Sobre el diseño de pruebas orales

Carmen Pastor
Instituto Cervantes Múnich

En este artículo hablaremos sobre el diseño de una prueba para la evaluación de la interacción oral que se adecue a las necesidades y recursos de nuestro entorno, entendiendo como prueba oral aquella en la cual se anima a una persona a hablar y es entonces evaluada en base a esa intervención.

1. Introducción

Durante la comunicación oral las personas que intervienen intercambian su papel continuamente y pasan de hablante a oyente de forma fluida y rápida. Eso hace que la conversación suela ser un sistema de comunicación de dos vías.



Figura 1

Las situaciones en las que solo una persona habla y otros solo escuchan (como en una conferencia, un discurso, un responso, etc.) se dan en menor medida y son situaciones que requieren otras habilidades, por loque es necesario pensar bien si son habilidades que queremos evaluar o mejor dicho si estaban incluidas en nuestro currículo.

La característica del papel interactivo hablante-oyente distingue una buena prueba oral de otros tipos de pruebas de lengua; como las de comprensión oral, lectura o escritura que presentan una serie de preguntas para conseguir una serie de respuestas y que no son claramente interactivas en este sentido. Así pues, cuando al modelo de la *figura 1* le añadimos a alguien que observa *y* evalúa el habla producida, obtenemos un proceso de evaluación que lo convierte en prueba (Underhill, 2003: 2).

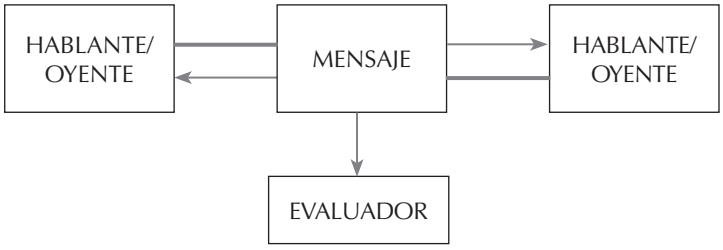


Figura 2

Más adelante hablaremos de modelos de examen, técnicas y criterios de puntuación entre los que podemos elegir para confeccionar una prueba oral útil que se adapte a nuestro entorno de enseñanza, pero antes quisiera detenerme en ese concepto de utilidad que Bachman y Palmer (2002: 17) apuntan como la característica más importante de una prueba de evaluación. Ellos definen esa *utilidad* como una función con cinco propiedades: *fiabilidad, validez, autenticidad, impacto y viabilidad*.

1. La prueba debe ser fiable. Esto tiene mucho que ver con los evaluadores; decimos que una prueba es fiable si los resultados que ofrece tienen estabilidad, es decir, si evaluando siempre a la misma persona en la misma situación obtuviéramos los mismos resultados. Se distinguen tres tipos:

— *fiabilidad interevaluadora,* se da si evaluadores diferentes corrigen el mismo examen y coinciden en la valoración siguiendo los baremos de corrección correspondientes;
— *fiabilidad intraevaluadora,* se da si un mismo evaluador corrige un mismo examen en diversas ocasiones y da una única valoración;
— *fiabilidad extrínseca,* se da si condiciones circunstanciales de administración y ejecución de la prueba (calendario, horario, lugar, presentación, etc.) no tienen incidencia en los resultados.

2. La prueba debe ser válida. La validez es la propiedad más importante y su idea es muy básica: ¿evalúa la prueba lo que se supone que debería evaluar? En nuestro caso, ¿está la prueba evaluando la capacidad del examinado para comunicarse oralmente en español? Se distinguen cuatro tipos de validez:

— *validez de contenido,* se da si la prueba evalúa exactamente los objetivos que había previsto;
— *validez de predicción,* se da si la prueba tiene capacidad de predicción de las conductas que tendrán los examinados en la vida real;
— *validez recurrente,* se da si una prueba obtiene resultados semejantes a otras pruebas iguales validadas previamente, o si diversas ediciones de una misma prueba obtienen puntuaciones iguales con los mismos examinados;
— *validez de constructo,* se da si la prueba es coherente con la teoría lingüística *y* pedagógica de la institución que diseña y administra las pruebas, es decir, que se correlaciona con el currículum educativo y con las prácticas de clase.

3. La prueba debe ser auténtica. Bachman y Palmer (1996: 23-26) distinguen entre *autenticidad situacional y autenticidad interaccional*. La primera exige que las tareas de una prueba guarden relación con actividades —en este caso, de interacción oral— que se desarrollan en la vida real o que son posibles en ella. Es

decir, cuanto más se parezca una tarea evaluadora a alguna actividad lingüística que ocurre de forma natural en la vida cotidiana, más auténtica será la primera.

La autenticidad interaccional se da si entre la tarea de evaluación y el examinado «las operaciones que se les exigen a los candidatos se corresponden con los procesos que ocurrirán de forma natural al interactuar con estos textos» (Bordon, 2000: 81). Es decir, que una tarea será interaccional —o tendrá autenticidad interaccional— si exige del examinado el conocimiento, destrezas y habilidades necesarias que necesitaría para actuar en una situación real similar de uso de la lengua.

4. La prueba debe intentar proporcionar un impacto positivo. Un tipo de impacto que pueden tener las pruebas de evaluación en la enseñanza de idiomas es el denominado *washback*. Daniel Cassany (1993: 10) cuenta como

> un amigo, profesor de inglés en una academia y redactor de los exámenes finales, [...] había conseguido cambiar la didáctica de las clases de sus colegas cambiando el examen final de curso. Había diseñado un examen mucho más comunicativo y funcional que el anterior; como los profesores de la academia querían que sus alumnos aprobaran, no tuvieron más remedio que cambiar el enfoque de sus clases.

Esto es precisamente el efecto *washback:* la influencia de la prueba en la enseñanza, aprendizaje e incluso en el currículo del centro. Al desarrollar una prueba debemos intentar proporcionar un impacto *washback* positivo, que sea beneficioso y no lo contrario.

Otro tipo de impacto a evitar son las desviaciones *(bias)* que pueden aparecer cuando una prueba favorece a unos examinados pero desfavorece a otros. Esto puede pasar cuando se utilizan, por ejemplo, técnicas que les son familiares a unos pero no a otros (como la descripción de historietas), o en las que unas personas se sienten más cómodas y otras más incómodas (como los juego de roles). De ahí que se recomiende la utilización de técnicas diferentes, conocidas por los examinados y con instrucciones simples o incluso en el idioma de la persona evaluada (Bachman y Palmer, 1996; 29-35).

5. La prueba debe ser viable. La prueba debe ser práctica y posible en la situación en la que se administra. Hay que preguntarse: ¿Con qué recursos cuento (personas, tiempo, espacio, equipos técnicos)? ¿Qué restricciones tengo a causa de los medios e instalaciones, el tiempo disponible (tanto para la administración de la prueba, como para su corrección y puntuación)? ¿Quién evalúa? ¿Cuál es la estructura de los cursos?

A continuación revisaremos algunos modelos de pruebas, de técnicas y de criterios de evaluación para el diseño de una prueba oral útil, que intente abarcar en todo momento estas propiedades recién descritas.

2. **Modelos de pruebas**

En la actualidad hay mucho interés en la evaluación del dominio oral de la lengua puesto que la enseñanza de idiomas extranjeros, especialmente en los primeros estadios, está más dirigida hacia la expresión y comprensión oral. Obviamente esto debería verse reflejado en la evaluación, pero en muchas ocasiones no se da por diversos factores.

Por una parte, la prueba oral suele ser más difícil de diseñar, administrar y puntuar (espacio, horarios, profesores...). Por otra parte, en este tipo de pruebas se necesitan examinadores bien entrenados, tanto para su papel de evaluadores como para el de entrevistadores [1]. Además, la prueba oral puede ser menos fiable, debido a la valoración subjetiva [2] que implica y que puede variar de una persona a otra, pero al mismo tiempo puede ser mucho más válida ya que cuando evaluamos la capacidad de una persona de actuar en una lengua extranjera queremos saber hasta qué punto es capaz de comunicar con otros y no cómo maneja un objeto artificialmente construido llamado prueba oral.

