# LA EVALUACIÓN ORAL: INTERACCIONISMO Y CONDUCTISMO

Alfonso Martínez Baztán
*Centro de Lenguas Modernas de la Universidad de Granada*

RESUMEN

*Esta comunicación va a dar cuenta de varios modelos de evaluación de la lengua oral centrándose en los conductistas y los interaccionistas mostrando los diferentes discursos que generan. Algo necesario porque la diversificación de tests y motivos por los que un candidato acepta someterse a una prueba de nivel lingüístico hacen que quienes las realizan y utilizan deban conocer claramente las dificultades y limitaciones de cada modelo para saber elegir el más conveniente y usarlo apropiadamente.*

Palabras clave: evaluación oral, interaccionismo, conductismo, estrés comunicativo, elicitación, análisis del discurso.

ABSTRACT

*This communication will account for several models of oral language testing focusing on behaviorists and interactionists and show the different discourses they generate. Something necessary because the diversification of tests and reasons for which a candidate accepts to undergo a test of linguistic level make that those who make and use them must clearly know the difficulties and limitations of each model to know how to choose the most convenient and use it appropriately .*

Keywords: oral evaluation, interactionism, behaviorism, communicative stress, elicitation, discourse analysis.

## 1. DIFERENTES MODOS DE EVALUAR LA EXPRESIÓN ORAL

Ante la creciente oferta de tests para evaluar las lenguas y para evaluar la lengua oral (no confundir con evaluar la lengua oralmente -Martínez, 2011), ante los numerosos modelos diseñados para diferentes objetivos, aunque no necesariamente opuestos ni excluyentes (constructivismo, interaccionismo, fines específicos, para medir la competencia comunicativa o solo aspectos de la competencia lingüística -Widowson, 2003; Fulcher, 2011) y ante el poder creciente de estos en nuestra sociedad (Sohamy, 2001), consideramos necesario clarificar el panorama intentando conocer los efectos que producen para elegir adecuadamente el que se necesite y saber a qué tipo de evaluación nos enfrentamos.

Como ya apuntamos (Martínez, 2008) los tests "no son coches que se cogen indistintamente para, con diferente comodidad, llegar al mismo lugar con minutos de diferencia: son medios que conducen a lugares diferentes, aunque puedan entrecruzarse. Y es importante saber dónde están en cada momento y hacia dónde van, para poder usarlos según nuestras necesidades."

Hasta ahora los principales modos de evaluar la lengua hablada han sido conductistas basados en la premisa de autenticidad de que el resultado obtenido se asemejará a la lengua usada en la vida real. Así son muchos de los utilizados desde los años 50, junto a los constructos lingüísticos derivados del modelo de lengua de Bachman (1990) y Bachman y Palmer (1996). Estos se caracterizan por describir unas tareas o ítems que provoquen una reacción lingüística o comportamiento observable del que se deduce la existencia de una habilidad (Alderson y Banerjee, 2002).

Pero esta no es la única perspectiva. Fulcher (2003) apunta que el conductismo cree que el constructo lingüístico viene definido por tareas puesto que las tareas y sus contextos están indisolublemente ligadas a la habilidad que provocan. Pero que hay otras teorías discrepantes. La teoría de los rasgos lingüísticos estima que las tareas son irrelevantes ya que los rasgos lingüísticos son independientes de los contextos en que se generan. Así, las notas de un test se atribuirían fácilmente a las características del candidato, a su conocimiento y a procesos internos antes que al contexto. Alderson y Banerjee (2002) explican que esta perspectiva es tradicional en la evaluación educativa, aunque la consideran simplista pues el contexto es un factor importante en la actuación y en la valoración de la lengua.

Para el interaccionismo, finalmente, ambas perspectivas son erróneas: hay que tomar elementos de todas partes. Sus tests son los más complicados de di-

señar y de utilizar, pues encuentran la dificultad de captar y medir las características de la habilidad lingüística que proceden de los rasgos, del contexto y de la interacción (Alderson y Banerjee, 2002), aunque bien planeados y dirigidos pueden ser muy eficientes.

Cada modelo obtiene discursos distintos.

## 2. EFECTO DEL MÉTODO: INFLUENCIA EN EL DISCURSO GENERADO

Muchos trabajos han tratado del carácter social de la construcción del discurso (Swain, 2001; Fulcher, 2003b; Deville, 2006), de la variabilidad debida al contexto (Wertsch, 1991), de los problemas de estructuración de las entrevistas y tipo de actuación que posibilitan o de la influencia del entrevistador (Lazaraton 2002; Ross, 1998; Brown, 2003). Todos ellos apuntaban a que los candidatos actúan de forma distinta. Recientes análisis profundizan en ello (Canagarajah, 2006; Deville, 2006; Martínez, 2008; Warren, 2008; Brooks, 2009; García Laborda, 2015).

