Procedimientos para asegurar la validez y la fiabilidad de la evaluación de la interacción oral para la certificación lingüística de ELE

Anna Doquin de Saint Preux
Susana Martín Leralta
Universidad Antonio de Nebrija

Resumen

La interacción oral requiere de una evaluación directa y subjetiva de la actuación del candidato. Para asegurar la validez y fiabilidad de la prueba, se necesita llevar a cabo una serie de procedimientos. Estas cuestiones se abordan en la presente comunicación a la luz de los resultados obtenidos tras el diseño e implantación del Diploma LETRA (*Lengua española para trabajadores inmigrantes*), proyecto de investigación de la Universidad Nebrija y la Comunidad de Madrid. A partir del análisis correlacional de las calificaciones obtenidas por los candidatos en cada una de las dos evaluaciones (holística y analítica), y de las encuestas realizadas a los evaluadores, se mostrará la incidencia de los procedimientos adoptados sobre los resultados de la prueba evaluadora.

1. Introducción

El presente estudio se enmarca dentro del proyecto de certificación lingüística de español inicial (A2-n) para trabajadores inmigrantes, realizado por la Universidad Nebrija y financiado por la Dirección General de Inmigración de la Comunidad de Madrid entre los años 2008 y 2013[1]. El proyecto se ha desarrollado en las siguientes fases: análisis de necesidades, diseño y pilotaje del examen de certificación Diploma LETRA (*Lengua española para trabajadores inmigrantes*), implantación del examen (convocatorias de noviembre de 2010, noviembre de 2011 y mayo de 2013) y transferencia de resultados de investigación[2].

De acuerdo con los expertos, para asegurar la calidad de un examen deben mantenerse los criterios de validez, fiabilidad y transparencia (Luoma, 2004). Con el propósito de lograrlos, a la hora de diseñar la prueba de evaluación de la interacción oral del examen LETRA se han llevado a cabo los siguientes procedimientos: una definición precisa del constructo, un sistema de doble evaluación del candidato (holística y analítica), el establecimiento de criterios de evaluación y descriptores de calificación, y una formación específica para los evaluadores, de cara a asegurar un comportamiento homogéneo en su actuación como interlocutores, así como una interpretación común de los criterios y descriptores. Con estos procedimientos se pretende minimizar la variación e inconsistencia propias de las pruebas de evaluación subjetiva.

1. Proyecto llevado a cabo por el Grupo de investigación LAELE (Lingüística aplicada a la enseñanza de lenguas extranjeras), dirigido por la Dra. Marta Baralo Ottonello.
2. Pueden consultarse los resultados parciales de las diferentes fases en Escudero y Guerra (2010), Martín (2011 y 2012) y Baralo y Estaire (2011).

En el trabajo que aquí se presenta nos proponemos los siguientes objetivos: analizar los procedimientos llevados a cabo para el diseño e implantación de la prueba de evaluación de la interacción oral; someter a un análisis correlacional las calificaciones obtenidas por los candidatos en cada una de las dos evaluaciones (holística y analítica) en las dos primeras convocatorias del examen (total de 600 candidatos), y analizar los resultados de las encuestas de satisfacción realizadas a los evaluadores al final de cada convocatoria del examen. La finalidad estriba en conocer el grado de validez y fiabilidad que aportan a la prueba evaluadora los procedimientos llevados a cabo durante el diseño e implantación de la misma.

2. Fundamentos teóricos

Tal como indica Bachman (1990), la confección de exámenes de idiomas implica una concepción del uso comunicativo de la lengua y de la habilidad comunicativa, así como de una teoría de la medición. De la concreción de estos dos aspectos, como veremos más adelante, partiremos para la sistematización de los procedimientos que se van a llevar a cabo en el desarrollo e implantación del examen.

La evaluación de la destreza oral es un proceso que consta de diversos estadios, en cada uno de los cuales se actúa e interactúa para tomar una serie de decisiones que afectarán a los estadios siguientes (Luoma, 2004). Los procedimientos que se adoptan en cada uno de ellos se retroalimentan y deben revisarse al finalizar cada convocatoria del examen, de cara a adoptar los ajustes oportunos a la luz de los resultados de la implantación de las pruebas. Esta revisión debe hacerse, además, teniendo en cuenta las necesidades y dificultades percibidas por los diferentes agentes que participan (diseñadores del examen, evaluadores, examinandos, personal de administración de la prueba).