Vamos a ver algunos modelos generales de pruebas orales para seleccionar aquella o aquellas que se adecue/n más a nuestro contexto.

2.1 *Autoevaluación*

Esta es la forma de evaluación más fácil, barata *y* rápida, *y* puede constituir una prueba por sí misma o ser parte de otra. La autoevaluación, en contraste con la evaluación realizada por otras personas, es aquella en la que las valoraciones respecto a la competencia lingüística las hace la misma persona que se evalúa.

Cada uno de nosotros, al hablar con otras personas, evalúa constantemente su éxito comunicativo. Después de un periodo de tiempo tenemos una impresión general de hasta qué punto podemos comunicar en una lengua extranjera, y esa impresión general es la que queremos obtener en una autoevaluación. En este sentido, el estudiante está en la mejor posición para definir lo bien o mal que habla, mientras que una prueba oral se basa en una muestra de habla de diez minutos.

---

[1] Entendemos como entrevistador al interlocutor que interactúa con la persona que se examina y que está entrenado para desarrollar dicha función. Por evaluador entenderemos a la persona encargada de observar la producción oral del examinado y de calificarle.

[2] La evaluación subjetiva se diferencia de la llamada evaluación objetiva por el método de puntuación. La evaluación es objetiva si no requiere el juicio de un evaluador, como en un ejercicio de selección múltiple, en el que la respuesta correcta equivale, por ejemplo, a un punto. La evaluación subjetiva, en cambio, es una valoración que realiza un evaluador. En la evaluación directa (una entrevista personal, escribir una carta) generalmente se otorgan calificaciones según un valor o juicio. Eso significa que la decisión respecto a la calidad de la actuación del alumno se toma subjetivamente, teniendo en cuenta factores concretos y haciendo referencia a líneas generales o bien a determinados criterios. La ventaja de un enfoque subjetivo y directo es que puede verse la actuación de la lengua en conjunto, de forma compleja *(Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* —en adelante MCER—, 2002: 188-189).

Esta forma de evaluación capacita a los estudiantes a tomar más responsabilidades en su proceso de aprendizaje y cualquier programa de enseñanza que dé importancia a la autonomía del alumno debería entrenar a sus estudiantes para evaluarse a sí mismos. Pero la habilidad de un alumno para determinar el nivel de su propio dominio de la lengua tiene ciertos límites ya que carece de la experiencia y formación que el profesor o evaluador sí tiene para comparar al estudiante con un estándar externo.

Asimismo, la autoevaluación no parece ser muy fiable por sí misma debido a una variedad de factores que afectan al juicio de la gente. Estos factores pueden ser conscientes y deliberados como una sobrevaloración para impresionar y conseguir un trabajo, o infravaloración para quedarse en un curso con un nivel que no suponga ningún reto o esfuerzo. Los factores inconscientes incluyen diversos niveles de autoconfianza y percepción de los alumnos: un historial poco espectacular de logros académicos llevará al alumno probablemente a infravalorarse, las mujeres en general se valoran por debajo que los hombres, y la gente mayor por debajo de los jóvenes (Underhill, 1987: 23).

El potencial más importante de la autoevaluación está en su utilización como herramienta para la motivación y la toma de conciencia, pues ayuda a los alumnos a apreciar sus cualidades, a reconocer sus insuficiencias y a orientar su aprendizaje de una forma más eficaz. Un esfuerzo importante en esta dirección es el proyecto *Porfolio* del Consejo de Europa, sobre el que se puede obtener más información en la página del Consejo de Europa:

http://www.coe.int/t/dg4/portfolio/Default.asp?L=E&M=/main_pages/welcome.html (en inglés o francés), o bien en la del Ministerio de Educación español: http://www.mec.es/programaseuropeos/jsp/plantilla.jsp?id=se (última consulta, 3 de julio de 2007).

En la autoevaluación introspectiva suelen hacerse dos distinciones en los tipos de escala que se utilizan: *definidas vs. no definidas* y *generales vs. específicas*. Una escala *no definida* solo tiene descripciones en el principio y el final de la escala, mientras que una definida tiene descripciones para cada nivel. Una escala específica pide al estudiante que valore su actuación en una situación lingüística particular, mientras que una escala general se refiere al uso de la lengua solo en términos generales. He aquí algunos ejemplos:

Tipo 1: Escala general no definida

Invita al estudiante a valorarse en una escala de:
0 – No hablo nada de español
a
20 – Hablo español de forma fluida
No hay ni niveles intermedios de descripción ni ejemplos específicos.

Tipo 2: Escala específica no definida

Tampoco tiene descripciones intermedias de nivel, pero invitan al estudiante a considerar su probable actuación lingüística en una situación hipotética. Por ejemplo:

Imagina que quieres pedir que te cambien un par de zapatos en la tienda en que los compraste. ¿Hasta qué punto crees que podrías hacerlo en español?

0 – No podría en absoluto

10 – Lo haría sin dificultad

Tipo 3: Escala general definida

Este tipo de escala tiene descripciones específicas en cada nivel, pero estas se expresan en forma de habilidades lingüísticas generales más que con ejemplos específicos. Es el tipo de escala elegido para el *Portfolio*. Por ejemplo [3]:

Interacción oral

– Nivel A1: Puedo participar en una conversación de forma sencilla siempre que la otra persona esté dispuesta a repetir lo que ha dicho o decirlo con otras palabras y a una velocidad más lenta y me ayude a formular lo que intento decir. Planteo y contesto preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales.

– Nivel A2: Puedo comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo de información sobre actividades y asuntos cotidianos. Soy capaz de realizar intercambios sociales muy breves, aunque, por lo general, no puedo comprender lo suficiente como para mantener la conversación por mí mismo.

Tipo 4: Escalas específicas definidas

Estas suelen ir en forma de cuestionario con listas de funciones.

Por ejemplo:

Puedo:	*preguntar por el tiempo*	*sí/no*
	deletrear mi nombre en voz alta	*sí/no*
	invitar a una persona a cenar	*sí/no*
	aceptar una invitación	*sí/no*
	rechazar una invitación	*sí/no*

El número de «sí» da una puntuación general. Los objetivos del programa de enseñanza suelen utilizarse para diseñar las preguntas del cuestionario.

2.2. *Evaluación del profesor*

En vez de basarse en una prueba de 10 minutos, la evaluación del profesor se basa en 30 o incluso 100 horas de exposición a la lengua del estudiante en

[3] Tomado del *Portfolio en castellano* del Ministerio de Educación y Ciencia español: http://www.mec.es/programas-europeos/docs/guia_pel_primaria.pdf (última consulta, 5 de julio de 2007).

situaciones y actividades variadas. Es decir, en términos generales de evaluación, se basa en una mayor y mejor cantidad de muestras de lengua; en términos humanos, el lenguaje ha sido producido en el aula, donde el estudiante está más relajado que en una prueba oral. No debemos olvidar que a muchas personas un examen les crea una ansiedad extrema que puede afectar muy negativamente en su actuación. Un buen entrevistador anima a su estudiante a abrirse y a demostrar su dominio de la lengua pero hay un límite para lo que puede hacer en 10 minutos; con mucho más tiempo a su disposición, un profesor con éxito llega a conocer a sus estudiantes como personas. Es ese «conocimiento» lo que la evaluación del profesor intenta atrapar.

La evaluación del profesor puede llevarse a cabo teniendo en cuenta todo el curso o bien centrarse en un periodo específico de, por ejemplo, una semana o dos. Durante ese periodo el profesor intenta asegurarse de que cada estudiante ha tenido la misma oportunidad de hablar y les va evaluando con una escala de puntuación (que veremos más adelante).

En una evaluación continua[4] como esta, el juicio del profesor se forma gradualmente y no es una decisión repentina. Cada vez que el estudiante realiza una tarea en clase, el docente ha administrado en realidad una parte más de la prueba. El total de estas a lo largo del curso constituye una prueba completa que, al contrario de las pruebas orales breves, no estará influida por variaciones individuales como el nerviosismo, la enfermedad o la fatiga de un día en particular. Un diario detallado de la actuación en el curso por parte del profesor constituye una excelente prueba oral.

