EL PROBLEMA DE LA EVALUACIÓN DE LA PRODUCCIÓN ORAL INTERACTIVA EN EL AULA ELE

María Cecilia Ainciburu

Universidad de Siena

1. Introducción

Este trabajo tiene como objetivo reseñar una serie de criterios necesarios para construir las especificaciones de examen que garanticen, en lo posible, un muestreo relevante de la competencia oral de un alumno o de un grupo de alumnos. Este muestreo es necesario desde dos puntos de vista: si somos docentes, porque tenemos que saber lo que nuestros alumnos han aprendido -no importa si estamos trabajando en una institución o en clases privadas- y, como investigadores, porque nuestra experimentación se basa generalmente en un análisis de actividades didácticas, actividades a partir de las cuales extraemos nuestros datos empíricos.

La ocasión práctica para esta reflexión nace de la solicitud, por parte de la Facultad de Economía de la Universidad de Siena, de realizar una evaluación de los alumnos después de tres cursos de estudios del español como lengua extranjera (ELE). Tal evaluación sirve para otorgar a los alumnos un certificado universitario que atestigua, por un lado, su competencia lingüística plena del nivel B1 -según se define en el *Marco común europeo de referencia para las lenguas* [...] (MCER)- y, por otro, las competencias parciales de tipo profesional de nivel B2 (tal como se definen en los descriptores de la *Association of Language Testers in Europe* (ALTE) y, en este caso, aplicadas a los estudios económicos). Las reuniones de profesores de lengua de la Universidad, que tenían como intención la de homogeneizar los criterios de evaluación de las

Centro Virtual Cervantes

diferentes competencias, demostraron que era bastante simple obtener un amplio acuerdo sobre los objetivos generales y los géneros sobre los que evaluar la competencia escrita, mientras que evidenciaron grandes discrepancias de criterio cuando se trató de evaluar la destreza oral y, especialmente, la producción oral interactiva. Este trabajo condensa algunos de los puntos relevantes que debimos considerar y busca un marco teórico y unos datos experimentales que respalden nuestra práctica de evaluación certificativa.

Para ello:

a) Realizaremos una serie de consideraciones generales sobre el problema de la evaluación oral.

b) Distinguiremos los diferentes tipos de géneros orales naturales, con especial atención a los que implican interacción.

c) Señalaremos los géneros propios de la evaluación pedagógica, intentando poner de manifiesto sus aspectos positivos y negativos. En este sentido, apoyaremos las consideraciones teóricas con los antecedentes logrados en las pruebas realizadas para obtener datos acerca del género más apropiado para evaluar la producción oral interactiva[1].

d) Por último, avanzaremos una serie de conclusiones que pueden servir, ya sea para aplicar la evaluación de la producción oral interactiva al contexto académico, como para utilizarla en otros casos de *fines específicos*.

2. Evaluación de la producción oral

2.1. Consideraciones generales sobre el problema de la evaluación de la producción oral

Dentro del problema de la evaluación sumativa, juzgar el nivel de la producción oral de un alumno -lo decimos desde un principio-, representa el pro-

1. Se presentaron a los exámenes de lengua, nivel B2, 120 alumnos que fueron evaluados por medio de una entrevista (27 de febrero de 2006). Las lenguas evaluadas con entrevistas comparables fueron español, francés, inglés y alemán. Los treinta alumnos de cada grupo recibieron un cuestionario posterior a la evaluación. Los docentes de la Facultad que vieron las entrevistas, directamente o grabadas, y respondieron al cuestionario fueron 12. Los cuatro entrevistadores eran desconocidos para los alumnos.

Centro Virtual Cervantes

blema más espinoso. En el análisis lingüístico general, así como en la lingüística aplicada, el aspecto más estudiado de la producción de los alumnos extranjeros es el de la comunicación escrita. Sólo desde hace pocos años empezamos a tener una buena base para estudiar la interacción oral de los hablantes nativos[2], pero carecemos de un corpus de producción oral de alumnos extranjeros (*learners corpus*) al que podamos hacer referencia. Éste es un problema grave, porque la investigación progresa a partir de la existencia de *datos compartidos*, mientras que, para el español, se trata generalmente de bases a las que sólo puede acceder un investigador o un pequeño grupo. Tener una base amplia de transcripciones, o mejor aún, de grabaciones, de la producción oral de alumnos de distintos niveles permitiría estudiar el fenómeno a partir de datos más sólidos.