Estamos de acuerdo con Martínez Baztán (2011) en que para diseñar exámenes eficaces tenemos que adoptar un modelo de evaluación garantista, entendiendo por ello una serie de pasos: probar el diseño del examen en su proceso, conocer las posibilidades y limitaciones de las tareas, asegurar la transparencia hacia el candidato y formar adecuadamente a los evaluadores.

Si bien es conveniente que todos los exámenes gocen de estas garantías, en el caso de los exámenes de alto impacto resultan imprescindibles. Así sucede con los exámenes encaminados a obtener un diploma, ya que, como recuerda Bordón (2006), repercuten tanto en los beneficios para los candidatos como en el desarrollo de currículos y materiales de enseñanza.

A la hora de diseñar e implementar el Diploma LETRA, se han tenido en cuenta las directrices del *Marco Común Europeo de Referencia para las Lenguas* (2002), las recomendaciones de los expertos (Bachman, 1990; Alderson, Clapman y Wall, 1998; Luoma, 2004; Bordón, 2006; Martínez Baztán, 2007, entre otros) y los estándares de las agencias ALTE (*Association of Language Testers in Europe*) y EALTA (*European Association for Language Testing and Assessment*).

2.1. Definición del constructo

La definición del constructo parte de la consideración de Bachman (1990) anteriormente citada, de un concepto de habilidad comunicativa y uso de la lengua. Para llegar a una definición precisa, se siguen las tres líneas guía de Luoma (2004): orientación a la lengua, orientación a la comunicación y basado en la situación.

Por otra parte, en lo que respecta a la teoría de medición adoptada, nos encontramos con una prueba de evaluación del dominio y de la actuación, directa, subjetiva y con referencia a un criterio.

El constructo de la prueba de interacción oral del Diploma LETRA queda definido de la siguiente forma: participar en una entrevista sobre información personal e intereses laborales; mantener un monólogo sostenido acerca del ámbito profesional de interés y responder a cuestiones sobre sus acciones, herramientas, riesgos y buenas prácticas; mantener una conversación acerca de las instituciones de apoyo al trabajador inmigrante y la documentación relativa al ámbito administrativo laboral.

La definición del constructo, al igual que los procedimientos de evaluación y calificación, se recoge en las especificaciones de la prueba, documento de uso interno que facilita "información oficial sobre lo que un examen evalúa y cómo lo evalúa" (Alderson, Clapham y Wall, 1998: 14).

2.2. Sistema de doble evaluación del candidato

Para asegurar la fiabilidad de la evaluación, se lleva a cabo una doble valoración de la actuación del candidato. Por una parte, el evaluador que interactúa con él durante la entrevista, denominado "examinador", realiza una evaluación holística y otorga una única calificación relativa al conjunto de la prueba completa. Por otra parte, un segundo evaluador, observador no participante en la interacción, lleva a cabo una evaluación analítica por bandas en función de diez criterios de evaluación, a cada uno de los cuales otorga una calificación con semejante ponderación. La nota final resultante es la media aritmética de la calificación de ambos evaluadores.

Los criterios que aplica cada uno de ellos se detallan a continuación.

2.3. Establecimiento de criterios de evaluación y descriptores de calificación

Dado que las pruebas de interacción oral son de carácter subjetivo, para asegurar su fiabilidad se requiere la elaboración de criterios de evaluación "sólidos y bien definidos" (Bordón, 2006: 243).

A partir de la concepción comunicativa a la que nos hemos referido al hablar de la definición del constructo, para la evaluación del Diploma LETRA hemos definido tres criterios de evaluación para la evaluación holística, y diez criterios para la evaluación

analítica. Nos hemos basado para ello en el MCERL, así como en la bibliografía de referencia sobre evaluación de la interacción oral.

Los tres criterios de evaluación holística son la comunicación (entendiendo por ello y el cumplimiento de la intención comunicativa), la relación con el interlocutor (la capacidad para mantener la conversación) y el control lingüístico (utilización de recursos lingüísticos).

Los diez criterios de evaluación analítica se definen de la siguiente forma:

- Mantenimiento de la conversación: Participa en una conversación de forma sencilla para describir acciones y elementos de los ámbitos personal, laboral y administrativo. Se ayuda de estrategias compensatorias para la comprensión y la producción.