Las desventajas de la evaluación del profesor conciernen directamente a la fiabilidad:

- Cuanta más gente participe en el programa de evaluación, más difícil es confiar en que los resultados sean comparables.
- Puede que sea difícil organizar entrenamiento permanente para los profesores implicados. Un buen profesor no es automáticamente un buen evaluador.
- Al puntuar a los alumnos el profesor puede tender a comparar entre ellos, incluso cuando utiliza una escala de puntuación.[5]
- Los profesores construyen relaciones con los alumnos y esto les puede influir a la hora de puntuar. Del mismo modo, la relación del alumno con el profesor le puede llevar a ser más o menos participativo en un curso.

[4] «La evaluación continua es una evaluación que realiza el profesor, y puede que también el alumno, respecto a las actuaciones en clase, los trabajos y los proyectos realizados a lo largo del curso. La calificación final, por tanto, refleja todo el curso o semestre» (MCER, 2002: 185).

[5] Apuntamos este aspecto como desventaja ya que nos inclinamos claramente por la *evaluación con referencia a un criterio* en contraposición a la *evaluación con referencia a la norma.* Mientras esta última valora o posiciona a los alumnos con relación a sus compañeros (hablamos del «tercero de la clase» o «está entre los diez primeros de su promoción») o en relación al grupo de alumnos que realizan la prueba, en la evaluación con referencia a un criterio es evaluado meramente en función a su capacidad, sin tener en cuenta la capacidad de otras personas. Las escalas de nivel del MCER ofrecen una descripción de un contínuum a partir del cual se puede clasificar con referencia a un criterio. Un ejemplo de este tipo de evaluación son los Diplomas de Español como Lengua Extranjera (www.dele.cervantes.es).

Estos problemas no son insuperables y, teniendo en cuenta los beneficios de la evaluación del profesor, tanto personales como organizativos, vale la pena invertir en un buen entrenamiento de puntuación en el programa.

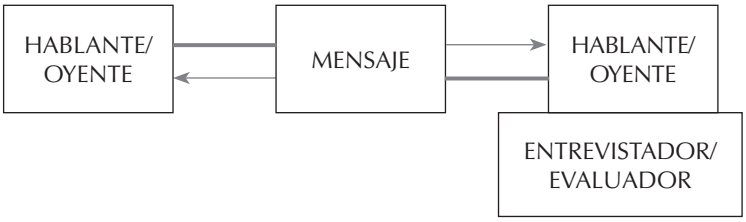
2.3. *Pruebas con examinador*

Si una prueba oral se define como aquella en la que el estudiante tiene que hablar, ¿con quién lo hace?

Veamos estas cuatro posibilidades:

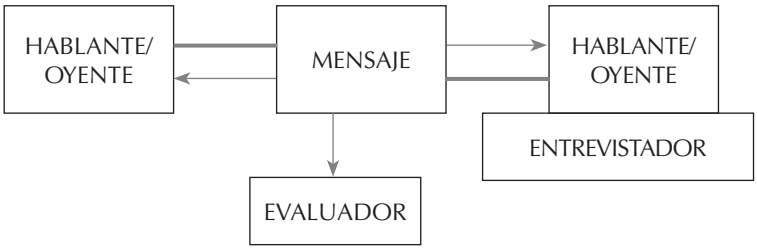
a) Puede hablar con un entrevistador que es también el evaluador.
b) Puede hablar con un entrevistador que no está involucrado en la puntuación.
c) Puede hablar con otro estudiante.
d) Puede hablar con un grupo de estudiantes.

a) Estudiante – entrevistador/evaluador



Una misma persona asume el papel de entrevistador y evaluador. Este es el modelo más común y económico, pero es difícil que una persona pueda concentrarse para evaluar de forma efectiva al mismo tiempo que intenta parecer interesado en lo que el estudiante está diciendo e implicarse seriamente en la comunicación con él. Este doble papel es muy agotador y requiere de descansos frecuentes.

b) Estudiante – entrevistador

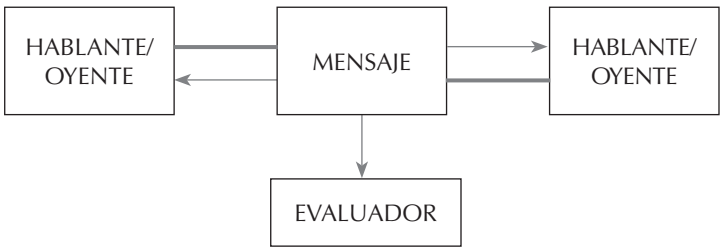


Aquí, la persona con la que habla el estudiante es diferente a la persona que le califica. El entrevistador puede concentrarse en la conversación, prestarle toda la atención al estudiante, animarle y ayudarle a hablar lo más fluidamente posi-

ble. Ciertamente, hablar con una persona sabiendo que otra le está observando puede resultar más difícil para algunos estudiantes. Por ello, es aquí fundamental la actuación del entrevistador y su capacidad de darle la impresión de estar interesado en lo que le cuenta, de ser amable, de conseguir que el alumno piense que está de su parte, y de aparecer como independiente del evaluador y bien entrenado.

En estos dos modelos el entrevistador controla el desarrollo de la prueba *y* puede por ello obtener el tipo *y* cantidad de muestra de lengua necesaria para llegar a una puntuación.

c) Estudiante – estudiante

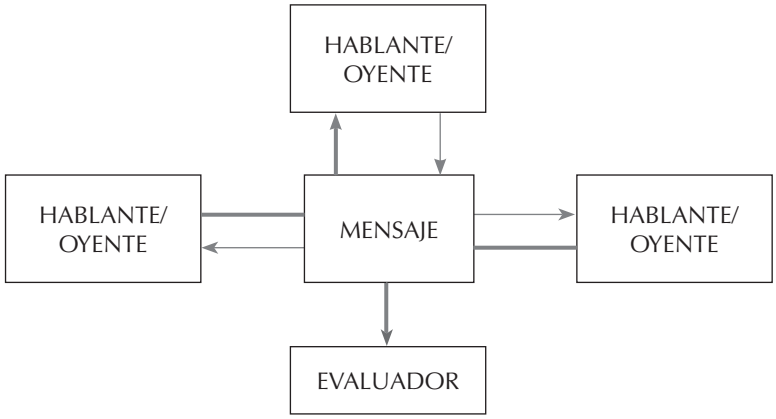


Dos estudiantes hablan entre ellos y llevan a cabo una tarea mientras el evaluador escucha sin intervenir. El evaluador puede prestar total atención a lo que se está diciendo ya que no debe preocuparse de mantener viva la conversación, por lo que al mismo tiempo se evita a un entrevistador que parezca agotado y desinteresado por estar preguntando lo mismo toda la mañana. Asimismo, el estudiante puede relajarse al estar hablando con alguien de su nivel en vez de con un profesor y/o nativo.

Al formar las parejas para este modelo es importante organizar pares equilibrados (fuertes con fuertes, tímidos con tímidos) tanto por su nivel de dominio lingüístico como por su personalidad. En principio los estudiantes suelen cooperar más que competir. De todos modos, si el evaluador tiene la impresión de que alguien no ha tenido una oportunidad justa para hablar, puede pedirle que vuelva a hacer la prueba con otra pareja.

En este modelo, y debido a que el profesor no interviene para controlar el proceso de la prueba, es fundamental que las instrucciones para los candidatos se hayan dado con antelación y que nos aseguremos de que las han entendido. De hecho es conveniente ensayar las técnicas que se utilizarán en clase (por ejemplo, juego de roles, descripción de historietas, etc.). Otro factor a tener en cuenta es el tiempo. Hay que decirles claramente de cuánto tiempo disponen (las pruebas en pareja o en grupo suelen alargarse si no se impone un límite) y de la importancia de que intervengan ambos.

d) Estudiante-grupo



Un grupo de tres a seis personas llevan a cabo una tarea conjunta. El examinador permanece en silencio y, preferiblemente, no se involucra. Al igual que en parejas, los estudiantes tienden a estar más desinhibidos y a hablar más espontáneamente al estar trabajando en parejas o grupos con otros estudiantes.