La primera condición para evaluar con claridad la producción oral, desde el punto de vista exquisitamente experimental, sería poder aislarla -natural o artificialmente- de las otras tres destrezas, pero esto resulta imposible. Siempre que tengamos interacción deberemos considerar también una fase de comprensión auditiva y, seguramente, aún en los géneros con turnos de palabra limitados o nulos (como la entrevista, la conferencia, etc.), deberemos contar con una serie más o menos extensa de instrucciones o consignas escritas y, a veces, con una planificación gráfica. En los géneros de conversación, o bien nos basamos en la intercomprensión, o bien proporcionamos fichas de simulación con vacío de información para realizar una tarea; pero, según veníamos diciendo, en los dos casos resulta imposible aislar completamente la destreza que queremos evaluar.

En la práctica evaluadora los requisitos básicos de toda prueba: la validez, la fiabilidad y la viabilidad pueden quedar aseguradas por un equipo de validadores expertos (como en el caso de las certificaciones o de los exámenes oficiales), pero eso no resuelve el problema de la corrección *in loco* del examen... Sin corpus extensos en los que basar la práctica es difícil entrenar a un docente (o a un grupo de docentes en el caso de instituciones y certificaciones) para que *su concepto de nivel intermedio* sea igual al de otro profesor.

Si sumamos a esta falta de entrenamiento la procedencia multicultural de los alumnos ¿estamos seguros de que el *nivel de supervivencia* que se les

---

2. Nos referimos aquí al Proyecto *Val.Es.Co* (*Valencia Español Coloquial*), dirigido por Antonio Briz, o al corpus oral de la Real Academia Española, constituido a su vez como recopilación de otras bases de datos (como el *ACUA* de la Universidad de Alcalá, dirigido por Ana Cestero, entre otros).

Centro Virtual Cervantes

exige a los alumnos chinos es el mismo que se les pide a los alumnos ingleses? La teoría de la evaluación dice que sí; la práctica docente indica que estamos más dispuestos a *dejar pasar* errores a los alumnos que consideramos que tienen una dificultad mayor para aprender la lengua.

No entramos ahora en la consideración del dialecto del profesor, que se espeja en la producción de cualquier alumno extranjero, pero bien sabemos que la articulación de los sonidos también tiene su peso en los exámenes y que los examinadores no dejan de tener prejuicios al respecto, como enseña la sociolingüística (cf. Moreno, 1989). Aún si nos quedamos con este nivel básico del que hablábamos antes, ¿tiene sentido hablar de un *nivel de supervivencia* en presencia de alumnos nativos de lenguas afines al español, como el portugués o el italiano? Volveremos a considerar estos interrogantes al final de este trabajo, pero revelamos que las dificultades propias de este tipo de evaluación derivan de los rasgos típicos de la oralidad que supone un grado mayor de interrelación entre hablantes, un conocimiento diferente de las realidades dialectales, de los patrones de entonación y pronunciación de una lengua, etc.

Todos estos problemas, referidos a la evaluación de la producción oral, se asocian a un dato básico de la situación evaluativa que es el hecho de que no podemos utilizar pruebas objetivas, como en la evaluación de la producción escrita o en la de comprensión oral. Además, en muchas situaciones no institucionales, ni siquiera podemos triangular los resultados utilizando más de un evaluador. Por esa razón decimos que los exámenes orales son los más subjetivos y los menos fiables por naturaleza. Sin embargo, reflexionar sobre el problema y hallar criterios comunes nos permitirá realizar una evaluación más objetiva o, por lo menos, lo más objetiva posible.

En el propósito de adentrarnos rápidamente en los temas más técnicos de la evaluación de la producción oral, no queremos olvidar un factor fundamental: el emocional. Creemos que todos estamos de acuerdo en que estas pruebas son las que quedan más sesgadas, tanto por el nerviosismo del alumno, como por su personalidad y demás valores idiosincrásicos. Para intentar una evaluación más atenta de la producción oral del alumno, tanto en la evaluación como en la toma de muestra de hablantes nativos, se realizan actividades previas *para romper el hielo*. Estas actividades se excluyen, tanto sea de la evaluación como de la muestra experimental, en el caso del análisis de la conversación.

