

El presente artículo corresponde a la conferencia presentada en la XVIII edición de Expolingua (Madrid, 7, 8, 9 y 10 de abril de 2005). Este trabajo, inédito, ha sido revisado por su autora para esta edición en marcoELE.

Mi objetivo es recoger, a partir de una revisión de bibliografía sobre la evaluación de la comprensión oral, un marco teórico y práctico desde el que desarrollar o revisar tareas para evaluar la comprensión oral en una prueba de aprovechamiento. Es decir, una prueba que nos indique cuánto ha aprendido un alumno al final de un curso de español. Una prueba cuyo objetivo sea “calificar y obtener información acerca del proceso de enseñanza-aprendizaje”. (Bordón, 2000: 84)

Nos centraremos en las tareas de comprensión que Gary Buck (2001: 167) llama comprensión oral *sin colaboración* (*non-collaborative listening*), es decir, la audición en situaciones en que el oyente no ayuda en la construcción del discurso, y no trataré aquí la comprensión oral llamada *con colaboración* o interactiva que suele ser evaluada junto con la expresión oral.

El propósito principal de ese marco de actuación, que recojo en 33 puntos, es que me sirva de guía hacia el desarrollo de una prueba con un nivel aceptable de validez, fiabilidad y viabilidad. Pero antes de seguir adelante debemos plantearnos dos preguntas muy importantes, ¿por qué evaluar la comprensión oral? y ¿por qué evaluarla separada de la expresión oral?

En general, la evaluación es un elemento presente en el proceso de enseñanza de una lengua extranjera por razones que pueden ser externas al propio aprendizaje (por ejemplo, obtener una descripción del propio nivel de competencia para presentarse a un puesto de trabajo), o puede ser parte integrante del proceso de enseñanza-aprendizaje (por ejemplo, para ver si el alumno está aprendiendo lo que debería, dónde está teniendo problemas, qué contenidos se deberían desarrollar en el curso, etc.).

En concreto, la evaluación de la comprensión oral se justifica porque forma parte de la competencia comunicativa (MCER, 2002: 60) y por “su potencial efecto positivo en las prácticas de clase” (Alderson y Bachman en Buck, 2001: xi). Por lo que respecta a la evaluación de la comprensión oral separada de la interacción oral, dos argumentos juegan un papel importante: por una parte, hay situaciones en que escuchamos sin participar, como al escuchar la radio, conferencias, anuncios en una estación de trenes, al oír casualmente una conversación, etc.; por otra, en muchas situaciones de enseñanza resulta inviable efectuar pruebas orales en los exámenes de aprovechamiento por razones prácticas y logísticas.

Una vez justificado el esfuerzo, vamos a ver cuáles son los pasos que, según los expertos en evaluación, debo tener en cuenta para elaborar tareas evaluadoras de la comprensión oral.

1. DEFINIR EL CONSTRUCTO

Antes de intentar medir un proceso tan complejo como la comprensión oral necesitamos entender cómo funciona. “La comprensión de aquello que estamos intentando medir es el punto de partida para la construcción de un test. Llamamos constructo a aquello que intentamos medir y nuestro test será útil y válido solo si mide el constructo correcto.” (Buck, 2001: 1) Lo que nosotros intentamos medir es la comprensión oral, así que vamos a definirla, ¿cómo funciona y qué la caracteriza?

La comprensión oral es el resultado de la interacción entre un número de fuentes de información que incluyen el *input* acústico, el conocimiento lingüístico, el conocimiento del mundo, el contexto de comunicación y el uso de los oyentes de cualquier información a su disposición que les parezca relevante para ayudarles a interpretar lo que el hablante está diciendo, es decir, para ir construyendo el significado del texto.