Para el evaluador es más difícil mantener el rastro de cinco o seis personas que de solo dos y, además, la interacción puede ser más larga (entre veinte o treinta minutos) para que este tenga bastante tiempo (y muestras) para puntuar a cada uno. Un consejo que suele darse es que el evaluador puntúe rápidamente a los estudiantes con una personalidad más prominente y se concentre en los miembros menos habladores después.

Existen también otros modelos de pruebas de los que no vamos a hablar aquí por razones de espacio como son las pruebas al teléfono o las pruebas grabadas. En estas últimas no está presente ni un entrevistador ni un examinado, sino que suelen hacerse en un laboratorio de lenguas donde varios estudiantes responden al mismo tiempo a una serie de estímulos grabados anteriormente que escuchan a través de auriculares. Sus respuestas se graban y se recogen para ser puntuadas.

Sin embargo, sí me gustaría mencionar el uso de la grabación de pruebas con objetivos muy diferentes. Me refiero a la posibilidad de grabar las pruebas realizadas en uno de los cuatro modelos mencionados anteriormente. En estos casos, la grabación se utiliza para uno o más de los siguientes objetivos:

— para programas de entrenamiento de examinadores (entrevistadores y evaluadores),
— para moderar la consistencia de la prueba,
— para la autoevaluación.

Las ventajas de grabar una prueba son:

a) La posibilidad de que una misma persona actúe como entrevistadora (durante la prueba) y como evaluadora (durante el visionado de la grabación) sin tener que hacerlo al mismo tiempo.
b) La puntuación no debe hacerse en tiempo real (es decir, en el mismo lugar y hora que la prueba).
c) Aumenta la fiabilidad si se puntúa la actuación con dos evaluadores, por ejemplo.
d) Es muy útil cuando se tienen problemas prácticos para reunir a estudiantes y examinadores en un lugar y una hora determinados.

Pero también tiene desventajas:

a) El examinador, si la grabación no es en vídeo, puede oírlo todo, pero no verlo, con lo que pierde algo tan fundamental como los aspectos de la comunicación no verbal.
b) En una prueba «en vivo» el examinador puede alargar o dirigir la prueba si tiene dificultades para evaluar algo; asimismo, puede preguntar si no ha entendido algo.
c) Pueden darse dificultades técnicas (mala calidad o la no grabación).
d) No hay contacto humano.

3. **Técnicas**

A continuación pasamos revista a once de entre las muchas técnicas utilizadas en pruebas orales de lengua, ordenadas —de manera muy general— en una escala de menor a mayor predictibilidad en la respuesta. Cada una de esas técnicas es más o menos adecuada para cada tipo de prueba: la gran mayoría de las pruebas utilizan dos o más técnicas con un equilibrio entre más y menos controladas.

3.1. *Discusión-conversación*

Esto es lo más natural del mundo: dos personas teniendo una conversación sobre un tema de interés común. Es también la más difícil de conseguir con éxito en el marco de una evaluación ya que se da solamente cuando ambas partes se sienten relajadas y aparece una fluidez en el intercambio que permite que la actividad (la conversación) sea dominante y su propósito oculto (la evaluación) quede temporalmente subordinado. La prueba oral alcanza así su mayor grado de autenticidad al dejar de ser una prueba.

En esta técnica es decisiva la capacidad del entrevistador para crear una buena atmósfera en tan poco tiempo, aunque también tiene mucho que ver su personalidad y la de los participantes. En general, solo alumnos con un alto dominio de la lengua se sienten lo suficientemente seguros como para tomar la iniciativa, hacer preguntas o expresar desacuerdo, funciones que requieren el dominio de habilidades muy específicas. También es una cuestión de personalidad; atreverse a tomar la

iniciativa en una situación en la que te están evaluando significa atreverse a tomar riesgos y el estudiante solo lo hará si sabe que se le considerará positivamente en vez de penalizársele. El instinto natural más bien nos lleva a mantenernos callados en un examen y a responder cuando se nos pregunta.

Una desventaja puede ser que las personalidades más extrovertidas y habladoras tengan ventaja ante los tímidos. De ahí que el entrevistador deba ser capaz de manejar tales situaciones y de ser sensible a la personalidad del estudiante. Una simple regla que se recomienda al entrevistador suele ser intentar no hablar demasiado y estar preparado a dejar silencios en los que el estudiante pueda pensar en lo que va a decir o incluso pueda reunir coraje para decir algo. Tampoco habría que olvidar que aspectos como los falsos comienzos, las declaraciones ambiguas o los farfulleos son características de la conversación ordinaria y no deben penalizarse o esperar un habla «de libro».

Más adelante veremos que esta técnica suele enlazarse a otras técnicas, por ejemplo, conversar sobre un tema expuesto en una exposición o una historieta.

3.2. *Exposición*

Esta técnica consiste en que el estudiante haga una presentación preparada anteriormente que puede durar entre cinco y diez minutos. Puede consultar notas, pero no leer la exposición. Al final de la presentación se espera que sea capaz de contestar a las preguntas que se le hagan. Esta es una técnica auténtica y comunicativa tanto para propósitos académicos como profesionales.

La elección del tema es aquí muy importante. Debería ser relevante para los objetivos del programa o las necesidades del alumno, y no debería ser demasiado especializado. Los temas pueden elegirse entre el alumno y el profesor. Muchos alumnos tienden a elegir un tema que les resulta muy familiar.

Al puntuar, el evaluador debe tener muy en cuenta si el tema entraña una dificultad por sí mismo, independientemente de la fluidez del hablante, y en las baremaciones suelen incluirse las distintas habilidades funcionales implicadas: exposición de datos, expresión de opiniones, exposición de argumentos, manejo de las preguntas que se le hacen, capacidad de resumen, etc. Así también pueden identificarse los puntos débiles para trabajarlos posteriormente.

Propuesta 1: Se le dan al estudiante unos cuantos temas para que elija uno (pueden darse también algunas preguntas que le ayuden a recoger ideas). Se le dan de 10 a 15 minutos para prepararse antes de la prueba y después hace una exposición de unos 3-5 minutos en la que podrá consultar sus notas pero no leer un guión. He aquí un ejemplo de algunos temas para un nivel B2 [6]:

[6] Tomado de: *Preparación al Diploma Intermedio Español Lengua Extranjera. Nivel B2*. P. Alzugaray, M. J. Barrios y C. Hernández. Madrid: Edelsa, 2006

Tema 1: ¿Comer en casa o en el restaurante?	Tema 2: Viajes, gustos y preferencias	Tema 3: Ecología
– Ventajas y desventajas. ¿Cuál de las dos opciones prefiere usted? ¿Por qué? – La sobremesa. – La comida el fin de semana y la comida los días laborables. – Horarios. – ¿Es más sana la comida elaborada en casa que la del restaurante? – ¿Qué tipo de platos suele pedir usted en un restaurante? – ¿Se comía mejor antes que ahora? – Razone su respuesta. – ¿Cómo son los restaurantes en su país?	– ¿Cómo y con quién prefiere viajar? ¿Qué tipo de alojamiento prefiere? – ¿En qué medio de transporte suele viajar? – ¿Qué opina de los vuelos baratos? – ¿Qué tipo de destino suele elegir: playa, montaña, ciudades monumentales...? – ¿Le gustan los viajes organizados o prefiere viajar por cuenta propia? – ¿En qué época del año suele viajar? – Consejos para un buen viaje.	– ¿Le preocupa la extinción de especies animales? – ¿Qué piensa de la edificación en o cerca de parques y reservas naturales? – ¿Qué piensa de los experimentos con animales para cosméticos, medicinas, etc.? – ¿Cuál es su opinión sobre los abrigos y los artículos de piel? – ¿Piensa que podríamos consumir menos energía? – ¿Qué piensa de los tipos de energía que hay actualmente? – ¿Qué piensa de los tipos de energía alternativa? – ¿Cómo podemos contribuir al ahorro energético?