Centro Virtual Cervantes

Para evaluar la producción oral, además, es necesario considerar su rasgo distintivo: su modalidad de transmisión física y semiótica, que no significa sólo su realización a través del canal fónico-auditivo sino, también, la existencia de un contexto extralingüístico común y la co-presencia de hablante e interlocutor en la mayor parte de los casos.

Según el MCER (cap. 9.2.2.1.) las fases prácticas de la evaluación de cualquiera de las destrezas que considerar son:

1. *Construcción*: en la que se diseñan las tareas de evaluación, tomando en cuenta escalas de las actividades comunicativas específicas para verificar los objetivos pautados en relación con el nivel de competencia requerido.

2. *Autoevaluación del alumno y evaluación por el profesor*: dependiendo del tipo de evaluación que se realiza (autoevaluación, certificación, diagnóstico, etc.). Para realizarla con criterios comparativos se realizará según el caso:

- *Lista de control*: en la que se recojan por medio de una lista los descriptores de un nivel concreto o sus componentes explícitos, como *sé presentarme; digo dónde vivo; digo mi dirección; digo la edad que tengo*, etc., y *pregunto a alguien cómo se llama; pregunto a alguien dónde vive; pregunto a alguien qué edad tiene*, etc. (ejemplos del MCER).

- *Parrilla*: donde el perfil de la competencia se define a través de un cierto número de categorías concertadas (*Conversación; Debate; Intercambio de información*) y de niveles de actuación esperados (B1+, B2, B2+) (ejemplos del MCER).

3. *Informe*: se realiza un documento que permita ver cuáles han sido los resultados globales obtenidos con la prueba. Generalmente incluye un análisis cuantitativo y cualitativo de los datos en relación con los objetivos perseguidos.

Por problemas de espacio nos dedicaremos aquí sólo a un aspecto de los citados, el de la construcción de la prueba. Creemos que una parrilla concertada de corrección es hoy una práctica difundida entre los docentes, mientras queda pendiente (en el ámbito de la evaluación sumativa) el problema del informe final. Suponemos que en las instituciones que se dedican establemente a la certificación se realizará un informe final en cada turno de exámenes y sería deseable que, al menos en el caso del DELE, fuesen públicos las parrillas y el informe final. No sólo como forma de control y de transparencia, sino para contribuir a la mejora general de las técnicas de evaluación de ELE. Pero no queremos dejar de señalar algo que sabemos a ciencia cierta: esto es, que los

Centro Virtual Cervantes

docentes universitarios europeos no suelen tener a su disposición los recursos técnicos para realizar este informe y, a veces, tampoco la formación estadística para realizar el análisis de los datos. Las instituciones universitarias, muy interesadas en evaluar el proceso didáctico, no deberían limitarse a analizar los datos finales con criterios de productividad general, sino de calidad integral del proceso.

La consideración de los aspectos teóricos y del estado de la cuestión constituyeron un paso previo a la planificación de los exámenes y a la experimentación siguiente. El marco teórico resultante de las lecturas efectuadas, luego de sopesar diferentes aspectos de la interacción oral, fue desmoralizante: muchos autores cuestionan la validez de las pruebas existentes, y eso no sería nada (podríamos inventar nuevos modelos) si no fuese porque también se afirma que el concepto de autenticidad no puede aplicarse al contexto de un examen y que esto invalida la transferibilidad de los datos resultantes de las investigaciones (Alderson y Banerjee, 2002; McNamara et ál., 2002; Ellis, 2003).

Todas las objeciones relativas a la obtención de modelos de evaluación eficaces se condensan cuando queremos hablar de la evaluación oral. Para sintetizar las dificultades más importantes, diremos que los aspectos sobre los que resulta más difícil obtener acuerdos -y por lo tanto, consensuar baremos de corrección- son:

1. La complejidad de aspectos que integran la destreza oral y el peso relativo que debe concederse a cada uno; la globalidad de aspectos que conlleva la actuación del alumno y la dificultad de analizarlos contemporáneamente.