### *2.1 Pros y contras de los diferentes tipos de tareas*

Abordamos tres tipos de tareas de test: monólogo, interacción en parejas y entrevista, centrándonos en esta última.

Cabe afirmar que los monólogos no generan un lenguaje realista y representativo del lenguaje del candidato, incluso en contextos académicos. Por experiencia como formador de evaluadores, en sesiones con evaluadores de niveles C1 y C2 de exámenes oficiales, comprobamos habitualmente que tras pedirles que se evalúen realizando las mismas tareas que plantean a los candidatos, muchos manifiestan no sentirse representados por la lengua que han usado en su ”examen”. A estas faltas de validez de respuesta y representatividad cabe añadir la falta de motivación y de validez de apariencia, o sea, la insuficiencia de las tareas para generar un contexto adecuado para el monólogo, a lo que se une un lastre en la fiabilidad por la variabilidad que ocasiona la necesidad de plantear opciones a elegir.

Respecto a las interacciones de candidatos por parejas, hay datos interesantes. Aunque ganan en validez de apariencia y realismo porque se asemejan a una conversación natural y a veces aumentan la empatía y la negociación entre

los participantes (He y Dai, 2006) incluso en ámbitos académicos (Taylor, 2000) y aunque produce frases más largas, más turnos y más comunicativos (Garacía Laborda, 2015), la variabilidad, la complejidad de la interacción, la imprevisibilidad y la falta de control son serias amenazas al constructo (Swain, 2001; May, 2006; Canagarajah, 2006). Surgen más elementos pragmáticos que en otros modelos (hasta 17 frente a 10 en muchos casos -Brooks, 2009-: peticiones de confirmación, ayuda y clarificación, referencias a ideas del compañero, etc.) pero son una desventaja pues las escalas son incapaces de medir la riqueza de lo que se produce (May, 2006). Por último, no todos los participantes saben manejar adecuadamente la interacción (Luoma, 2004) y los contextos que surgen pueden desorientarles (Fulcher, 2003). Demasiados elementos como para asignar un nivel de habilidad lingüística estandarizado.

El tercer modelo, la entrevista con un evaluador, plantea problemas como que la construcción del discurso está influída por la actitud y preguntas del entrevistador, quien modifica su discurso de maneras no estandarizadas, poco claras o inconsistentes (Lazaraton, 2002) que provocan efectos inesperados en los candidatos, quienes simplifican su discurso, esperan pasivamente a las preguntas o producen menos elementos discursivos (O'Sullivan, 2001; Brown, 2003), meramente reactivos, con frases más breves, mayor número de autocorrecciones y menor negociación (Brooks, 2009; García Laborda, 2015). Aunque estos déficits puedan evitarse mediante formación y entrenamiento para guiar la interacción y con esquemas de elicitación adecuados (Byrnes, 1989; O'Loughlin, 2001; Lazaraton, 2002; Martínez, 2011).

### *2.2 Interaccionismo y conductismo: génesis de diferentes discursos*

Estudios de entrevistas con métodos diferentes basadas en las mismas tareas o actividades comparables y a los mismos candidatos (Martínez, 2008; Warren, 2008) muestran que ante una entrevista realizada de forma natural y discursiva (interacionista), se producen la mitad de pausas y de menos duración que ante las mismas tareas/actividades en otra entrevista en la que se solicita al candidato simplemente hablar sobre ello (conductista). Cuatro veces más de pausas grandes (además de las normales de buscar ideas o palabras) pasando de pausas de 0.4 o 0.8 segundos a pausas de más de 2 segundos y en 7 ocasiones, frente a solamente 2 con el modelo de entrevista natural-discursiva.

Además, cuando pides a un candidato que hable de su vida o comente una frase interesante de un periódico y escuchas su respuesta aplicando las escalas

de medición, sin intervenir en su discurso (conductismo), encuentras respuestas vagas, pasivas o meramente reactivas -por falta de intención o contexto- como estas que no generan suficiente discurso ni apropiado:

1.- ¿Puedes hablarme de tu ciudad?

*- Mi ciudad es pequeña, no es tan interesante como X (nombre de ciudad) porque en X hay más cosas y en mi ciudad es muy aburrida.*

-Vale, tu ciudad no es como X. pero… ¿puedes contarme algo más?

2.- ¿Puedes hablarme de tus planes?

*- Eh, bueno, nunca he pensado en …, mi vida… es … es muy... Yo soy estudiante...*

- ¿De verdad no tienes planes? ¿No puedes contarme un poco más?

3. - Y la gente ¿cómo es diferente en X (nombre de estado) de otros estados?

*- En el norte, la gente, lo mismo. Pero en el sur hay, no lo mismo. Son vida más menos comparativa.*

- ¿Puedes explicarme por qué?