Propuesta 2: Se invita al estudiante a elegir un tema junto con su profesor. Se le da un día, una semana o incluso un mes para prepararlo. Lo expone en 5-10 minutos (ante la clase, por ejemplo).

3.3. *Discusión y toma de decisiones*

Esta técnica se utiliza mucho en los modelos de evaluación conjunta en un grupo de dos o más estudiantes, sin la participación de un entrevistador. Los estudiantes tienen que mantener y dirigir la discusión totalmente por sí mismos. La tarea suele implicar documentación escrita que se lee antes de la prueba, en esta hay que discutir y llegar a un consenso en el grupo. La discusión es lo que se evalúa y no la decisión final que se tome; por lo tanto, no hay ni una respuesta correcta ni «un ganador». Esto es algo que los estudiantes deben saber con antelación, y también que se les puntuará de acuerdo al modo en que se expresen, justifiquen sus opiniones y evalúen las de los demás.

El material que se utilice puede ser auténtico (por ejemplo, folletos) o bien escrito expresamente para la prueba. En cualquier caso deberían ser sencillos y relativamente breves, para que la prueba no se convierta en una prueba de comprensión lectora. También deben ofrecerse suficientes posibilidades distintas para que pueda haber discusión y tener cuidado en que una de las elecciones sea la mejor de forma muy obvia. Sería conveniente practicar esta técnica antes en cla-

se, no solo para prepararles para la prueba sino para acostumbrarles a la idea del trabajo centrado en el alumno e independiente del profesor.

El evaluador debe tener en cuenta que hay estudiantes más habladores y otros más tímidos. Como no puede intervenir en la discusión se intentará animar con antelación a los tímidos a hablar, pero de todos modos deberá basar su puntuación en intervenciones de al menos 10 frases (o 100 palabras) por alumno.

Propuesta 1: Discusión sobre una identidad a partir de efectos personales.

Se le entrega a un grupo de estudiantes un paquete de «efectos personales» (por ejemplo, dos por cada persona del grupo) de los que suele llevar una persona en sus bolsillos o en el bolso: tarjetas de crédito, fotografías, cartas, recibos, agenda, billetes de autobús, etc. La tarea consiste en hacer conjeturas sobre el tipo de persona a la que pueden pertenecer esas cosas (edad, profesión, estatus familiar, intereses, tipo de vida, etc.) y llegar a un consenso. Se les puede pedir que cada uno describa dos objetos y haga conjeturas a partir de ellos para repartir el tiempo de habla. Esta técnica es especialmente buena para evaluar la formación de hipótesis, conjeturas y negociación. Los estudiantes deben entender que no hay una respuesta «correcta» y que no gana quien haga la propuesta que se acepte finalmente sino que se evalúa a partir de la calidad de sus intervenciones.

Propuesta 2: Discusión a partir de folletos turísticos.

Se reparte a cada estudiante (dos o más) un folleto de una ciudad distinta. Antes de la prueba habrán tenido tiempo para leerlo y pensar en las ventajas que ofrece el destino turístico que describe su folleto para defenderlo ante su pareja/grupo. El objetivo de la tarea consiste en defender su destino como el ideal para unas vacaciones conjuntas y llegar entre todos a un acuerdo. De nuevo lo que se puntuará será la calidad de las intervenciones *y* de sus argumentaciones *y* contraargumentaciones.

3.4. *Juego de roles*

En esta técnica se le pide al estudiante que asuma un papel específico y que se imagine en ese papel en una situación determinada. Tiene que conversar con su/s interlocutor/es de forma adecuada al papel y a la situación. Se le dan las instrucciones poco antes de la prueba, que deberán explicar exactamente lo que debe hacer. En esta técnica se trabaja la habilidad de hacer preguntas, que muchas veces queda descuidada en las pruebas orales.

Los temas y las situaciones pueden elegirse dependiendo de la muestra de lengua que se quiera obtener. Por ejemplo:

- Funciones: quejarse, dar direcciones.
- Estructuras: narrar un accidente (pasados), conseguir información (hacer preguntas).
- Vocabulario por temas: reservar un hotel o un vuelo, hacer un pedido.

— Habilidad para hacer circunloquios, es decir, para hacerse entender a pesar de ignorar vocabulario específico. Aquí la situación elegida debe ser deliberadamente poco familiar.

Esta técnica puede adaptarse al nivel de la prueba a través de la elección de las situaciones a representar, que pueden ser más simples o más complejas. Una situación simple puede ser el estereotipo de una situación cotidiana como pedir algo en un restaurante o reservar una habitación de hotel. He aquí el ejemplo que se propone para el juego de roles de la parte oral del DELE-Nivel inicial (B1)[7]:

> EN LA RECEPCIÓN DE UN HOTEL. USTED HA RESERVADO UNA HABITACIÓN Y SE LO DICE AL RECEPCIONISTA.

Ejemplo de interacción para este caso:

> EXAMINADOR: — *Buenas tardes, ¿qué desea?*
> CANDIDATO: — *Buenas tardes, tengo una reserva (a nombre de...)*
> EXAMINADOR: — *¿A nombre de?* (Si no lo ha dicho antes el candidato)
> CANDIDATO: (Da su nombre si no lo ha dicho antes)
> EXAMINADOR: — *Un momento, por favor. ¿Era una habitación doble o sencilla?*
> CANDIDATO: — *Doble / Sencilla*
> EXAMINADOR: — *Muy bien, aquí está. Habitación 501. ¿Trae usted equipaje?*
> CANDIDATO: — *Sí, tengo una maleta.*

Una situación compleja, en cambio, debe añadir aspectos inusuales, pero no improbables, de la vida cotidiana: una entrevista de trabajo, relatar un accidente a la policía, describir los síntomas de una enfermedad a un médico, etc. Se trata de situaciones que requieren cierto grado de persuasión o en las que el alumno tiene que ponerse en una situación en la que no tiene experiencia.

Un posible problema al utilizar esta técnica puede ser el rechazo a participar de algunos estudiantes. Hay que pensar con antelación si es una técnica que puede ser aceptada, si no habrá que decidirse por otra técnica.

Propuesta: Juego de roles entre estudiantes

Dos estudiantes forman pareja en un juego de roles. Se les dan las instrucciones de forma escrita unos minutos antes de la prueba: cada

[7] En www.dele.cervantes.es, Modelos de examen, Recomendaciones para preparar las pruebas orales. Nivel inicial: http://diplomas.cervantes.es/docs/ficheros/200511300001_7_3.pdf

uno recibe uno de los roles y la tarea que tienen que llevar a cabo. Para que la producción de lengua entre ambos esté equilibrada pueden darse instrucciones como: «Haz al menos tres preguntas» o si el entrevistador lo considera necesario, puede pedirle al final a cada uno de los estudiantes que haga un resumen, que diga qué pasaría después o puede pedirle más información.

3.5. *Descripción y reconstrucción*

Esta técnica se utiliza sobre todo en el modelo de prueba estudiante-estudiante. Uno de ellos debe describir una imagen o una figura a su interlocutor, que debe ser capaz de reconstruirla a partir de su descripción, sin ver el original. El estudiante no puede ver la reconstrucción que está intentando hacer su compañero, por lo que no sabe hasta qué punto está teniendo éxito con su descripción.

Se puntúa el lenguaje usado por su precisión y falta de ambigüedad. El grado de éxito de comunicación puede medirse bien según el tiempo necesitado para completar el modelo o bien por el grado de exactitud con el que este se consigue reconstruir en un tiempo determinado. Los estudiantes cambian los roles y repiten la tarea con modelos diferentes. Se evalúa la muestra de lengua del estudiante que describe. Los materiales pueden ser construcciones como los «Lego», o dibujos de estos, o descripción de mapas de carreteras.

