

ELIA 2, 2001

PROBLEMAS EN LA EVALUACIÓN DE LA COMPETENCIA DISCURSIVA EN INGLÉS COMO L2.

Rubén Chacón Beltrán
Universidad de Sevilla

Nowadays, language testing is considered a fundamental component in foreign language teaching and learning. In this paper, some of the most representative features of language testing are described and special attention is given to the question of language test validity. Some well-known language-testing institutions use the native speaker (NS) as a stakeholder to evaluate the overall level in the foreign language, without taking into account the feasibility of such a concept. In this regard, some arguments are presented in order to demonstrate the ambivalence of the notion of “native speaker” in language learning evaluation. Particular attention is paid to international language-testing standards that, although well-known and widely used, may not be as appropriate as they are assumed to be. Some suggestions are made for the field of bilingualism studies.

1. Situación actual de la evaluación lingüística

La evaluación del conocimiento lingüístico en una L2 es un fenómeno complejo que ha cobrado especial interés en los últimos años. Numerosas son las publicaciones periódicas relacionadas con este tema que han surgido o han incrementado su importancia, como *Language Testing*, *Fair Test* y *Language Testing Update*. El mayor énfasis por averiguar cuál es la eficacia del aprendizaje por medio de pruebas de rendimiento, o pruebas de aptitud, se centra principalmente en las exigencias institucionales que instan, tanto a profesores como a alumnos, a cumplir con unos objetivos curriculares. En el aprendizaje institucional de una L2, la evaluación formal se vislumbra como un componente fundamental.

La evaluación, en cualquiera de sus modalidades, supone la recogida sistemática de información que posteriormente se utilizará para tomar decisiones relativas a la enseñanza y el aprendizaje, y suele tener consecuencias trascendentales para el futuro de los alumnos ya sea por la admisión o exclusión en un determinado curso, o la concesión de una nota

que representa el grado de progreso alcanzado. Aunque la evaluación del aprendizaje puede encauzarse tanto desde una perspectiva cualitativa como cuantitativa, es ésta última la que en la actualidad se practica con mayor frecuencia, sobre todo por su conveniencia al trabajar con grupos numerosos. Mientras que la evaluación cualitativa cuenta con la observación como instrumento de medida, la evaluación cuantitativa emplea la prueba o el *test*.

2. Características generales de la evaluación lingüística

En cualquier prueba de aptitud lingüística, dos son las cualidades esenciales a las que se debe prestar especial atención: la **fiabilidad** y la **validez**. La fiabilidad es la medida del grado en que una prueba proporciona resultados consistentes, es decir, se dice que **una prueba es fiable si da resultados similares cuando se administra en ocasiones diferentes o cuando la usan personas distintas** (Richards, Platt & Platt:1997). La fiabilidad se encuentra condicionada por la consistencia de la medición en diferentes momentos, con diferentes tipos de examen, y con diferentes evaluadores. Por otra parte, la validez es **el grado en que una prueba mide lo que se supone que tiene que medir**, es decir, el grado en que puede emplearse con buenos resultados para los fines deseados. Hay varios procedimientos estadísticos que pueden aplicarse a una prueba para estimar su validez, y con ello la bondad de la medición (Richards, Platt & Platt:1997). Se debe tener en cuenta que ambas cualidades son complementarias pues, mientras que la **fiabilidad** es una cualidad de la puntuación o calificación, la **validez** lo es de la interpretación y uso de las pruebas. Para que la evaluación suponga un instrumento preciso, ambas cualidades son necesarias, si bien en este trabajo nos centraremos en la validez como componente fundamental de una evaluación fehaciente.