- Cumplimiento de la intención comunicativa: Se hace entender y cumple adecuadamente los fines funcionales (la intención comunicativa) con un cierto grado de independencia del interlocutor.

- Fluidez: Se desenvuelve con enunciados breves, con pausas para buscar expresiones, articular palabras menos habituales, reformular y salvar la comunicación.

- Organización y cohesión/coherencia discursiva: Organiza su discurso de manera adecuada y lógica para describir acciones y elementos de los ámbitos personal, laboral y administrativo.

- Competencia sociopragmática: Establece contactos sociales básicos en un registro formal manifestando cortesía (empleo de expresiones de disculpa, agradecimiento, petición) y utilizando formas de tratamiento y saludos apropiados.

- Alcance y control de vocabulario: Domina un repertorio básico de palabras y frases sencillas relativas a sus datos personales y a situaciones concretas de los ámbitos administrativo y laboral (horarios y herramientas de una profesión, riesgos laborales que entraña y buenas prácticas en el trabajo) que le permite aportar la información requerida.

- Alcance y control gramatical: Muestra un control limitado de unas pocas estructuras gramaticales sencillas y de modelos de oraciones dentro de un repertorio básico o aprendido, que le permiten expresar su intención comunicativa (afirmativa, negativa, interrogativa, exhortativa), aunque con errores sistemáticos propios del nivel.

- Control fonológico/pronunciación: Su pronunciación está marcada por el acento extranjero, pero es comprensible. Deletrea de forma suficientemente clara para ser comprendido.

- Control fonológico/entonación: La entonación es adecuada a su intención comunicativa y discrimina la intención comunicativa del interlocutor por la entonación que emplea.

- Referentes socioculturales: Es capaz de identificar instituciones/oficinas del ámbito administrativo-laboral y centros de apoyo a las actividades del trabajador inmigrante como agente social. Es capaz de explicar su función así como algunas gestiones que se realizan en los mismos. Es capaz de identificar documentos pertinentes para algunas de esas u otras gestiones, explicar su función y los procedimientos para su obtención.

De acuerdo con Eres y Rádis (2010: 32), los criterios claros, válidos y fiables son una ventaja de los exámenes de certificación. Sin embargo, dicen las autoras, precisamente por tratarse de patrones y categorías de medición fijos, no contemplan la idiosincrasia de los candidatos.

A la hora establecer y aplicar los criterios de evaluación para la prueba de interacción oral del Diploma de Lengua española para trabajadores inmigrantes, se han perseguido la garantía de la validez y la fiabilidad, pero se han considerado los factores que, de acuerdo con Raga (2003), intervienen en la interacción comunicativa: formales (usos verbales, cortesía verbal, aspectos paralingüísticos, turnos de palabra, lenguaje no verbal...), y significativos (valores psicológicos, emotivos, individuales y sociales), teniendo en cuenta, según indica el autor, las posibles divergencias derivadas del hecho de que se trate de interacciones interculturales.

Para garantizar la fiabilidad de la evaluación, es imprescindible describir la actuación lingüística correspondiente a cada banda de la escala de calificación aplicada. En el caso del Diploma LETRA, la escala va de 1 a 4, considerando que 1 es la peor calificación.

Tal como se recoge en Martín (2012), a pesar de que no es habitual emplear descriptores para la escala de calificación de la evaluación holística, durante la fase de pilotaje del examen LETRA llevada a cabo en nuestro proyecto de investigación, se observó la necesidad de contar con ellos, de cara a asegurar que todos los evaluadores tuvieran la misma imagen de la actuación lingüística correspondiente a cada banda de la escala de calificación, percepción difícilmente alcanzable sin unos descriptores. Con la fase de validación del examen LETRA se demostró el acierto en la decisión, puesto que se obtuvo una correlación alta y significativa entre los evaluadores, e igual volvió a suceder en la primera convocatoria del examen.