En la versión más simple de esta técnica solo se obtiene lenguaje unidireccional: un alumno describe a quien reconstruye. En otra versión, el constructor puede hacer preguntas para comprobar o clarificar instrucciones; de este modo se obtienen muestras de preguntas descriptivas.

Esta es una buena técnica para generar y comprobar la descripción detallada de objetos físicos, requiere un buen dominio del léxico del color, dimensión, ubicación y secuencia en la descripción. Se puede utilizar desde niveles iniciales.

Propuesta: Leer un mapa

Un estudiante describe a otro una ruta predeterminada a lo largo de un mapa. El segundo estudiante tiene que seguir la ruta a lo largo de otro mapa, igual pero sin marcar, con un lápiz. Al final de la descripción tiene que señalar claramente el punto final en el mapa. Se pueden usar mapas reales o imaginarios, y se evalúa la habilidad de dar direcciones de forma clara y precisa. Para hacer la tarea más fácil, se dan más nombres de calles o puntos de referencia.

3.6. *Dar respuestas adecuadas*

Se le da al estudiante un número determinado de situaciones cortas, cotidianas y sin relación entre ellas. Las situaciones se le pueden dar escritas en un papel, se le pueden leer, o ambos. El estudiante debe imaginarse en la situación y dar lo que considere que sería la reacción verbal correcta. Las situaciones deben estar des-

critas de modo que sean fáciles de entender y la respuesta requerirá una o dos frases como mucho. Están diseñadas para obtener lenguaje funcional, por ejemplo, preguntar por información, disculparse o rechazar educadamente una invitación. En esta técnica pueden observarse también posibles malentendidos o carencia de dominio de formas culturales.

Esta técnica necesita un examinador con un dominio de la lengua y de la cultura hispana igual a la de un nativo, para poder valorar la adecuación de la respuesta rápidamente. Podría también prepararse una lista de posibles respuestas adecuadas, pero es realmente difícil confeccionar una clave que las abarque todas. Se puede puntuar en una escala como la siguiente:

0— Inapropiado o seriamente incorrecto.
1— Pertinente pero no del todo aceptable.
2— Apropiado y correcto.

La técnica es rápida y fácil de utilizar: varias situaciones y sus respuestas requieren solo un par de minutos para administrar y puntuar. Sin embargo, se dan algunos problemas:

— Hay que tener mucho cuidado en crear situaciones que no sean ambiguas y en las que solo un tipo de respuesta sea la apropiada.
— Puede favorecerse a estudiantes con una procedencia cultural similar a la de los hispanohablantes.
— El español es una lengua internacional y hay que tener muy en cuenta que las respuestas pueden responder a hábitos culturales de una zona determinada del mapa hispanohablante.

Algunos ejemplos serían:

— *Usted está en el despacho de un cliente en el que hace mucho calor, ¿qué le diría?*
— *Usted entra en un ascensor en el que hay una persona fumando, lo cual le molesta muchísimo, ¿qué le diría?*

Propuesta: Dar respuestas convencionales

Se le pide al estudiante que dé respuestas adecuadas con fórmulas convencionales. Por ejemplo:

– *Siento haberle hecho esperar (No tiene la menor importancia)*
– *¿Qué tal está? (Muy bien, gracias. ¿Y usted?)*
– ¿Qué tal el fin de semana? (Estupendo, gracias. Fuimos a esquiar)
– ¿Le importa que abra la ventana? (No, no, claro que no)
– Muchas gracias (De nada)

3.7. *Preguntas y respuestas*

Esta técnica suele consistir en una serie de preguntas sin conexión entre sí, graduadas de menor a mayor dificultad; es decir, de preguntas como *¿Cómo te llamas?* o *¿Dónde vives?* hasta llegar a frases complejas como *Si no hubieras venido hoy a hacer este examen, ¿qué habrías hecho?* Esta técnica es adecuada para las pruebas de clasificación o para evaluar niveles iniciales. Es una técnica poco auténtica.

Los listados suelen ser del siguiente tipo (y pueden alargarse hasta el infinito):

— *¿Cómo te llamas? ¿Cómo se escribe?*
— *¿Cómo te encuentras?*
— *¿Qué tal es tu nivel de español?*
— *¿Te gusta hablar en español?*
— *¿A qué te dedicas?*
— *Háblame un poco de tu familia.*
— *¿Puedes contar a partir de 20?*
— *¿Sabes qué hora es?*
— *¿A qué fecha estamos? ¿Sabes qué día de la semana es hoy?*
— *¿Qué tiempo hace?*
— *¿Dónde has aprendido español?*
— *Cuéntame tres cosas que hiciste ayer.*
— *¿Qué estabas haciendo ayer a estas horas?*
— *¿Dónde trabajas/estudias?*
— *¿Te gusta trabajar/estudiar allí?*
— *¿Qué haces en tu tiempo libre?*
— *¿Qué harás hoy cuando salgas de aquí?*
— *¿Qué piensas hacer en tus próximas vacaciones?*
— *¿Cuáles son tus planes para el futuro?*

En la confección del listado puede sernos útil recordar que las preguntas de *quién/qué/cuándo/dónde* son más sencillas que las preguntas de *cómo/por qué, y* que el grado de dificultad de la respuesta depende también del tipo de pregunta. Este sería un orden de menor a mayor grado de dificultad:

— preguntas de respuestas sí/no
— preguntas fácticas (sobre hechos sencillos)
— preguntas descriptivas
— preguntas narrativas
— preguntas especulativas
— preguntas hipotéticas
— preguntas de opinión

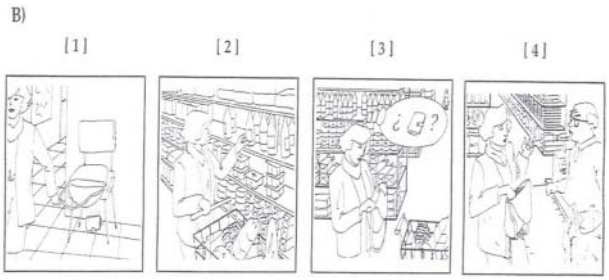
Variante: Pedir que haga la pregunta.

En todos los niveles es importante asegurarse de que los alumnos saben también preguntar y no solo responder. El entrevistador puede proponer que las haga con

modelos como el siguiente: *Ahora me gustaría que me hicieras algunas preguntas. Pregúntame mi nombre, por mi profesión, mi familia, mis intereses, etc.*

3.8. *Imágenes e historietas*

Las historietas consisten en varias viñetas (de 4 a 12) que cuentan una historia simple pero que permite al alumno dar su propia interpretación sobre las personas y los acontecimientos que aparecen. En esta técnica se le da al alumno una imagen o una historieta (bien en la misma prueba, o bien con antelación con tiempo para prepararse en niveles inferiores). Entonces el entrevistador le pide que la describa y, si se trata de una historieta, que además narre la acción. Cuando termina la descripción (y narración), el entrevistador tiene la oportunidad de hacerle preguntas destinadas a obtener la información que le interesa: por ejemplo, la comprobación del dominio del léxico. Este es un ejemplo [8]:



Las historietas están dibujadas o se eligen específicamente para este objetivo, excluyendo objetos o acciones que son difíciles de describir o muy marcados culturalmente. Las imágenes o historietas pueden explotarse a dos niveles:

a) Descripción de personas, objetos y situaciones que llevan a una interpretación.
b) Discusión sobre un tema subyacente o más general.

Así pues, y como apuntábamos ya más arriba, a esta técnica se le puede enganchar la técnica de discusión-conversación (3.1). Su ventaja es que los estímulos visuales son una vía económica y efectiva de proveer un tema de conversación sin darle al estudiante las palabras o frases que necesitará, por lo que no requiere de comprensión escrita ni auditiva, pero el problema es la dificultad de encontrar imágenes e historietas adecuadas.

Propuesta 1: Imágenes similares.