2. El carácter social que se vehicula en cada intercambio lingüístico y la dificultad objetiva para el examinador de desincorporar sus propios prejuicios de la evaluación general.

Una vez que el docente toma acto de las dificultades objetivas y subjetivas que los investigadores relevan, la necesidad de evaluar la competencia oral de un alumno es algo a lo que, a pesar de todo, no puede renunciar. Probablemente, como sugiere Ellis (2003), el único camino posible es el de la transparencia de los objetivos perseguidos, de los presupuestos teóricos asumidos y de los procedimientos que se adoptan para integrar el proceso de evaluación.

Bachman y Palmer (2003) indican que es necesario tomar distancias respecto a una sistematización absoluta de los diferentes aspectos de la produc-

Centro Virtual Cervantes

ción lingüística de un alumno y, sobre todo, del alcance cognoscitivo y operativo que nos ofrece una evaluación realizada con los métodos con los que contamos hoy. La investigación en el campo de la evaluación es, en este sentido, la menos desarrollada en el campo de la didáctica de las lenguas extranjeras. En este trabajo, el punto más relevante lo constituye la intención de desenmascarar una pretensión de objetividad que cada examinador tiene. En realidad, cada método de evaluación es, por naturaleza, subjetivo y sujeto a la combinación de una serie de factores ambientales y sociales que no pueden explicitarse completamente.

En la propuesta de Bachman y Palmer (2003), se señalan seis cualidades propias de un examen que contribuyen a la buena práctica de la evaluación: a) validez de construcción de una prueba; c) fiabilidad; d) impacto; e) interactividad; f) autenticidad; g) practicabilidad.

Estas cualidades no se colocan en una escala jerárquica o de prioridad operativa, pero todas ellas deberían considerarse: mientras se proyectan, dejando en claro cuál es la relación entre ellas, en la construcción de las parrillas de corrección y, sucesivamente, en el proceso de validación final.

2.2. Tipos de actividades comunicativas orales naturales y géneros correspondientes

El MCER propone una clasificación de las actividades comunicativas según éstas sean unidireccionales o bidireccionales, es decir, que impliquen sólo la producción o también la interacción, y da a las segundas un interés mayor.

> Las actividades de expresión tienen una función importante en muchos campos académicos y profesionales (presentaciones orales, estudios e informes escritos) y se les otorga un valor social determinado (las valoraciones realizadas de lo que se ha presentado por escrito o de la fluidez en la articulación de presentaciones orales). En la *interacción*, al menos dos individuos participan en un intercambio oral o escrito en el que la expresión y la comprensión se alternan y pueden de hecho solaparse en la comunicación oral. No sólo pueden estar hablando y escuchándose entre sí dos interlocutores simultáneamente; incluso cuando se respeta estrictamente el turno de palabra, el oyente por lo general está ya pronosticando el resto del mensaje del hablante y preparando una respuesta. Aprender a interactuar, por lo tanto, supone más que aprender a comprender y a producir expresiones habladas. Generalmente se atribuye gran importancia a la interacción en el uso y el aprendizaje de la lengua dado su papel preponderante en la comunicación. (MCER, 2002: 2.1.5.)

Centro Virtual Cervantes

La mayor importancia atribuida a las manifestaciones de interacción se justifican, entonces, con un interés social y con razones de mayor complejidad de las operaciones cognitivas utilizadas para desempeñarlas. Siguiendo esta división, podemos individualizar formas diferentes de situación comunicativa, propias de la vida real, en las que se verifican formas de comunicación unidireccionales o bidireccionales. Siguiendo esta división tales actividades se caracterizan como actividades de interacción y de expresión, como en el Cuadro 1 (MCER, 2002: 9.2.1.):

Cuadro 1: Fragmento de la sinopsis "categorías de las actividades comunicativas"

	Interacción (Espontánea, turnos breves)	*Expresión* (Preparada, turnos largos)
Orales	Conversación Debate informal Cooperación centrada en el objetivo	Descripción de su especialidad académica

El Consejo de Europa indica como *reales*, dentro del esquema propuesto las actividades del Cuadro 2, citadas a lo largo de los capítulos 2 y 9:

Cuadro 2: Elaboración personal de las formas propuestas por el MCER (2002: 9.2.1.)