Los descriptores de la escala de calificación de la evaluación holística corresponden a cuatro bandas, pero solo mostramos dos como ejemplo, para no exceder los límites de esta comunicación:

- Respuesta satisfactoria (calificación: 3). Comunicación: El candidato cumple adecuadamente los objetivos comunicativos de la prueba aunque evidenciando ocasionalmente un cierto grado de dificultad. A pesar de esta dificultad, su men-

saje es suficiente para satisfacer la necesidad comunicativa planteada. Relación con el interlocutor: Para su actuación, el candidato depende ocasionalmente de la cooperación del interlocutor. Un interlocutor acostumbrado a tratar con hablantes de otras lenguas puede comprenderle con cierto esfuerzo y concentración. Control lingüístico: Manifiesta cierta dificultad en la utilización de los recursos lingüísticos básicos de un nivel A2-1 y correspondientes a un repertorio reducido: necesita algunas pausas y reformulaciones, a menudo demuestra dudas y comete numerosos errores, la mayoría propios de su nivel.

- Respuesta insatisfactoria (calificación: 2). Comunicación: El candidato tiene serias dificultades para reaccionar comunicativamente a la situación planteada. En algunas ocasiones su mensaje se relaciona sólo vagamente con dicha situación, no satisfaciendo por tanto la necesidad comunicativa planteada. Relación con el interlocutor: Para su actuación, el candidato depende totalmente de la cooperación del interlocutor. Un interlocutor acostumbrado a tratar con hablantes de otras lenguas necesita esforzarse al máximo para comprenderle. Control lingüístico: Utiliza de forma muy limitada y poco controlada los recursos lingüísticos básicos de un nivel A2-1 y correspondientes a un repertorio reducido: de forma continua comete errores, necesita pausas y reformulaciones y demuestra dudas.

Respecto a la evaluación analítica, se describen cuatro bandas correspondientes a cada uno de los diez criterios de evaluación empleados. Dado que, por razones de espacio, no podemos recoger aquí los descriptores de escala completa[3] ya que es demasiado extensa, incluimos los descriptores correspondientes al criterio de mantenimiento de la conversación, a modo de ejemplo:

- Mantenimiento de la conversación: (1) No es capaz de participar en la conversación, se limita a la emisión de palabras sueltas, sin activar estrategias compensatorias que le ayuden a transmitir el mensaje. Carece de recursos para comprender al interlocutor. (2) Participa en la conversación con gran dificultad, no logra ofrecer satisfactoriamente la información requerida sobre los ámbitos personal, laboral o administrativo. Las estrategias que emplea no son suficientes para mantener la conversación. Muestra continuamente dificultad para comprender al interlocutor. (3) Participa adecuadamente en la conversación, aunque de forma sencilla, ofreciendo suficiente información sobre los ámbitos personal, laboral y administrativo. Hace uso eficaz de algunas estrategias compensatorias para la comprensión y la producción. (4) Participa en la conversación de manera sencilla ofreciendo amplia información sobre los ámbitos personal, laboral y administrativo. Activa un repertorio amplio de estrategias para comprender y otras que le ayudan a transmitir el mensaje de manera completa.

Los criterios de evaluación y los descriptores de calificación han sido sometidos a un proceso de pilotaje y validación, que ha servido para reelaborarlos hasta su versión definitiva. La validación se realizó con una muestra de 184 informantes entre diciembre

3. Pueden consultarse los descriptores completos en el trabajo de Baralo y Estaire (2011).

de 2009 y febrero de 2010 en 10 centros (ONG, CEPI y asociaciones) de la Comunidad de Madrid, León y Santander.

Los criterios de evaluación y los descriptores de calificación, en su versión definitiva, se revisan antes y después de cada convocatoria del Diploma LETRA, en el marco del curso de formación de evaluadores.

2.4. Formación de los evaluadores

Como se indica en nuestro estudio anterior (Martín, 2012), el principal requisito para la implementación de una prueba de interacción oral es la formación de los evaluadores que van a participar en el proceso, puesto que el papel y el comportamiento del interlocutor durante la realización de la prueba es una variable que puede afectar a la realización lingüística del candidato y a la puntuación del examinador.

De acuerdo con Karavas y Delieza (2009), la inconsistencia y la variación son los principales factores que afectan a la validez y a la fiabilidad de la prueba. Los aspectos en los que se producen, según los autores, son la laxitud o dureza en la aplicación de la escala de calificación, la interpretación de las escalas de calificación y los criterios de evaluación, la manera de adaptarse al nivel lingüístico de los candidatos, el apoyo que les ofrecen y la elicitación de la realización comunicativa de los candidatos. Para evitarlo, se recomienda que se lleve a cabo una formación específica de los evaluadores que facilite oportunidades sistemáticas de entrenamiento inicial y de actualización.