Se necesitan dos copias de 3-4 imágenes similares; por ejemplo, paisajes o cuadros. El estudiante elegirá una de las imágenes y se la describirá a otro estudiante o al entrevistador, que deberá decidir cuál de las imágenes similares está describiendo. Se evalúa la fluidez

[8] Fuente: www.dele.cervantes.es, Modelos de examen, Recomendaciones para preparar las pruebas orales. Nivel intermedio: http://diplomas.cervantes.es/docs/ficheros/200511300001_7_4.pdf

y riqueza de la lengua producida y no su capacidad discriminatoria visual.

Propuesta 2: Ordenar imágenes para crear una historia.

Se le da al estudiante una historieta con las viñetas recortadas. El estudiante decide en qué orden deberían ir y va contando la historia en el orden que ha elegido. Otro estudiante o el entrevistador va escuchando y va ordenando la historia con otro juego de viñetas desordenadas. El objetivo es darle más alcance a su imaginación y creatividad al permitirles asumir el control sobre los conocimientos que describen.

Propuesta 3: Narrar una historia en grupo.

Para una evaluación en grupo. Se recorta una historieta y se le dan 2-3 viñetas a cada estudiante del grupo. Las viñetas estarán numeradas y se trata de que vayan contando la historia describiendo lo que hay y pasa en su viñeta cuando les toca el turno. Los candidatos no deben ver las viñetas de los demás, así que deberán estar atentos a lo que digan para seguir con el hilo de la historia. Es importante practicar esta técnica en clase con anterioridad.

Propuesta 4: Nombrar objetos.

Se trata de un simple test de vocabulario.

Propuesta 5: Generar tiempos diferentes.

Se añade a la historieta un reloj o una agenda que hagan que el alumno deba contar la historia en un tiempo verbal determinado.

3.9. *Dar instrucciones, descripciones, explicaciones*

El estudiante describe con cierta longitud un objeto conocido, un sistema o un procedimiento cotidiano. La descripción es fáctica (ni abstracta ni teórica) y se consigue que el estudiante produzca un discurso relacionado sobre un tema determinado al mismo tiempo que le permite cierta libertad de expresión sin necesitar mucho tiempo de preparación (aspecto que lo diferencia de la *exposición)*.

Se le da al estudiante una lista entre algunos temas entre los que puede elegir, y unos minutos de preparación. Algunos ejemplos:

— *¿Cómo preparas una taza de buen café o de buen té?*
— *Describe cómo se prepara tu plato favorito.*
— *Da instrucciones para enviar un correo electrónico con adjunto.*

Lo importante es que los temas elegidos requieran que el estudiante produzca al menos entre seis u ocho frases. Si empieza a hablar de otras cosas hay que advertirle de que solo se darán puntos por lo relevante al tema.

3.10. *Resumir o reconstruir una historia a partir de un estímulo*

a. A partir de un estímulo auditivo

El estudiante escucha un pasaje corto o una historia en una grabadora o en un vídeo, después se le pide que lo resuma y reconstruya. Las instrucciones deben enfatizar que es la calidad y no la cantidad de lo reconstruido lo importante, y que debe utilizar sus propias palabras más que intentar recordar palabras o frases del pasaje. Esto debería verse reflejado en el sistema de puntuación que debería valorar la capacidad de parafrasear y de reconstruir los puntos principales de la historia o del pasaje.

Aunque las frases son breves —de seis a diez frases— son demasiadas para poder ser memorizadas completamente, por lo que la técnica apuesta por la capacidad-comprensión-procesamiento-producción. Al ser esta una mezcla de destrezas puede decidirse no utilizarla para evaluar la expresión oral, aunque muy pocas veces va esta desunida de la comprensión.

También puede utilizarse un pasaje en el que no se habla sino que solo se oyen sonidos que sugieren una historia [9]. El uso de grabaciones tiene también la ventaja de disminuir la carga en los modelos de examen en los que el entrevistador es también el evaluador.

Variante: Discutir el contexto.

Después de escuchar el estímulo, se invita al candidato a especular sobre el contexto: ¿Quién habla con quién, dónde, cuándo? ¿Se usa un lenguaje formal o informal? Aquí es más apropiado usar varias audiciones cortas que una larga [10].

b. A partir de un estímulo escrito

El estudiante lee un pasaje o series de pasajes cortos para sí mismo. Después se le pide que cuente lo leído con sus propias palabras. No hay un tiempo fijo para la parte de la lectura, pero no puede consultar el texto escrito a partir del momento en que empiece a contar la historia.

También puede reducirse el texto escrito ofreciéndole gráficos y pidiéndole que los interprete. Esta técnica se puede hacer en todos los niveles.

La diferencia con la técnica de estímulo auditivo es que el alumno tiene tiempo para releer y determinar el tiempo de lectura. En la audición se hace en tiempo real y por tanto escapa a su control. El entrevistador puede preguntar si hay alguna palabra que no se entienda.

Variante: Tomar notas.

---

[9] Hacemos una propuesta de material para esta técnica en Carmen Pastor, 2005: 13.

[10] Ofrecemos un ejemplo de este tipo de actividad en Carmen Pastor, 2005: 14.

Especialmente útil para fines académicos. Los estudiantes hacen una pequeña exposición o descripción a partir de unas tarjetas preparadas previamente. Todos los estudiantes reciben los mismos apuntes para asegurar que se les evalúa por su nivel de lengua y no por su conocimiento especial del tema. Se pueden hacer reconstrucciones o comparaciones de teorías diferentes o bien explicaciones de datos.

4. **El sistema de puntuación**

El sistema de puntuación es una parte vital de la prueba oral. Como apuntábamos más arriba, las pruebas que requieren una evaluación subjetiva —es decir, el juicio de un evaluador— son menos fiables, lo cual hace que sus resultados puedan ser de menos confianza. Para evitar esto se podría optar por pruebas objetivas, pero eso en destrezas como la expresión oral suele convertirlas en poco válidas. Para conseguir una puntuación fiable, hay una serie de cuestiones a tener en cuenta.

4.1. *Sobre evaluadores y entrevistadores*

a. Dos evaluadores producirán una puntuación más fiable que uno, y tres más que dos, pero atención: la regla a más evaluadores más fiabilidad no funciona. El problema está en que cuantos más haya, mayor dificultad tendremos en asegurar la fiabilidad interevaluadora. La solución ideal es disponer de un pequeño número de evaluadores que hayan sido entrenados cuidadosamente y poder contar con dos para cada prueba[11]. Una tercera opinión, en caso de disconformidad entre ellos, también subirá el grado de fiabilidad (a partir, por ejemplo, de una grabación).

b. No todos los buenos profesores son buenos evaluadores (ni buenos entrevistadores como apuntábamos en otro apartado). Algunas personas tienden a dar puntuaciones más consistentes (fiabilidad intraevaluadora) que otras, tanto respecto a sus propias puntuaciones como a las de otros compañeros. El entrenamiento para llegar a dicha consistencia suele hacerse a través de sesiones en las que un grupo de evaluadores en entrenamiento puntúan pruebas grabadas y discuten entre ellos los resultados, hasta ir llegando a una «representación ideal» en su cabeza sobre qué actuación de un alumno corresponde a qué puntuación.

c. Otra posibilidad, cuando solo se dispone de un evaluador, es grabar la prueba y dejar que este mismo u otro evaluador la puntúe después. Pero es importante tener en cuenta que cuando se puntúa desde una grabación se suelen dar puntuaciones más bajas que en vivo (Underhill, 1987: 94). Esto podría deberse a que:

— El evaluador no está implicado en la situación real y pierde mucha comunicación no verbal.

[11] En Alderson *et altri* (1995) se detallan los pasos a seguir para la formación *y* la supervisión de la fiabilidad de los examinadores.

— Pueden darse problemas para entender enunciados aunque se trate de un buen equipo.
— Tiene más tiempo para escuchar los errores del estudiante, incluso de volver atrás para comprobar dichos errores.
— Los entrevistadores en vivo siguen su propio estilo para recoger la información que les llevará a hacerse un juicio de valor. Los evaluadores de grabaciones dependen de la técnica de otra persona.

d. Hacer protocolos de evaluación también ayuda: la gente sabe con antelación qué pasos hay que ir dando, lo que ahorra tiempo y desconcierto.