La competencia discursiva representa uno de los cuatro componentes de la competencia comunicativa que describe la habilidad para producir discurso, tanto oral como escrito, comprensible y coherente en los diferentes géneros (ej. cartas de negocios, ensayos científicos,...) (Richards, Platt & Platt, 1997). Para la elaboración de pruebas destinadas a medir la competencia discursiva, ya sea a gran escala o para grupos pequeños, se deben tener en cuenta varios factores determinantes pues la evaluación es

intrínsecamente problemática debido a que persigue la medición de capacidades cognitivas y metacognitivas. Además, debemos tener en cuenta que, de entre todos los componentes de la competencia comunicativa, la competencia discursiva representa posiblemente la capacidad menos estudiada en el aprendizaje de lenguas por las dificultades que ello entraña. La evaluación de la competencia discursiva trata tanto capacidades mentales como contextuales, y por ello es necesariamente indirecta, incompleta, imprecisa, subjetiva y relativa (Bachman, 1990). Analicemos cada una de estas características por separado:

A. Indirecta. La actuación de los alumnos en las pruebas o *tests* supone un indicador indirecto de las características o cualidades subyacentes que, como lingüistas, nos interesa medir. Por ejemplo, para averiguar si el aprendiz es capaz de utilizar adecuadamente un vocabulario previamente introducido en clase, se le puede pedir que escriba una redacción, o que haga un ejercicio de rellenar espacios en blanco. Sin embargo, en este tipo de pruebas intervienen otras cualidades individuales completamente ajenas al manejo de la L2, como pueden ser destrezas especiales para escribir de forma creativa, o la capacidad de tratar y relacionar elementos léxicos descontextualizados.

B. Incompleta. La competencia/actuación que se puede observar a partir de una prueba representa únicamente un pequeño porcentaje de las capacidades lingüísticas que posee el individuo. En la mayoría de los casos se usan pruebas escritas de una o dos horas de duración para determinar el conocimiento global del alumno en la L2. En lo concerniente a la destreza oral, ese tiempo se puede ver reducido a una entrevista de 10 ó 15 minutos.

C. Imprecisa. Una tarea no siempre mide eficazmente la destreza o habilidad que pretende analizar. Puede ocurrir que la tarea se sitúe por encima o por debajo del nivel de competencia del alumno, y que proporcione información poco precisa. Este problema es más frecuente de lo que sería deseable, ya que en evaluación cuantitativa raras veces se hacen pruebas personalizadas sino que, por el contrario, se busca un punto medio de dificultad, y en relación con éste se valora la competencia y/o actuación del grupo.

D. Subjetiva. Quienes diseñan las pruebas constantemente toman decisiones subjetivas en cuanto al diseño y al contenido de las mismas, ya sea en lo referente a la temática o las destrezas a las que se concede mayor importancia. Del mismo modo, las pruebas lingüísticas están basadas bien en contenidos curriculares o en teorías del aprendizaje, las cuales, a su vez, también son subjetivas.

E. Relativa. La relatividad de las pruebas lingüísticas es una de las características en las que quisiera centrarme. A pesar de que se suele medir en términos absolutos, la estimación de las habilidades lingüísticas en una L2 es relativa y parcial. Normalmente se contrastan las habilidades productivas en la L2 con una escala numérica progresiva. Con frecuencia, se toma al Hablante Nativo Culto (HNC) como criterio absoluto de habilidad lingüística y se le sitúa en el punto más alto de la escala. Esto se hace pasando por alto las tremendas variaciones individuales apreciables en cuanto a cohesión del discurso, organización de ideas, o comportamiento sociolingüístico, etc... Sin embargo, la experiencia demuestra que, como en cualquier otra habilidad cognitiva, no existen términos absolutos y por ello no hay un individuo que posea el conocimiento al 100%, fuera de un contexto teórico e idealizado. Difícilmente se podría caracterizar a un individuo con una competencia discursiva absoluta ni en la L1 ni en la L2.

