencias significativas	O'Loughlin, 2002
La influencia del entrevistador sobre el candidato	En variables como la cantidad de soporte que se le proporciona a los candidatos, el tipo de <i>rapport</i> (empatía, relación) que se establece entre ellos o la forma en que siguen las instrucciones relevantes a su rol.	Young y Milanovic, 1992; Lazaraton, 1996; McNamara y Lumley, 1997; O'Loughlin, 2002.

En conjunto, no sólo se trata de asegurar la validez de constructo al diseñar los instrumento y/o técnicas evaluativas, sino también de investigar, mediante diversas aproximaciones, cuáles son los elementos o variables relacionadas, con el fin de poder mejorar la validez general, mediante estudios de validez convergente, concurrente y predictiva.

#### IV. Conclusiones

Diseñar un sistema de evaluación de la competencia oral es una tarea compleja, cuya dificultad comienza en la definición del constructo a evaluar. Los estudios y referencias (incluidos los aportados en el MECR, 2002), ponen de manifiesto que la tarea no sólo es compleja, sino que su desarrollo debe ser cuidadoso, y extremadamente bien cuidado tanto a nivel del planteamiento teórico de la variable a evaluar como en relación a la elaboración de las técnicas que utilizemos para su evaluación.



Buena parte de las dificultades que podemos encontrar se pueden abordar si analizamos la competencia oral, desde un planteamiento micro-analítico, diferenciando la *comprensión oral* de la *expresión oral*. En el primer caso, las soluciones pueden ser más fácilmente resueltas, mientras que en el segundo la dificultad se mantiene.

Somos conscientes que las aportaciones de la investigación, algunas de las cuales se han ido reseñando a lo largo de este trabajo, suponen un apoyo indudable para mejorar la evaluabilidad de esta competencia; sin embargo, también es cierto que el carácter del constructo a evaluar implica dificultades intrínsecas que van a requerir esfuerzos especiales en su investigación si queremos mejorar sus modos de evaluación.

La evaluación de la competencia oral no está exenta de problemas asociados. El efecto *washback* impacta de forma directa en la enseñanza (Wall, 2000; Alderson y Benerjee, 2001; Fernández Alvarez y Sanz, 2005; Muñoz y Álvarez, 2010), máxime cuando el origen se produce en relación a pruebas de certificación/admisión –como son las PAU españolas o las pruebas de nivel.

La creciente tendencia en España de la enseñanza plurilingüe, junto con la presencia cada vez mayor de estudiantes de otras culturas, hace que sean necesarios estudios sobre la forma en que se evalúa la enseñanza de la L2. Se identifican tres áreas principales como clave de este cambio de paradigma:

- Se debe garantizar la evaluación correcta y equitativa de estas competencias (Solano-Flores y Trumbull, 2003).
- Es necesario apoyarse en tests adaptativos, facilitados por las nuevas tecnologías de la información y la comunicación.
- En el marco de un uso formativo de la evaluación, es necesario aprovechar las aportaciones de los enfoques basados en autoevaluación del lenguaje. La participación del estudiante en el proceso de evaluación propicia que tenga una visión clara de éste, aumente su implicación en el proceso y favorezca su rendimiento.

Por último, queremos hacer mención de la equidad en la evaluación. Las dificultades de la competencia oral como constructo a evaluar, nos llevan a enfatizar este aspecto que afecta a la validez, y es centro de interés en una perspectiva ética de la evaluación. Aunque ha sido atendido, precisamente por estos motivos, desde diversas perspectivas (Xi, 2010), lo cierto es que es necesario que en España se produzca un cambio al abordar este tipo de evaluaciones. No sólo se trata de mejorar la validez de las PAU o de las pruebas de nivel, sino en general de la evaluación de estudiantes, y ello pasa necesariamente por prestar una especial atención al estudio de las pruebas en sí mismas y a las competencias del evaluador en el área evaluada. Si no analizamos la calidad de las pruebas, estamos dejando de lado la atención a la equidad, la justicia y la ética en la evaluación.