Se ha debatido mucho sobre cómo se aplican estos conocimientos o cómo actúa la mente sobre la onda sonora que escuchamos y cómo llega a comprender su significado. Los dos puntos de vista más importantes son el modelo de abajo-arriba (*bottom-up*) y el modelo de arriba-abajo (*top-down*). El modelo abajo-arriba asume que la mente procesa en un orden definitivo y de una sola dirección: primero identifica los sonidos, después las palabras que constituyen esos sonidos, después la frase compuesta por esas palabras, después hay un análisis del contexto semántico para llegar a la comprensión literal del significado lingüístico básico, y por fin el oyente interpreta ese significado literal en términos de la situación comunicativa y añade el significado pragmático que le da como resultado lo que el hablante quiere decir. (Buck, 2001: 2)

Pero el procesamiento de diversos tipos de conocimiento no ocurre en una secuencia fija, sino que más bien puede ocurrir de forma simultánea o en cualquier otro orden necesario. Así, el conocimiento sintáctico puede ayudar a identificar una palabra o el conocimiento del contexto puede ayudar a interpretar el significado. Podemos entender el significado de una palabra antes de decodificar su sonido porque tenemos muchos tipos de conocimiento - incluido el del mundo que nos rodea- que nos crea expectativas acerca de lo que vamos a oír. Este es el llamado modelo de arriba-abajo que “aborda el enunciado en su totalidad, de forma global, y descomponiendo en unidades cada vez más pequeñas.” (Poch 2000: 12)

Para Dolores Poch (ibídem: 14) “un modelo que dé cuenta de la comprensión auditiva debe ser global y considerar que existe una interrelación íntima y constante entre los diversos niveles lingüísticos que componen el enunciado.” Y

para Buck (2001: 3) la comprensión auditiva también es un proceso de arriba-abajo en el sentido que los varios tipos de conocimiento implicados en la comprensión del lenguaje no se aplican en un orden determinado sino que pueden interactuar e influirse mutuamente.

Otro aspecto sumamente importante para definir la comprensión oral es el input del oyente, es decir, el lenguaje hablado. Este tiene tres características particularmente importantes para el constructo de la comprensión oral: (a) el habla está codificada en forma de sonido; (b) es lineal y se produce en un tiempo real, sin oportunidad para la repetición; (c) es lingüísticamente diferente al lenguaje escrito. Vamos a verlas con más detenimiento.

a) El primer contacto con el enunciado llega en forma de onda sonora, pero el oído humano no percibe por completo sus características físicas de intensidad, frecuencia y duración de los sonidos. Esto no suele ser ningún problema en la lengua materna, pero no ocurre lo mismo ante enunciados de una lengua extranjera. Un ejemplo es el hecho de que “los hablantes necesitan que los enunciados emitidos en lengua extranjera presenten mayor intensidad que los emitidos en su lengua materna. En ese sentido es fácil, por ejemplo, darse cuenta de que cuando estamos viendo un programa de televisión en una lengua que no es la nuestra, necesitamos siempre aumentar el volumen del aparato para poder seguir lo que se está diciendo con la misma comodidad con la que seguiríamos en nuestra lengua materna a un volumen más bajo.” (Poch 2000: 7)

Por otra parte, “los hablantes, en función de la situación de comunicación, modifican sus realizaciones de tal manera que articulan con menos precisión cuando la situación es informal y suprimen algunos sonidos en terminaciones o en grupos consonánticos, llegando, en ocasiones, a eliminar sílabas enteras. [...] Un ejemplo de ello lo constituye la pronunciación de palabras tan frecuentes en español como *entonces* o *también*, que, en la mayoría de los casos, son realizadas como *toces* o *mbién* por no mencionar casos de supresión de consonantes en expresiones como *es cierto*, que suele realizarse como *ecierto*, o como *los lunes*, que suele realizarse *lolunes*.” (ibídem: 10) En estos casos, los oyentes nativos restituyen todo lo que no está presente físicamente en la onda sonora -pero que sí lo está en la intención de realización del hablante- gracias al contexto global en que se desarrolla la comunicación.

Asimismo, el oyente identifica lo lingüísticamente pertinente de lo que no lo es. Poch (ibídem: 8), citando a Sorin, señala que el sistema auditivo humano es capaz de realizar tres funciones fundamentales encaminadas a conseguir la comprensión de los mensajes:

- Distinguir el habla “útil” del resto del entorno sonoro. Es decir, comprende incluso con ruido o música de fondo, mientras hablan otras personas.
- Codificar adecuadamente las diferencias fonéticas individuales, es decir, las características acústicas diferentes de hablantes distintos.