3. El HNC como modelo en la evaluación

En la actualidad, se puede apreciar una clara tendencia a estandarizar todo lo concerniente a la evaluación del aprendizaje lingüístico. Prueba de ello son las diversas organizaciones creadas recientemente con el fin de mejorar los procesos evaluadores y, en última instancia, elaborar pruebas de aptitud lingüística con alcance internacional. Tras esta tendencia también se encuentra el objetivo de crear unos criterios *estándar*, que tengan validez universal independientemente de la lengua con que se trate. Por ello, se han establecido asociaciones internacionales como ILTA (*International Language Testing Association*) o ALTE (*Association of Language Testers in Europe*). A pesar de los esfuerzos realizados por estas asociaciones, la evaluación del inglés como L2 se encuentra influida en gran medida por las tendencias imperantes en los EE.UU. En la base de dicha corriente,

encontramos el enfoque de la competencia/actuación (*Proficiency Movement*) que se emplea particularmente para evaluar el uso oral de la lengua inglesa. El *Proficiency Movement* abarca tanto la enseñanza como la evaluación de lenguas extranjeras y se encuentra representado por la **ILR** (*Interagency Language Roundtable*), institución precursora de las **ACTFL Guidelines** (*American Council on the Teaching of Foreign Languages*), que representan las directrices empleadas en la actualidad por instituciones estatales para la evaluación lingüística. Las aspiraciones de este movimiento son altas, como queda patente en la siguiente cita de Liskin-Gasparro:

The guidelines promise to be a “common yardstick” for the foreign-language profession that will allow us to organize much needed proficiency-based curricula, materials, evaluation instruments, and articulation plans around well-defined levels of language skill. (Liskin-Gasparro, 1984:13)

La principal ventaja de las *ACTFL Guidelines* radica en que proporcionan una prueba de evaluación en la que el rendimiento se mide de acuerdo con un **criterio establecido** previamente (*criterion referenced test*). Por este motivo, el estudiante debe alcanzar un nivel de competencia-actuación mínimo para pasar dicha prueba, es decir la puntuación se interpreta como indicador de los logros obtenidos en relación con una serie de habilidades fijadas con anterioridad, al igual que ocurre con las pruebas para obtener el carnet de conducir. En contraste con éstas, se encuentran las de **norma de grupo** (*norm referenced test*) en las que se compara el rendimiento de un alumno con el rendimiento global de otro grupo cuyas puntuaciones se han establecido como norma base, por ejemplo el *TOEFL*, para el que se establece una norma en relación con la puntuación obtenida por un grupo seleccionado de alumnos.

La escala utilizada por la *ILR* (*Interagency Language Roundtable*) para medir el grado de competencia discursiva oral de los aprendices, oscila entre los niveles 0 y 5, entre lo que determinan como 0 ó la carencia absoluta de habilidad alguna para comunicarse en la lengua, y 5 que se aplica a quien interactúa funcionalmente igual a un HNC de la lengua meta. El tiempo

requerido para este tipo de pruebas suele oscilar entre los 10 y los 40 minutos de intervención por alumno. Además, en su descripción se especifica que el nivel máximo (5) o equivalente al HNC **se obtiene a largo plazo**, a través de la familiarización con varios tipos de registros idiomáticos y grupos sociales en un período extenso que transcurre desde la infancia a los estudios de postgrado. En dichas directrices para la evaluación, se matiza que la experiencia acumulada a lo largo de años de aplicación de dichas pruebas, demuestra la **imposibilidad de todos los Hablantes Nativos (HN) para alcanzar el nivel máximo** que normalmente corresponde a individuos con una titulación académica. Paradójicamente se especifica que la mayoría de los HN probablemente tiene una competencia equivalente al nivel 3.

4. Problemas con el HNC como modelo de competencia discursiva

Las referencias al HNC como estándar son explícitas en las definiciones de ILR (*Interagency Language Roundtable*). Las directrices de *ACTFL Guidelines*, que suponen una evolución de la anterior y que han sido perfeccionadas a lo largo de años de aplicación, han evitado las referencias al HNC aunque éste continúa presente de forma implícita. Autores como Valdman (1989) y Barnwell (1989) comparten este punto de vista. La realidad es que incluso las *ACTFL Guidelines* continúan arraigadas en la creencia de que el HN se puede caracterizar sin ningún problema y que indiscutiblemente debe estar presente entre los criterios de evaluación.