Bidireccionales o de interacción	*Unidireccionales o de expresión*
Realizar una transacción Conversación casual Discusión informal Discusión formal Debate Entrevista Negociación Realizar un proyecto común Cooperación para alcanzar un objetivo práctico	Leer un texto escrito en voz alta Hablar ayudándose con un soporte escrito o visual Recitar un papel Hablar espontáneamente Cantar

Centro Virtual Cervantes

Estos tipos de discurso deben tenerse en cuenta a la hora de realizar la evaluación de la competencia oral, tanto unos como otros, porque independientemente de su complejidad cognitiva, para que un examen sea válido debe tener como objetivo varios ejemplos de comportamiento lingüístico, dado que el alumno necesita tener una competencia general en todos ellos, aún cuando sus necesidades lo lleven a privilegiar más una que otra:

> Se va aceptando paulatinamente que la evaluación auténtica requiere el muestreo de una serie de tipos de discurso relevantes; por ejemplo, respecto a los exámenes de expresión oral, una prueba elaborada recientemente es ilustrativa al respecto. En primer lugar, hay una *conversación* simulada que funciona como introducción y después se da un *debate informal* sobre temas por los que el examinando declara su interés. A esto le sigue una fase de *transacción* en forma de actividad -ya sea cara a cara o simulada- de búsqueda de información por teléfono. A continuación, se desarrolla una fase de *expresión* basada en un *informe* escrito en el que el examinando ofrece una descripción de su área de especialidad académica y de sus planes. Por último, hay una *cooperación* centrada en el *objetivo,* una tarea de consenso entre los examinandos. (MCER, cap. 9.2.1.)

Siguiendo la clasificación de Anderson y Ciliberti (2002), que proponen la etiqueta de *géneros monológicos y dialógicos*, en vez de uni y bidireccionales, pondremos atención no sólo en la naturaleza de esta comunicación, sino también en el grado de copresencia de sus participantes y en la libertad de toma de palabra que el género supone.

Cuadro 3: Elaboración personal de las formas propuestas por Anderson y Ciliberti (2002: 96)

	Intercambio comunicativo	*Grado de presencia*	*Libertad de los turnos de palabra*
Conversación natural	sí	cara a cara	completa
Conversación telefónica	sí	mediada	completa
Debate, entrevista, interrogatorio y similares	sí	cara a cara	concertada
Conferencia o discurso (monólogo)	no	cara a cara	inexistente
Transmisiones radiofónicas o televisivas	no	mediada	inexistente

Centro Virtual Cervantes

Queda claro que a una menor libertad de los turnos de palabra corresponde un mayor grado de formalidad entre los hablantes y una mayor posibilidad de planificación del discurso (Cuadro 3). Paralelamente, los hablantes nativos utilizan con mayor frecuencia los géneros que suponen una mayor libertad en la toma de turno de palabra, mientras que algunos de estos hablantes pueden pasar una vida entera sin realizar discursos formales.

Desde el punto de vista de la didáctica de las lenguas nos parece obvio que debemos dedicar la mayor parte de nuestro empeño a enseñar, practicar y evaluar las formas de comunicación más frecuentes y naturales, o sea la conversación en sus diferentes modos. Así nos acercamos al método comunicativo, a la realización de tareas y, en forma más específica, al enfoque léxico y a la enseñanza de fragmentos prefabricados de lengua (*chunks*) que facilitan la comunicación y aumentan la fluencia de la lengua hablada. Contrariamente, los estudiosos de la conversación sugieren que la conversación natural es uno de los géneros más difíciles de aprender y que, incluso, algunos nativos se muestran poco eficaces en la realización de este género. Si sumamos las dificultades de un género que todos damos por *aprendido* a las complicaciones propias de la situación pragmática para un extranjero, la situación se pone aún más engorrosa. ¿Cómo evaluaremos un género del cual no conocemos todas las vertientes y que se supone, deberíamos enseñar en niveles avanzados?