## Referencias

- Alderson, J. y Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34, 213-236.
- Alderson, J. y Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- Bonk, W. y Ockey, G. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20(1), 89-110.
- Bonk, W. J. y Van Moere, A. (2004). *L2 group oral testing: the influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on individual scores*. Documento presentado en el Language Testing Research Colloquium.
- Bordón, T. (2007a). La evaluación de la actuación oral en aprendices de EL2 mediante cuadros o escalas de niveles. En E. Balmaseda (Coord). *Las destrezas orales en la enseñanza del español L2-LE*. (pp. 97-112). Logroño: Asociación del Español como lengua extranjera (ASELE).
- Bordón, T. (2007b). La evaluación de la expresión e interacción orales. En B. Montmany y N. Sans (Eds.). *XVI encuentro práctico de profesores de ELE* (pp.18-23). Barcelona: International House Barcelona y Difusión.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, 15(1), 45-85.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Carey, M. D., Mannell, R. H. y Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28(2), 201-219.
- Chapelle, C. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Coniam, D. (1998). Interactive evaluation of listening comprehension: How the context may help. *Computer Assisted Language Learning*, 11(1), 35-53.
- Crystal, D. (1997). *A dictionary of linguistics and phonetics*. Oxford: Basil Blackwell.
- Fernández Álvarez, M. y Sanz Sainz, I. (2005). Metodología para el diseño de una prueba de inglés en selectividad. En H. Herrera Soler y J. García Laborda (Coord.). *Estudios y criterios para una selectividad de calidad en el examen de inglés*. Valencia: Universidad Politécnica de Valencia.

Figueras, N. (2008). El MCER, más allá de la polémica. *Monográficos Marco ELE*, 7. Recuperado de <http://marcoele.com/descargas/evaluacion/03.figueras.pdf>

Fulcher, G. y Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing*, 20(3), 321-344.

Jornet, J. M. y González-Such, J. (2009). Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, 16, 103-123.

Jornet, J. M., González-Such, J. y Suárez, J. M. (2010). Validación de los procesos de validación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, 19, 11-29.

Jornet, J. M. y González-Such, J. (2010). *Métodos de Construcción de Pruebas*. Material de aula, Métodos de Investigación y Diagnóstico en Educación. Valencia: Universidad de Valencia.

Jornet, J. M., González-Such, J., Suárez, J. M. y Perales, M. J. (2011). Diseños de procesos de evaluación de competencias: consideraciones acerca de los estándares en el dominio de las competencias. *Bordón*, 63(1), 125-145.

Kaga, M. (1991). Dictation as a measure of Japanese proficiency. *Language Testing*, 8(2), 112-24.

Kim, Y. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26(2), 187-217.

Kunnan, A.J. (ed.) (1998). *Validation in language assessment*. Mahwah, NJ: L. Erlbaum.

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

Leech, G. (1983). *Principles of pragmatics*. Londres: Longman.

Liskin-Gasparro, J. E. (2007). La evaluación de las destrezas orales: Perspectivas norteamericanas. En E. Balmaseda (Coord). *Las destrezas orales en la enseñanza del español L2-LE* (pp. 19-37). Logroño: Asociación del Español como lengua extranjera (ASELE).

Lynch, B. y McNamara, T. (1998). Using G theory and Many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15, 158-80.

Magnusson, D. (1990). *Teoría de los tests*. México: Trillas

Martín, L. y Ramírez, M. (2007). La evaluación de las destrezas orales en la educación superior: estudio de dos universidades británicas. En E. Balmaseda (Coord). *Las*

*destrezas orales en la enseñanza del español L2-LE* (pp. 811-822). Logroño: Asociación del Español como lengua extranjera (ASELE).

McNamara, T. F. y Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing*, 14, 142–51.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). Nueva York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-56.

Ministerio de Educación, Cultura y Deporte (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Secretaría General Técnica del MECD-Subdirección General de Información y Publicaciones y grupo Anaya.

Muñoz, A. P. y Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49.

Nunnally, J.C. (1978). *Teoría Psicométrica*. México: Trillas.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, 19, 169-192.

Papageorgiou, S. (2010). Investigating the decision-making process of standard setting participants. *Language Testing*, 27(2), 261-282.

Pinilla, R. (2007). La competencia estratégica del estudiante de ELE en su producción oral: algunas consideraciones para su evaluación. En E. Balmaseda (Coord). *Las destrezas orales en la enseñanza del español L2-LE* (pp. 89-96.). Logroño: Asociación del Español como lengua extranjera (ASELE).

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463-481.

Scott, M. L., Stansfield, C. W. y Kenyon, D. M. (1996). Examining validity in a performance test: The listening summary translation exam (LSTE). *Language Testing*, 13, 83-109.

Solano-Flores, G. y Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of english-language learners. *Educational Researcher*, 32(2), 3-13.

Teasdale, A.; y Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing*, 23(4), 411-440.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.

Weigle, S. (1998). Using FACETS to model rater training effects. *Language Testing*, 15, 263-87.

Weir, C.J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22(3), 281-300.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170.

Young, R. y Milanovic, M. (1992). Discourse validation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 403-24.