*- En el norte hay mucho comercio y industria. En el sur, no. Hay playas y campo.*

Como vemos en el caso tres, la repregunta no logra su objetivo ni mejora la primera respuesta. Se necesita otro modelo de intervención.

*2.3 Análisis del discurso: causas y soluciones*

Los tests interaccionistas no se conforman con ofrecer tareas diseñadas para un fin y observar el discurso del candidato. Intervienen en su génesis mediante un "saber hacer" igual que un periodista, abogado o sicólogo bucean en situaciones cambiantes o las generan para averiguar la existencia de lo que desean obtener. Algo nada sencillo pero tampoco imposible (y obviamente innecesario ante candidatos acostumbrados a ámbitos académicos que saben amoldarse a cada tarea pues conocen si deben esforzarse por la gramática, vocabulario o determinada organización textual).

Evaluar abiertamente la lengua de candidatos de procedencias dispares y formaciones diversas y elicitar suficiente cantidad de datos para realizar una medición de sus habilidades con pruebas y no indicios requiere más herramientas y el conomiento del objetivo de cada tarea y de los modos de preguntar y repreguntar (para construir y modular el contexto que permita que se produzca un discurso real y no un discurso artificial "para la evaluación").

En los tres casos señalados, las repreguntas del entrevistador no pueden hacerse por curiosidad sino que deben encaminarse a devolver a la tarea o pregunta su sentido y valor. La curiosidad no ayudará en ningún caso. El sistema de elicitación interactivo debe ligarse al objetivo de las preguntas y a la estructura de la entrevista, para lograr devolver sentido a la tarea inicial cuando esta falla. Y usarlo con profesionalidad de buen entrevistador que se da cuenta de qué está pretendiendo hacer el candidato y si eso era lo que se pretendía con la tarea/pregunta.

Veamos nuestros ejemplos: en el caso 1 el candidato intenta comparar ciudades cuando solo le pedimos que describa la suya; y los casos 2 y 3 no saben qué responder (no por falta de habilidad) porque no captan contexto o este es general e improductivo. Repreguntar insistiendo solo les colapsa más.

¿Cómo hacerlo? Analizando la intención del candidato a través del contenido de su discurso. Las respuestas son inadecuadas: en el caso 1 se embarca infructuosamente en elaborar una respuesta de más nivel (comparar) que el propuesto (describir); la solución pasa por replantear el asunto volviendo al terreno deseado. Las preguntas cerradas son el mejor método para ello pues centran los temas y devuelven a lo concreto permitiendo hablar tranquilos (habla que se valorará según la habilidad del candidato). En el caso 1 debemos evitar la comparación mediante la pregunta "¿Cuál es tu ciudad?" y luego volver a pedir su descripción. En el 2, debemos "sacarle" de lo abstracto preguntando quizás "¿Tienes algún plan para tus vacaciones?". En el 3, debemos evitar pedir clarificaciones pues requieren mayor elaboración discursiva de lo pretendido (solo una comparación o cierto análisis).

### *2.4 Estrés comunicativo y muestra evaluable*

Todo esto nos plantea dos aspectos más, fundamentales: el estrés comunicativo y la creación de una muestra de lengua evaluable (Martínez, 2011).

El estrés se puede dar, incluso en nativos, ante situaciones difíciles como hablar ante un superior, alguien que intimide, queriendo ocultar información,

etc. Y su efecto es un colapso que puede hacer decaer el nivel del lenguaje o de su contenido informativo hasta cometer errores en los que incluso falle la gramática. Por ello no todas las muestras de discurso obtenidas en una evaluación oral son muestras evaluables y habrá que seleccionar las representativas del nivel de los candidatos y no las producidas por colapso, por ejemplo al arriesgar más de lo que requiera una tarea o actuando bajo un estrés imposibilitador. Actuando con profesionalidad, estas muestras fallidas deberían ser reorientadas o, al menos cuando estemos calificando, identificarlas y desecharlas, evitando que influyan en la nota final que no debe verse influída por elementos ajenos o elementos no considerados en el constructo inicial.

Veamos 2 ejemplos de la misma persona hablando durante el mismo examen, uno de los cuales no es evaluable:

*1. No hay nada en X (ciudad), es solamente bares, solamente. No cultura. Comercios centrales, sí, y bares sí y nada más. Un poquito cine, uno museo, pequeños parques, pequeñitos, y nada más. Es una playa muy bien pero es en el sur de X. Es una hora y media entre A (ciudad) y B (ciudad).*

*2. Es, es un ciudad muy antiguo y tiene… calles muy … muy pequeña y tiene … ah… tiene una lugares de ah… católicos y … los ah… tienes muy grandes ah… torres y ... el ela el lugar es en… es en una... montaña y hay un río en en ... Hay un río y entonces ah… la tierra alto y muy alto es lejo y lo río y alto muy tierra. Ah…y tiene una sistema de escaletores paran personas ah… escalar ah… alto arriba montaña.*

Este segundo extracto está generado al exigir al candidato una tarea superior a sus posibilidades. El colapso se aprecia en sus dudas, titubeos, pausas y vueltas atrás, un lenguaje no apto para la calificación: tampoco hay que dejar que genere un efecto halo que puede lastrar la consideración de las habilidades reales del candidato. Para calificar el nivel de este candidato tenemos la muestra 1, no la 2. Recordemos que se debe evaluar en positivo, sobre lo que se puede hacer, no sobre lo contrario.