La situación comunicativa que llamamos *conversación natural* se decanta en una diferente estructura y utilización de los recursos estrictamente lingüísticos, como demuestran los análisis realizados por el grupo *Val.Es.Co*. La lengua de la conversación, respecto a los usos formales y al uso escrito, se caracteriza por la simplificación del paradigma verbal y la reducción del número y de la variedad de los conectivos, por una diferente relación entre el léxico y la gramática dado que resulta más frecuente el uso de la deixis y de las llamadas *palabras gramaticales* y menos frecuente la ocurrencia de palabras lexicalizadas. Esta constatación debería tener consecuencias fundamentales en la calificación de las pruebas orales, donde sin embargo, prevale a veces un canon del escrito o de *exposición académica*, especialmente en los esquemas de evaluación de la riqueza léxica, de la variedad de estructuras utilizadas o de la utilización de estructuras paratácticas.

Un problema bien evidente, cuando programamos una evaluación oral interactiva, es saber quién será el interlocutor del alumno evaluado. En este sentido, existen dos prácticas posibles: o se diseña una actividad entre pares y se evalúan contemporáneamente dos alumnos (interacción horizontal) o el

Centro Virtual Cervantes

alumno realiza la actividad con un evaluador (interacción vertical). Ambas posibilidades tienen sus ventajas y sus desventajas. Sin dudas, la actividad entre pares es algo que semeja más al género de la conversación natural y, por eso, resulta preferido en muchos casos. Por otra parte, las certificaciones utilizan un evaluador como contraparte. En el caso específico de los alumnos universitarios, si bien se aconseja ampliar las competencias a lo cotidiano, es probable que el alumno se encuentre en algún momento en el país extranjero (en ocasión de un Erasmus o de un traslado por motivos de trabajo) y que tenga que hablar en condiciones de interacción vertical. No podemos asegurar que utilizará un registro más que otro y estamos seguros, sobre todo, que uno no es más real que el otro.

2.3. Tipos de actividades comunicativas orales pedagógicas

Las actividades que utilizamos para enseñar un tema en clase y para evaluar suelen tener la misma estructura. No siempre resultan *auténticas* si definimos ese término en modo estricto, pero lo cierto es que utilizan una serie de operaciones cognitivas que constituyen la base de los comportamientos lingüísticos que los hablantes utilizan en el mundo real. Nadie discutiría, para ser gráficos, el hecho de que antes de conducir un avión un aspirante a piloto deba usar una cabina virtual de vuelo. Las actividades interactivas que pueden utilizarse en la evaluación son[3]:

1. Conversación académica o interacción casi libre.

2. Interacción dentro de un esquema de actuación (de situación, *role-play*, etc.).

3. Interacción guiada por medio de fichas (actividades del tipo de las que se encuentran en los libros *Acuerdos*, *De dos en dos* o *Pretextos para hablar*).

El primer tipo, la conversación académica, es más un modelo al que se aspira que una forma real de conversación. Ya sea porque no se produce espontáneamente (la situación de examen no lo permite), ya sea porque el alumno tiene como interlocutor a un docente. Como ya dijimos, esta situación de interacción vertical se produce también en la vida real, pero en ese caso el

3. Un cuadro completo de las formas de evaluación, incluidas las formas utilizadas para la producción oral fonológica, en Bordón (2004).

Centro Virtual Cervantes

docente no *conduce* lingüísticamente el juego. Dentro de este tipo de evaluaciones interactivas se coloca la entrevista, género preferido en las certificaciones, que requiere un gran entrenamiento por parte del examinador y que no proporciona resultados deseados cuando el examinador coincide con el profesor con el que el alumno está familiarizado. Quienes defienden este tipo de examen (normalmente las certificaciones lo adoptan como forma única de evaluación oral), lo hacen subrayando su aproximación al género natural y la posibilidad de que, en niveles altos de competencia, se produzca realmente una conversación. Quienes rechazan la entrevista, aducen que las muestras de lengua que produce son difícilmente analizables y, casi seguramente, no comparables, aún dentro de un grupo de alumnos más o menos homogéneos.