Con el propósito de controlar estos aspectos, en el marco del proyecto de investigación del Diploma LETRA, aparte de tener en cuenta las recomendaciones de las investigaciones precedentes, se ha llevado a cabo un estudio exploratorio encaminado a analizar las diferencias de los evaluadores y la incidencia de su comportamiento en realización lingüística de los candidatos y la calificación obtenida (Amador, 2011). Los resultados obtenidos se han considerado a la hora de diseñar las ediciones posteriores del curso de formación de evaluadores, concretamente, a la hora de ofrecer guiones y marcos de actuación a los examinadores.

La formación de los evaluadores del Diploma LETRA consiste en un curso inicial de 4 créditos ECTS (100 horas de duración) que se imparte de manera semipresencial, combinando la enseñanza online con sesiones presenciales, más un curso de actualización antes de cada convocatoria del examen. Los objetivos del curso son conocer las especificaciones de la prueba, familiarizarse con la estructura y la tipología de tareas, comprender las capacidades evaluadas, conocer el material necesario para realizar el examen y aprender a manejarlo (material gráfico, cuadernillo del examinador, hojas de evaluación), conocer el protocolo de administración del examen y los papeles de los diferentes participantes (administrador, preparador, examinador/entrevistador y evaluador), conocer y manejar los criterios de evaluación holística, conocer y manejar los criterios de evaluación analítica, aplicar la escala de calificación de la evaluación holística a exámenes reales grabados, aplicar la escala de calificación de la evaluación analítica a exámenes reales grabados, someter a acuerdo entre jueces las evaluaciones realizadas

para llegar a un entendimiento común de los criterios de evaluación y las bandas de calificación, analizar la actuación de diferentes examinadores/entrevistadores en pruebas orales reales grabadas, y realizar un simulacro de prueba de interacción oral con un informante/candidato potencial del examen de certificación.

3. Estudio empírico: análisis correlacional de las calificaciones de la doble evaluación

Para comprobar la validez y la fiabilidad de las pruebas orales, hemos sometido las calificaciones de las dos evaluaciones (analítica y holística) a un análisis estadístico correlacional. En cada una de las dos primeras convocatorias del Diploma LETRA (noviembre de 2010 y noviembre de 2011) hubo un total de 300 candidatos (n=300). Para cada una de las pruebas hemos aplicado una correlación de Pearson entre las 300 puntuaciones de la evaluación analítica con las 300 puntuaciones de la evaluación holística.

Como se puede ver en las tablas 1 y 2, en ambos casos hemos obtenido un índice de correlación muy alto y muy significativo: 0,947 en la primera convocatoria y 0.891 en la segunda. Lo que nos permite afirmar con mucha seguridad que los candidatos tienen puntuaciones parecidas en las dos evaluaciones.

Convocatoria 2010		Interacción oral Evaluación holística	Interacción oral Evaluación analítica GENERAL
Interacción oral Evaluación holística	Correlación de Pearson	1	,947**
	Sig. (bilateral)		,000
	N	300	300
Interacción oral Evaluación analítica GENERAL	Correlación de Pearson	,947**	1
	Sig. (bilateral)	,000	
	N	300	300

1. Resultados del análisis de correlación entre las evaluaciones holística y analítica de la prueba de interacción oral. Primera convocatoria del examen LETRA (2010)

Convocatoria 2011		Interacción oral Evaluación holística	Interacción oral Evaluación analítica GENERAL
Interacción oral Evaluación holística	Correlación de Pearson	1	,891**
	Sig. (bilateral)		,000
	N	300	300
Interacción oral Evaluación analítica GENERAL	Correlación de Pearson	,891**	1
	Sig. (bilateral)	,000	
	N	300	300

2. Resultados del análisis de correlación entre las evaluaciones holística y analítica de la prueba de interacción oral. Segunda convocatoria del examen LETRA (2011)

Para reforzar la fiabilidad de la prueba, también hemos analizado los resultados de las encuestas de satisfacción realizadas a los evaluadores al final de cada convocatoria del examen. En la convocatoria de noviembre de 2010, 13 de 35 evaluadores contestaron a la encuesta (N= 13) y el la de noviembre de 2011, 20 de 35 (N= 20). Presentamos aquí las respuestas a la pregunta siguiente: "*Con la información recibida fue relativamente sencillo puntuar las pruebas orales a pesar de su naturaleza compleja*" en la que tenían que contestar con una puntuación del 1 al 5 en una escala de Likert (1: Completamente en desacuerdo – 5: Completamente de acuerdo).