Como ya se indicó anteriormente, los ideadores de este tipo de prueba posiblemente se encuentran en lo cierto al preferir un **criterio establecido** a una prueba de **norma de grupo** por proporcionar mayor objetividad en los resultados. Sin embargo, teniendo en cuenta lo complicado que resulta medir cualquier tipo de actividad cognitiva y el hecho de que difícilmente se puede caracterizar a un individuo con una habilidad lingüística absoluta y completa, es improbable que nadie se sitúe en lo alto de dicha escala. A pesar de estas consideraciones, el HN se emplea como modelo comparativo, convirtiéndose así en un criterio inadecuado. La experiencia nos demuestra que independientemente de la variedad o registro lingüístico por la que se opte, el grupo de HN no es homogéneo. A esto deberíamos añadir el debate suscitado, desde hace ya más de dos décadas,

sobre el tratamiento del concepto HN en el campo de la lingüística (Coulmas, 1981). Dicha dialéctica ha girado en torno a la necesidad de proporcionar una definición integral del HN que sea útil, tanto en el ámbito teórico como aplicado, y que englobe matices psicolingüísticos, sociolingüísticos y pragmáticos. Por el momento, más de 20 años después de que se iniciara dicho debate, no hay indicios para pensar que la completa caracterización del HN esté próxima, lo que se debe a la gran cantidad de enfoques que se pueden adoptar, y a la interdisciplinariedad necesaria para conseguir un concepto sólido. Asimismo, resulta muy difícil, tanto desde un punto de vista teórico como práctico, definir un nivel de conocimiento lingüístico absoluto o modélico (Bachman y Savignon, 1986; Bachman, 1990), o que al menos presente homogeneidad entre los miembros de una comunidad lingüística. Esta postura se encuentra corroborada por autores como Hamilton *et al* (1993) quienes identifican una tendencia bastante generalizada en los últimos años para desarrollar pruebas de aptitud dirigidas a HNN, basadas en la comparación con el discurso del HN. Tras esta práctica se encuentra nuevamente la idea de que el grupo de HN es homogéneo y fácilmente caracterizable. Numerosos son los trabajos dedicados a contrastar la actuación de nativos y no-nativos en pruebas diseñadas para medir la competencia/actuación en la L2, es decir, pruebas diseñadas para HNN aplicadas a HN (Alderson, 1980). En estos trabajos de investigación se demuestra empíricamente que la actuación de los HN es muy heterogénea y que además la competencia/actuación de los propios HN, se encuentra lejos del nivel superior. Cuando menos, podríamos calificar estos resultados de paradójicos, ya que contrastan con los resultados que cabría esperar pues, como señala McNamara (1996), el grado de competencia-actuación está sistemáticamente relacionado con las destrezas cognitivas individuales de cada persona. Weir (1988) realizó una serie de pruebas lingüísticas, utilizando este tipo de test, y encontró claras diferencias entre la producción de HN y HNN. Los HN consiguieron resultados que oscilaban entre el 76% y el 88% en tareas de lectura y entre el 65% y el 88% en tareas de escritura. Estos resultados están lejos del nivel 100% y aún más, cuestionan la imparcialidad y la validez de las pruebas que emplean el criterio mencionado anteriormente. Los resultados de este estudio cuestionan igualmente la noción de que exista una interacción modélica entre los

propios HN, ni a un macro-nivel, como pueden ser variedades o registros lingüísticos, ni tampoco a un micro-nivel como pueden ser situaciones cotidianas.