4.2. *Baremos de puntuación*

Para puntuar una prueba de evaluación subjetiva como es la de la expresión oral, se suelen distinguir dos tipos de baremos que, aunque utilizan los mismos criterios, se organizan de forma diferente:

a. Baremos holísticos

Algunas pruebas tienen como objetivo evaluar el dominio general de la lengua. En estos casos se utilizan los llamados baremos sintéticos u holísticos, que establecen grados o niveles de calidad a partir de descripciones globales. Normalmente este tipo de baremos dispone solo de *escalas* cortas de 4-6 grados. Para utilizarlos, el evaluador lee el escrito una o más veces y elige, entre las opciones, la que se adapta a la representación que se ha formado. Este sería un ejemplo [12]:

Muestra de una escala holística		
18-20	Excelente	Inglés natural con errores mínimos y realización completa de la tarea encomendada.
16-17	Muy bien	Más que un conjunto de frases simples, con buen vocabulario y estructuras. Algunos errores básicos.
12-15	Bien	Realización simple aunque precisa de la tarea encomendada, con algunos errores serios.
8- 11	Suficiente	Razonablemente correcta aunque torpe y no comunicativa O tratamiento natural y justo del tema, con algunos errores.
5-7	Flojo	Vocabulario y gramática no adecuados al tema.
0-4	Muy flojo	Incoherente. Los errores muestran la ausencia de conocimientos básicos de inglés.

b. Baremos analíticos

Otras pruebas, en cambio, intentan evaluar habilidades específicas. En este caso se utilizan baremos analíticos, que aíslan categorías lingüísticas para puntuarlas por

[12] Se toma aquí el ejemplo (en su versión traducida) que Alderson, Clapham y Wall (1998: 108) tomaron de *UCLES International Examinations in English as a Foreign Language General Handbook,* 1987. Aunque hubiésemos preferido incluir aquí los baremos utilizados en la prueba oral de los DELE, estos no se encuentran aún publicados.

separado. La puntuación por categorías se combina muy a menudo con escalas para cada categoría. El evaluador se basa entonces en una descripción de la actuación «ideal» en cada categoría a cada nivel de la escala. En el DELE-Nivel intermedio, por ejemplo, se utiliza una baremación analítica para la calificación de la prueba oral en la que se incluyen las siguientes categorías: pronunciacón *y* acento, léxico, gramática, adecuación al registro *y* la situación, interacción y fluidez. En cada categoría existe una escala de 1 a 4 entre la que elige el evaluador[13].

Las categorías que se deciden incluir dependen de lo que se ha decidido trabajar en el currículo de enseñaza. Algunas de las habilidades que más comúnmente se evalúan son gramática, léxico, pronunciación, entonación, acentuación, estilo *y* fluidez, adecuación *y* contenidos. Pero sean las que sean las que se elijan, es importante recordar que un evaluador no puede seguir la huella de tres o cuatro categorías al mismo tiempo, por lo que es mejor no incluir demasiadas (suelen haber de 6 a 8) y diseñar la prueba con diferentes técnicas en las que puedan ir observándose algunas categorías por separado.

También es posible combinar formulaciones analíticas y sintéticas en un mismo baremo. Por ejemplo, es posible evaluar el éxito de un objetivo comunicativo con el sistema sintético y analizar la pronunciación, gramática, léxico y adecuación con un baremo analítico.

4.3. *Escalas de puntuación*

Una escala de puntuación se compone de una serie de descripciones breves de diferentes niveles de una habilidad lingüística[14]. Su objetivo es describir brevemente lo que un alumno «ideal» puede hacer en cada nivel, de forma que sea fácil para el evaluador decidir cuál es la puntuación que puede darle al estudiante. Es decir, que la escala le ofrece al examinador una serie de descripciones y él elige la más adecuada para describir la actuación del alumno.

Es importante no usar más niveles de los necesarios. Cuantos menos, más fácil será evaluar y más alta será la fiabilidad. Una escala podrá funcionar bien solo si el examinador puede estar escuchando y/o hablando con el estudiante al mismo tiempo que retiene en su memoria dichas escalas. También se conseguirá mejorar una escala a través de la experiencia *y* del error; conservando las partes que funcionan *y* tirando o remodelando las que no.

Fundamental es también basar la puntuación sobre el modelo de lengua que muestre el estudiante: su constante, y no sobre momentos de brillantez o sobre un serio error que aparece de forma aislada. Así como asegurarnos en todo momento que estamos evaluando lo que queríamos, y no otra cosa; es decir, que si queremos

[13] Como comentábamos en la nota 12, no nos es posible incluir los criterios de calificación de la prueba oral del DELE por no estar aún publicados. En Underhill (1987: 98-99) se da un ejemplo en inglés.

[14] Un ejemplo de este tipo de escalas son las ya muy conocidas del MCER o las de DIALANG (www.dialang.com) o ALTE (www.alte.org) que se encuentran también en MCER, 2002: 222-233 y 242-247 respectivamente.

evaluar expresión oral no tendrá mucho sentido quitar puntos por una respuesta en la que el alumno hace una interpretación de una imagen que no esperábamos.

4.4. *Ponderación*

Si en un sistema de puntuación damos los mismos puntos a todas las categorías, estaremos dando por supuesto que todas tienen la misma importancia. Sin embargo, en la mayoría de programas de evaluación oral se da una importancia diferenciada a las categorías evaluadas. Cuando damos más importancia o peso a algunas de ellas, las ponderamos.

Un sistema para ponderar que no crea confusión o agotamiento en los evaluadores (es mentalmente más fácil puntuar) es aquel en el que se le da la misma cantidad de puntos a cada categoría pero en la que después se multiplica por cifras diferentes, dependiendo de la mayor o menor influencia en el resultado final que se le quiera dar.

5. **A modo de conclusión**

Resumiendo, al diseñar una prueba oral parece evidente la necesidad de empezar tomando buena nota de cuáles son nuestros objetivos y cuáles las necesidades de nuestros estudiantes. A continuación, deberemos ser conscientes de nuestros recursos (de tiempo, personal, equipos, salas) y restricciones antes de decidirnos por un modelo de examen. Una vez tomadas estas decisiones podremos ocuparnos de las técnicas más adecuadas y buscar los materiales en concordancia con los contenidos que deseemos evaluar. Asimismo, habrá que decidirse por un sistema de puntuación y entrenarlo con los evaluadores, además de formar a los entrevistadores en las técnicas de evaluación elegidos.

Estupendo sería si pudiésemos pilotar nuestra prueba antes de ponerla en funcionamiento, aún mejor sería ir corrigiendo nuestro programa de evaluación oral continuamente, pero lo mejor sería si conseguimos diseñar una prueba en la que nuestro candidato llegue a olvidar que está en un examen y se implique espontáneamente en una conversación, en la que pone en escena todo su dominio del español.

Bibliografía

— Alderson, J. Ch., C. Clapham y D. Wall (1995): *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press (Versión publicada en español: Exámenes de idiomas. Madrid: CUP, 1998).
— Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 2001.
— Bachman, L. F. y A. S. Palmer (1996): *Language Testing in Practice*. Oxford: OUP, 2002.
— Bordón, T. (2000): «La evaluación de la comprensión auditiva en aprendices de ELE». *Carabela,* n.º 49, pp. 77-101, Madrid: SGEL.
— Cassany, D. (1993): *Reparar la escritura*. Barcelona: Graó, 2000.

— Consejo de Europa (2000): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (MCER). Madrid: Anaya (También disponible en http://cvc.cervantes.es/obref/marco).
— Pastor, C. (2005): «La Cenicienta de las destrezas: la comprensión oral» en *Der fremdsprachliche Unterricht Spanisch,* n.º 8. Friedrich Verlage.
— Hugues, A. (1989): *Testing for Language Teachers*. Cambridge: CUP, 2003.
— Underhill, N. (1987): *Testing Spoken Language*. Cambridge: CUP, 2003.