En nuestro caso, la experiencia de prueba de la entrevista como modo de evaluación en ámbito académico del nivel B2 nos dio como resultado principal la facilidad de administración del examen: el entrevistador tiene que estar muy entrenado para ser eficaz, pero el alumno puede no estarlo. Probablemente, por esa razón, fue la entrevista la forma de examen preferida por los alumnos (41%) y la menos por parte de los docentes (26%) (resultados en el Cuadro 4). Entre las razones proporcionadas por los estudiantes[4]: "es la más apropiada en la Universidad", "puedes conducir la conversación hacia lo que sabes", "el profesor, si quiere, puede ayudarte". Entre las razones, opuestas, de los docentes: "No se puede ser igual en las primeras y en las últimas entrevistas", "los resultados no se pueden comparar", "depende demasiado de la personalidad del alumno".

El segundo tipo, la interacción dentro de un esquema de actuación, puede responder a formas de uso real de la lengua (con todas las salvedades dadas por la situación de examen que ya hemos hecho para hablar de la entrevista).

Los datos que nos proporcionaron los cuestionarios que realizamos en nuestra universidad fueron poco confortantes. Éste fue el tipo de prueba que estudiantes y profesores coincidieron en evaluar como la menos deseable (respectivamente el 27% y el 32% de las preferencias). No fue un resultado previsto, porque suponíamos que la actuación en *situaciones* es un recurso bien conocido en ámbito de la didáctica de las lenguas y la simulación, uno de los preferidos de la didáctica con fines específicos empresariales y econó-

4. Los cuestionarios estaban hechos en italiano, por lo que todas las respuestas que se incluyen han sido traducidas.

Centro Virtual Cervantes

micos. El tema elegido en todas las lenguas era *la entrevista de trabajo*. Las razones del los alumnos fueron: "uno queda a merced del compañero que le toca", "no es apropiado para la universidad", "yo preguntaba mucho y mi compañero respondía sólo sí o no". Como puede verse aquí la interacción horizontal es fuente de problemas y, para los alumnos, no resultó la preferida. Los docentes argumentaron en contra por razones similares a las de los alumnos.

El tercer tipo de actividad, la interacción guiada por medio de fichas, necesita una buena preparación previa del material por parte de los docentes, luego, la fase en la que se procede al examen -la administración- es fácil sólo si los alumnos están acostumbrados a trabajar con este tipo de fichas. El hecho de proporcionar una base común para la actuación es un arma de doble filo: por un lado, tiende a emparejar las actuaciones de los alumnos en tiempo y producción lingüística; por otra, puede proporcionar -con las instrucciones o con el material de trabajo- un *input* gramatical o de léxico excesivo.

En el caso de nuestra experimentación, este tipo de interacción fue la preferida por los docentes (40%) y la segunda predilecta en el caso de los alumnos (32%). En todos las lenguas, el material proporcionado requería discutir en base al currículo de tres candidatos en mérito a la vacante de un cargo empresarial y los alumnos habían realizado ejercicios de ese tipo en clases. Los alumnos observaron que: "es similar a lo de la clase, por lo que, por suerte, no hay sorpresas", "prefiero no tener que actuar", "si no te acuerdas algo puedes "llenar" el espacio leyendo un poco". Los docentes observaron que "es más fácil dar a todos cosas similares y que unos temas no sean más difíciles que otros", "lo que los alumnos producen es similar", "si la función requiere una construcción, el alumno no tiene otra salida".

Como se ha podido ver, las características positivas y negativas de cada una de las formas de evaluación, más allá de las preferencias, queda clara en las justificaciones de los estudiantes y de los docentes que intervinieron en la experimentación.