Widdowson (1998) señala que los diálogos y textos incluidos en manuales para la enseñanza del inglés dentro del enfoque comunicativo, con frecuencia describen una situación alejada de la realidad que no se corresponde con lo que ocurriría en un contexto cotidiano; por ejemplo, el diálogo de una familia a la hora del desayuno. En una situación real los interlocutores intercalan en su discurso oral gestos, frases incompletas e incluso bostezos, con lo cual el contexto y la información paralingüística juegan un papel decisivo. Otro autor que comparte este punto de vista es Fulcher (1987), quien analiza una conversación entre cuatro personas con amplia formación académica, tres de las cuales son HN de la lengua inglesa mientras que la cuarta es un HNN del inglés. Fulcher (*ibid.*) graba y analiza la conversación mantenida por los cuatro interlocutores y demuestra la gran cantidad de repeticiones, dudas, tartamudeos, aparentes errores gramaticales,... que están presentes en su discurso oral. Este autor concluye que tanto las escalas de evaluación como los materiales didácticos elaborados bajo esta concepción, responden a lo que especialistas en evaluación y diseñadores de pruebas **piensan que ocurre** en situaciones comunicativas reales. Sin embargo, las pruebas reflejan aquello que sus autores creen que ocurre, o debería ocurrir, en situaciones cotidianas, y no lo que realmente acontece. Algunos autores critican la excesiva confianza puesta en las pruebas para medir la competencia discursiva oral, establecidas a nivel mundial y que por su temprana implantación han creado tradición en el campo de la evaluación lingüística. El problema radica en que se trata de instrumentos imprecisos e ineficaces para medir lo que pretenden medir y por tanto carecen de la validez necesaria para tal efecto. Según estos autores, el principal peligro es que una vez implantadas en el sistema educativo son difíciles de modificar o, llegado el caso, erradicar. Por ello se produce un fenómeno al que tanto Lantolf y Frawley (1992) como Traynor (1985) coinciden en llamar “the tail wagging the dog”, es decir, el rabo moviendo al perro.

Otra circunstancia que merece considerarse es la amplia difusión que

llegan a tener este tipo de pruebas en el ámbito mundial. De hecho, aunque inicialmente diseñadas para el contexto sociopolítico de los Estados Unidos, llegan a adoptarse en otros contextos sin ser sometidas a revisión o adaptación alguna.

5. El HNC como objetivo idealizado en la enseñanza de lenguas

Llegado este punto, sería importante destacar algunas matizaciones con respecto al uso generalizado de términos como HN y HNN. Normalmente, se les entiende como términos antagónicos y se presupone la homogeneidad de los mismos, aún más, se les considera calificativos fácilmente utilizables para clasificar individuos. Firth y Wagner (1997) exponen la necesidad de llevar a cabo profundas modificaciones en la conceptualización de las investigaciones recientes relativas a la adquisición de segundas lenguas. Señalan estos autores que las teorías y metodologías de investigación en la adquisición de la L2 reflejan un desnivel importante entre orientaciones cognitivas por un lado y orientaciones sociales y contextuales por otro. Esta situación, y el marcado énfasis e interés por factores cognitivos en oposición a factores contextuales, han hecho que se identifique al aprendiz de la L2 como **un comunicador deficiente deseoso de conseguir la competencia-actuación de un HN idealizado**. Estos autores postulan que conceptos fundamentales como **HNN, aprendiz de lenguas e interlengua** representan elementos monolíticos en el ámbito de la adquisición de lenguas segundas, pues son entendidos de forma simplificada presuponiendo la existencia de un HN idealizado.