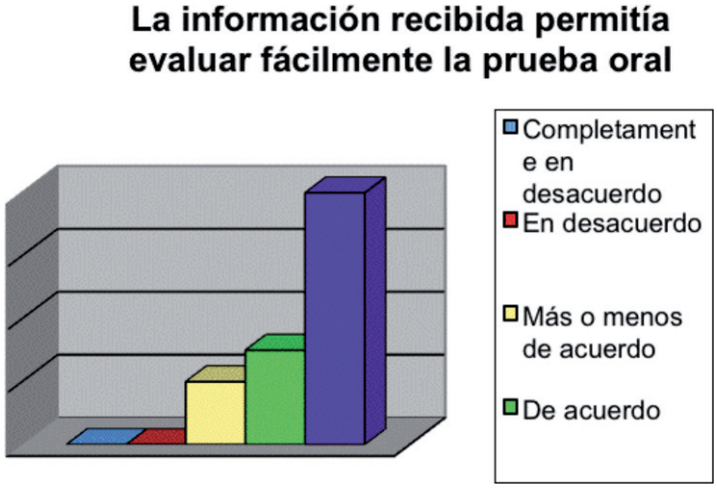

Gráfico 1: Respuestas de los evaluadores (n=13)- Primera convocatoria del examen LETRA (2010)

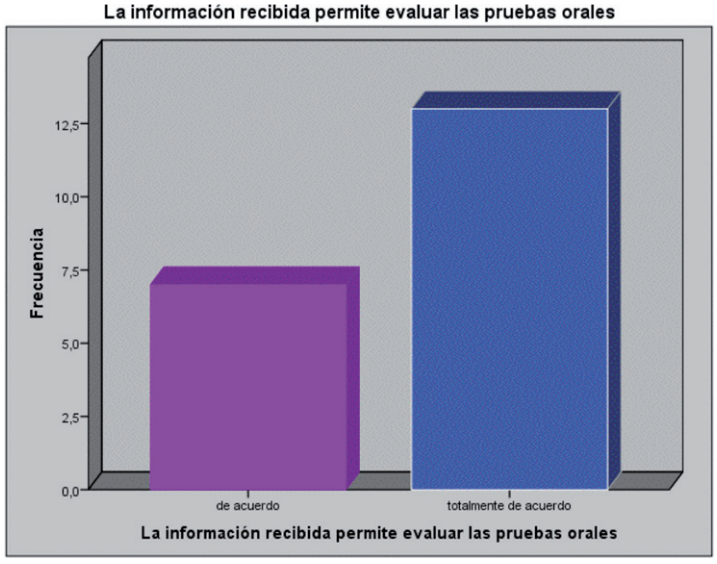

Gráfico 2: Respuestas de los evaluadores (n=20)- Segunda convocatoria del examen LETRA (2011)

4. Conclusiones

El análisis de estos resultados nos permite concluir que las pruebas de evaluación oral son fiables y que nuestra investigación tiene validez interna. Podemos comprobar, mediante este tipo de análisis, que los fundamentos teóricos y los procedimientos seguidos tanto para el diseño de la prueba oral del Diploma LETRA, como de sus materiales de evaluación, se revelan adecuados para evitar la inconsistencia y variación ente los evaluadores.

Estamos de acuerdo con Martínez Baztán (2011) en que evaluar es una necesidad social de gran importancia, que debe estar sometida a un escrutinio y control académico, sociolaboral y económico que le dé credibilidad y certifique su competitividad y excelencia. Asimismo, consideramos que compartir las prácticas evaluativos y sus resultados contribuirá a avanzar en esta área de estudio.