Centro Virtual Cervantes

Cuadro 4: Resumen de los datos obtenidos a partir de los cuestionarios posteriores al examen

	Preferencia alumno	Preferencia docente
Entrevista	41%	26%
Role-play	27%	34%
Fichas	32%	40%

3. Conclusiones

Un problema que consideramos aquí fue el de la comparabilidad de los exámenes. En modo intuitivo deberíamos suponer que si los mismos alumnos se someten en la misma fecha a los tres tipos de evaluación los resultados que se obtienen deberían ser similares. Los resultados se obtuvieron realizando el promedio entre las correcciones de dos docentes diferentes que trabajaban con parrillas concertadas (Cuadro 5). Dichas parrillas se construyeron utilizando los descriptores correspondientes a la *interacción* del *Manual preliminar para la evaluación de la Comunidad Europea* (2003) (en adelante *MPE*). Los resultados fueron los siguientes:

Cuadro 5: Calificaciones obtenidas en el examen

	Porcentaje de aprobados	Porcentaje de notas altas (más de 27/30)
Entrevista	77%	59%
Role-play	71%	19%
Fichas	69%	32%

Se note que el porcentaje de alumnos aprobados es mucho más alto (unos ocho decimales) en la entrevista respecto a la actividad de fichas y, aun más, una diferencia de 27 puntos en el porcentaje de los alumnos que obtienen una calificación alta (más de 27/30). Dado que la diferencia es tan alta y significativa, tiene que existir algún tipo de razón fuerte que explique esta diferencia.

Centro Virtual Cervantes

Es fundamental explicar el hecho de que las notas no se distribuyan en modo homogéneo en la muestra.

La explicación que nos hemos dado, a partir de la descripción que hemos realizado en los puntos anteriores, es que existen dos factores que discriminan en su combinación las actividades y que podrían explicar la distribución de los resultados; a saber: el tipo de interacción, horizontal o vertical, y el apoyo escrito. En la combinación esquematizada en el Cuadro 6:

Cuadro 6: Combinación de factores en las actividades

	Interacción horizontal	Interacción vertical	Apoyo escrito
Entrevista		✓	
Role-play	✓		
Fichas	✓		✓

Podríamos concluir que funcionan como factores positivos ya sea la interacción vertical como el uso de un apoyo escrito (fichas)[5]. Desgraciadamente el número acotado de informantes y la cantidad de factores que intervienen en la situación de examen, no nos permiten obtener resultados más sólidos.

En modo marginal, los resultados confirman la utilidad de efectuar una evaluación que contenga, al menos, dos tipos de actividad diferente, como ya sugerido por el MCER y explicitado en el segundo parágrafo de este trabajo. En la evaluación de destrezas integradas encontraremos una respuesta más justa, dado que en la vida de todos los días no aislamos las destrezas. Los estudios de didáctica a partir de destrezas aisladas resultan, sin embargo, útiles para comprender los factores que entran en juego cada vez que activamos una destreza. Si se nos permite el parangón, primero la aritmética y, luego, la geometría.

5. Un análisis factorial, realizado sobre un número mayor de alumnos y referido a tres fechas de examen, proporcionó en el mes de julio, resultados positivos para esta hipótesis. Obviamente se comparó el mismo nivel y tipo de evaluación, pero los alumnos examinados no eran los mismos.

Centro Virtual Cervantes

Bibliografía

ALDERSON, J. C. y J. BANERJEE: «Language Testing and Assessment (Part 2)», *Language Teaching,* 35, 2002, 79-113.

ANDERSON, L. y A. CILIBERTI: «Monologicità e di(a)logicità nella comunicazione accademica», *Sul dialogo. Contesti e forme di interazione verbale*, (Ed. BAZZANELLA, C.), Guerrini Studio, 2002, 92-105.

BACHMAN, L.: «Some Reflections on Task-Based Language Performance Assessment», *Language Testing*, 19 (4), 2002, 453-476.

ELLIS, R.: *Task-Based Language Learning and Teaching*, Oxford: Oxford University Press, 2003.

MCER: INSTITUTO CERVANTES (Trad.): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación,* Madrid: MECD-Anaya, 2002. (COUNCIL OF EUROPE: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe, 2001), [en línea] <http://cvc.cervantes.es/obref/ marco/>

MCNAMARA, T., K. HILL y L. MAY: «Discourse and Assessment», *Annual Review of Applied Linguistics*, 22, 2002, 221-242.

MORENO, F.: «Sociolingüística y educación», *Resla*, 5, 1989, 9-21.

YUAN, F. y R. ELLIS: «The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production», *Applied Linguistics*, 24/1, 2003, 1-27.

Centro Virtual Cervantes