Un supuesto bastante difundido entre los especialistas en enseñanza de lenguas extranjeras es que todo aprendiz tiene como objetivo parecerse tanto como sea posible al HN. Sin embargo, Shepherd (1990) sugiere que ante la pregunta de si el objetivo final de los aprendices de lenguas extranjeras es parecerse a un HN tanto como sea posible, la respuesta inmediata de la mayoría sería afirmativa. Al pensar detenidamente en el esfuerzo necesario para convertirse en un HN (o pseudo-nativo) de la lengua meta, así como las implicaciones culturales que ello conllevaría, un gran número de estos aprendices cambiaría de opinión. En tal caso, posiblemente

se debería hacer una distinción entre fines instrumentales y fines integradores en el aprendizaje de lenguas extranjeras. Se debería matizar que el HNC, en sentido genérico, probablemente no sea el único modelo disponible para los aprendices de lenguas, ni tampoco que todo aprendizaje tenga como objeto último al HNC.

6. El modelo del HNC y el bilingüismo

Un aspecto con el cual algunos especialistas en bilingüismo como Romaine (1989) se muestran en total desacuerdo, es el hecho de que la lingüística teórica actual toma como punto de partida un comportamiento monolingüe. Por ello, tanto el análisis como la organización del conocimiento bilingüe se suelen contrastar con un modelo monolingüe. Tras esta tendencia, se encuentra nuevamente la postura chomskiana que parte de un HN ideal perteneciente a una comunidad de hablantes completamente homogénea, y que “conoce su lengua a la perfección”.

Desde mi punto de vista, la tendencia que considera al HNC como objetivo primordial del aprendizaje no es sensible a la realidad lingüística de nuestro mundo, pues se da la circunstancia de que el bilingüismo se está convirtiendo, cada vez más, en la norma en vez de la excepción. El bilingüismo, o incluso el multilingüismo, representan la situación más cotidiana en diversas regiones de España, en el resto de Europa y del mundo. Por ello, las pruebas de evaluación que toman como referencia al HNC perteneciente a una comunidad monolingüe homogénea, muestran un alto grado de parcialidad. Son diversos los tipos de bilingüismo que se pueden encontrar, dependiendo del grado de uso que se tenga en cada lengua, de las circunstancias en el aprendizaje de las mismas, o del dominio alcanzado en cada lengua. Sin embargo, una de las situaciones más frecuentes es aquella en la que el manejo de una de las lenguas es superior a la otra, posiblemente porque los sujetos bilingües habitualmente usan cada una en contextos sociales diferentes y se produce un desequilibrio en cuanto a la cantidad de tiempo o al registro lingüístico. Teniendo en cuenta estos factores, resultaría problemática la asignación de la cualidad HN a un bilingüe en cada una de estas lenguas, si bien sería más conflictivo aún, privarle de la afiliación a una de ellas.

Para finalizar con la relación existente entre el HNC y el bilingüismo, se debería resaltar el paralelismo existente entre los términos “HN” y “bilingüe equilibrado”. Según Romaine (1989), el concepto “bilingüe equilibrado” es tan escurridizo como el del “HN”, por tratarse de un concepto perteneciente a la perspectiva teórica que toma al comportamiento monolingüe como punto de referencia.

7. Conclusiones

McNamara (1996) constata que no se ha investigado suficientemente la actuación del HN y por ello las escalas de evaluación que lo utilizan, implícita o explícitamente, carecen de suficiente validez. Por ello, propone que se efectúen mayores esfuerzos para proporcionar validez a las escalas de evaluación, de las que además se debe excluir al HN. En ningún momento debería confundirse la fiabilidad con la validez, pues al intentar proporcionar validez a la prueba por medio de la comparación con la competencia discursiva del HNC, se asume en este último una fiabilidad de la que carece, lo que a su vez resta validez al proceso evaluador.