De cara a mejorar las herramientas de evaluación y la formación de evaluadores, las futuras líneas de investigación que nos planteamos pasan por estudiar la correlación por criterios de evaluación analítica (recordemos que en el Diploma LETRA se manejan diez criterios de evaluación analítica), para determinar si en alguno de ellos la correlación es baja y, por lo tanto, hubiera que revisar su formulación o abordar su interpretación de manera específica en la formación de evaluadores.

De cara a mejorar el diseño de la enseñanza, nos proponemos profundizar en el estudio de algunos parámetros evaluados en la evaluación analítica (mediante la observación de las grabaciones del examen) según ciertas variables individuales de los informantes, de cara a obtener conclusiones útiles para la enseñanza específica para esos aprendientes. Estudiaremos, por ejemplo, las estrategias de interacción oral utilizadas por los candidatos según su lengua materna, grado de alfabetización y nivel académico, para contrastar estos resultados con la calificación obtenida en la evaluación analítica. Asimismo, queremos analizar también la adecuación sociopragmática en la interacción oral por los inmigrantes según su lugar de procedencia, y contrastar los resultados con la calificación obtenida en la evaluación analítica.

Bibliografía

ALDERSON, J. Charles, Caroline CLAPHAM y Diane WALL (1998): *Exámenes de idiomas. Elaboración y evaluación*, Madrid: Cambridge University Press.

AMADOR LABELLA, Isabel (2011): *La certificación lingüística para trabajadores inmigrantes: análisis de la interacción oral y pautas didácticas*, Trabajo de Fin de Máster inédito, Madrid: Universidad Nebrija.

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

BARALO OTTONELLO, Marta y Sheila ESTAIRE (2011): «Variables socioculturales y comunicativas en el diseño curricular de una certificación de español para trabajadores inmigrantes», *Lengua y Migración*, 3/2, 5-42.

BORDÓN, Teresa (2006): *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*, Madrid: Arcolibros.

CONSEJO DE EUROPA (2002): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Ministerio de Educación, Cultura y Deporte, y Anaya.

ESCUDERO DOMÍNGUEZ, Inmaculada y Rosario GUERRA INFANTE (2010): «La observación etnográfica como método para el análisis de necesidades comunicativas de la población inmigrante», en M.ª Rosario Caballero y M.ª Jesús Pinar (coords.), *Modos y formas de la comunicación humana,* vol. 2, 263-268.

ERES FERNÁNDEZ, I. Gretel y Livia M. RADIS BAPTISTA (2010): *La enseñanza de lenguas extranjeras y la evaluación*, Madrid: Arcolibros.

KARAVAS, Evdokia y Xenia DELIEZA (2009): «On site observation of KPG oral examiners: Implications for oral examiner training and evaluation», *Apples-Journal of Applied Language Studies*, 3/1, 51-77.

LUOMA, Sari (2004): *Assessing Speaking*, Cambridge: Cambridge University Press.

MARTÍN LERALTA, Susana (2011): «Certificación lingüística de nivel inicial para inmigrantes en contexto laboral: ejemplo de una prueba del examen DILE», *Lengua y Migración*, 3/1, 89-104.

MARTÍN LERALTA, Susana (2012): «La evaluación de la interacción oral para la certificación lingüística en L2», en *Actas del XVI Congreso Internacional de la ALFAL*, Alcalá de Henares: Universidad de Alcalá.

MARTÍNEZ BAZTÁN, Alfonso (2007): «Consideraciones sobre la evaluación de la lengua de personas inmigradas adultas y adolescentes», *Linred. Lingüística en la Red*, [en línea], http://www.linred.es/monograficos_pdf/LR_monografico1_articulo1.pdf

MARTÍNEZ BAZTÁN, Alfonso (2011): *La evaluación de las lenguas: garantías y limitaciones*, Granada: Octaedro Andalucía.

RAGA GIMENO, Francisco (2003): «Para un análisis empírico de las interacciones comunicativas interculturales», en Grupo CRIT, *Claves para la comunicación intercultural. Análisis de interacciones comunicativas con inmigrantes*, Castellón de la Plana: U. Jaume I, 17-35.

Fuentes electrónicas:

ALTE: *Código de práctica de ALTE*: http://www.alte.org/attachments/files/code_practice_es.pdf

EALTA: *Directrices para una buena práctica en el desarrollo de pruebas y la evaluación de lenguas:* http://www.ealta.eu.org/documents/archive/guidelines/Spanish.pdf