### *2.5 Discusión*

Sabiendo que la nota obtenida es resultado de diversas variables, de las que solo una es la habilidad lingüística, concluimos que evaluar la lengua oral ne-

cesita constructos más elaborados, escalas que capten la riqueza generada, y formación: conocimiento de los objetivos de cada tarea y preguntas; métodos de elicitar que obtengan lo mejor del candidato y con instrucciones que no prohiban explorar las habilidades en profundidad; y entender el proceso de una entrevista y lo que en ella pase. Es necesario no generar nerviosismo o desorientar al candidato y saber atender al estrés que se genera, detectarlo y minimizarlo de forma que el resultado sea una muestra evaluable, suficiente y válida.

Tarea complicada pero posible, igual que para periodistas, policías o sicólogos. Trabajamos el mismo campo de actuación con la lengua y no debemos escudarnos en complicaciones falsamente irresolubles.

## REFERENCIAS BIBLIOGRÁFICAS


Alderson, J. Ch. y Banerjee, J. (2001), "Language Testing and Assessment (Part 1)", *Language Testing*, 34: 213-37.

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

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

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

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

Byrnes, H. (1989), "Strategies for eliciting speech" en *The ACTFL Oral Proficiency Intreview*, (K. BUCK, Ed.)Yonkers, Nueva York: ACTFL.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3*, 229–242.

Deville, C., y Chalhoub-Deville, M. (2006). Old and new thoughts on test score variability: Implications for reliability and validity. In M. Chalhoub- Deville, C. A. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp. 9–25). Amsterdam: John Benjamins.

Brooks, L. (2009), Interacting in pairs in a test of oral proficiency: Co-constructing a better performance, *Language Testing*, 26 (3), 341–366

Fulcher, G. (2003). *Testing second language speaking.* London: Pearson.

Fulcher, G. y Marquez Reiter, R, (2003b), Task difficulty in speaking tests, *Language Testing* 20 (3), 321-334

Fulcher, G., Davison, F, Kenn, J., (2011), Effective rating scale development for speaking tests: Performance decision trees, *LanguageTesting* 28 (1) 5–29

García Laborda, J., Otero de Juan, N., Bakieva, M., (2015), "Co-Participation In Oral Paired Interviews: Preliminary Findings Of The OPEN-PAU Project", en *Procedia - Social and Behavioral Sciences 191*, 559 – 563 [pdf], [en línea: 7/04/2017) www.sciencedirect.com/science/article/pii/S1877042815028748

He, L., y Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, *23*, 370–401.

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

Luoma, S. (2004), *Assessing Speaking*, Cambridge: Cambridge University Press.

Martínez Baztán, A. (2008), *La evaluación oral, una equivalencia entre las "guidelines" de ACTFL y algunas escalas del MCER.* Editorial de la Universidad de Granada. [pdf], [en línea: 7/04/2017) https://hera.ugr.es/tesisugr/17457853.pdf

Martínez Baztán, A. (2011), La evaluación de lenguas: garantías y limitaciones, Octaedro, Barcelona.

May, L. (2006). An examination of rater orientations on a paired candidate discussion task through stimulated verbal recall. *Melbourne Papers in Language Testing, 1*, 29–51.

O'Loughlin, K. (2001). *The equivalence of direct and semi-direct speaking tests.* Cambridge: Cambridge University Press.

Ross, S. (1998), "Divergent Frame Interpretations in Language Proficiency Interview Interaction", en YOUNG, R. / HE, A. W. (eds.), *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam/Philadelfia: John Benjamin: 333-353.

Shohamy, E. (2001), *The power of tests: A critical perspective on the uses of language tests*. London: Longman.

Swain, M. (2001). Examining dialogue: Another approach to content specifications and to validating inferences drawn from test scores. *Language Testing, 18*, 275–302.

Taylor, L. (2001). The paired speaking test format: Recent studies. *University of Cambridge ESOL Examinations Research Notes, 6*, 15–17.

Warren, K. (2008), *La Evaluación Oral y el Estrés comunicativo: Identificar,*

*localizar y explicar casos de estrés comunicativo en dos modelos de examen distintos.* [Memoria inédita del II Máster de ELE del Departamento de Lengua Española de la Universidad de Granada. Director: Martínez Baztán, A].

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

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press