Las investigaciones futuras destinadas a indagar en la valoración de la competencia discursiva de los aprendices deben abandonar esa perspectiva teórico-analítica para intentar averiguar cuáles son las aspiraciones del aprendiz. Para ello, se deben dejar a un lado modelos idealizados y tomar como punto de partida la competencia-actuación real de los aprendices. Es preciso entenderles como miembros activos que pertenecen a un entorno social en el que se usa la lengua de diferentes maneras y, además, tomar en consideración los factores contextuales y pragmáticos que forman parte de la comunicación. El aprendizaje, y la evaluación del mismo, deben ir encauzados hacia la comunicación e interacción social con todo tipo de interlocutores y no a su comparación con modelos idealizados. La competencia oral del HNC es indiscutible y necesariamente diferente a la del HN por la marcada diferencia en cuanto al tiempo dedicado al uso de la lengua en cuestión, y por ello carece de sentido la comparación *vis a vis* de ambas para fines evaluadores. Una comparación que muy frecuentemente responde a unos criterios de política lingüística que intenta favorecer o

preservar la hegemonía de un colectivo determinado, o de una variedad lingüística concreta.

Referencias bibliográficas

- Alderson, J.C. 1980. 'Native and nonnative speaker performance on cloze tests.' *Language Learning* 3:59-76.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L.F. & S.J. Savignon. 1986. 'The evaluation of communicative language proficiency: A critique of the ACTFL Oral Interview.' *The Modern Language Journal* 70/4:380-390.
- Barnwell, D. 1989. 'Proficiency and the native speaker.' *ADFL Bulletin* 20/2:42-46.
- Coulmas, F. 1981. *A Festschrift for Native Speaker*. The Hague: Mouton Pub.
- Firth, A. y J. Wagner. 1997. 'On discourse, communication, and (some) fundamental concepts in SLA research.' *The Modern Language Journal* 81/3:285-300.
- Fulcher, G. 1987. 'Tests of oral performance: the need for data-based criteria.' *ELT Journal* 41/4:287-291.
- Fulcher, G. 1996. 'Invalidating validity claims for the ACTFL oral rating scales.' *System* 24/2:163-172.
- Hamilton, J. (et al.). 1993. 'Rating scales and native speaker performance on a communicatively oriented EAP test.' *Language Testing* 10/3:337-353.
- Lantolf, J.P. y W. Frawley. 1988. 'Proficiency: Understanding the construct.' *Studies in Second Language Acquisition* 10/2:181-195.
- Lantolf, J.P. y W. Frawley. 1992. 'Rejecting the OPI - Again: A Response to Hagen.' *ADFL Bulletin* 23/2:34-37.
- Liskin-Gasparro, J. 1984. 'The ACTFL proficiency guidelines: a historical perspective' en Higgs, T.V. (ed.) 1984: *Teaching for Proficiency: The Organizing Principle*. Lincolnwood, IL: National Textbook Company.

- Liskin-Gasparro, J. 1996. 'Circumlocution, Communication Strategies, and the ACTFL proficiency guidelines: An analysis of student discourse.' *Foreign Language Annals* 29/3:317-330.
- McNamara, T.F. 1996. *Measuring Second Language Performance*. London: Longman.
- Richards, J.C., J. Platt y H. Platt. 1997. *Diccionario de Lingüística Aplicada y Enseñanza de Lenguas*. Barcelona: Editorial Ariel.
- Romaine, S. 1989. *Bilingualism*. Oxford: Blackwell.
- Shepherd, J. 1990. 'La fluidez oral del inglés (lengua materna): Implicaciones pedagógicas para la evaluación oral del estudiante de inglés (segunda lengua).' en F. Garrudo Carabias y J. Comesaña Rincón (eds.) 1990: *Actas del VII Congreso Nacional de Lingüística Aplicada*. Sevilla: AESLA.
- Traynor, R. 1985. 'The TOEFL: An Appraisal'. *ELT Journal* 39/1:43-47.
- Valdman, A. 1989. 'The problem of the target model in proficiency-oriented foreign language instruction.' *Applied Language Learning* 1/1:33-51.
- Weir, C.J. 1988. *Communicative language testing with special reference to English as a FL*. Exeter: University of Exeter.
- Widdowson, H. 1998. 'Real English: Whose reality?'. Sesión Plenaria en las *XIV Jornadas Pedagógicas para la Enseñanza del Inglés*. GRETA: Granada.