- Garantizar que dicha codificación es resistente frente a la “degradación” de la onda sonora. Esa degradación se da por ejemplo al hablar por teléfono ya que la banda de frecuencias que transmite el aparato se sitúa entre 300 Hz y 3.400 Hz mientras que los sonidos del habla abarcan un espectro que va desde los 50 Hz hasta los 10.000 Hz por lo menos.

“Estas funciones, que el oyente realiza de manera prácticamente automática cuando se trata de una lengua materna, parecen plantear muchas más dificultades cuando deben realizarse en lengua extranjera.” (ibídem: 8)

b) El habla tiene lugar en tiempo real, oímos el texto una sola vez. Por supuesto podemos pedirle a quien habla que repita lo que ha dicho, pero normalmente no obtenemos una repetición palabra por palabra, sino una reformulación de lo dicho. El hablante suele darse cuenta en estas ocasiones de que hay un problema, por lo que ayuda a quien escucha reformulando o dando ejemplos. Incluso si nos repiten con las mismas palabras, la entonación cambia al repetir la información. En el uso normal del lenguaje tenemos solo una oportunidad para comprender lo oído y eso tiene dos consecuencias: la primera es que el oyente debe procesar el texto a la velocidad determinada por el hablante, que suele ser bastante rápida; la segunda es que los oyentes no pueden volver a oír el texto (como en la escritura), todo lo que le queda está en la memoria, una memoria imperfecta de lo que se ha oído.

Para poder entender a los hablantes a la velocidad de unas tres palabras por segundo (que suele ser algo muy normal), el procesamiento del oyente debe ser casi totalmente automático. Ángel López (2002: 18) apunta que la gran diferencia entre la comprensión oral en la L1 y en la LE estriba en la conciencia del proceso que se da desde que las ondas sonoras llegan al oído hasta que el cerebro receptor comprende, es decir, asimila lo que acaba de oír. El nativo tiene una comprensión suficiente y un proceso inconsciente; el aprendiz de LE, en cambio, una comprensión insuficiente y un proceso consciente. Precisamente la enseñanza de la comprensión oral en la LE consiste en lograr que ese proceso vaya siendo cada vez menos consciente, más automático, para aproximar la comprensión del oyente de LE a la del nativo.

De hecho, los oyentes con un procesamiento menos automático o más consciente, comprenderán menos. Lynch (1998) apunta que “cuando la velocidad crezca, no tendrán tiempo suficiente para procesarlo todo, así que empezarán a prestarle más atención al procesamiento de la información gramatical y léxica, y menos atención a la interpretación, mucho más amplia, del significado. Cuando la velocidad se hace aún más rápida, el oyente no tendrá suficiente tiempo para procesar ni siquiera la información gramatical y léxica y empezarán a perder parte del texto. A una cierta velocidad, su procesamiento tenderá a romperse completamente, y dejarán de comprender.

c) El lenguaje hablado es lingüísticamente diferente del escrito. El oral es un discurso con indecisiones, pausas, falsos comienzos, repeticiones e incluso errores gramaticales y de elección de palabras, no está planificado y se

caracteriza por unidades de ideas cortas, ensartadas de forma relajada, con un vocabulario y unas estructuras discursivas característicos.

Daniel Cassany (1987: 31-34) hace una relación detallada de las diferencias tanto contextuales como textuales entre el código oral y el escrito:

Diferencias contextuales:

- canal auditivo vs. el visual;
- recepción sucesiva vs. la simultánea;
- comunicación inmediata vs. la diferida;
- comunicación efímera vs. la duradera;
- uso de muchos códigos no verbales vs. poco uso de estos (disposición del espacio, textura del soporte, etc.);
- dependencia del contexto vs. la independencia del contexto;
- interacción durante la emisión del texto vs. la no interacción durante la composición del texto

Diferencias textuales:

- adecuación: el texto oral, por ejemplo tiende a marcar la procedencia dialectal mientras que en el texto escrito se neutralizan las marcas de procedencia y se usa una lengua más estándar;
- coherencia: somos más redundantes, introducimos más digresiones y datos irrelevantes en la comunicación oral que en la escrita;
- cohesión: oralmente utilizamos pausas y la entonación, elementos paralingüísticos, códigos no verbales y referencias exofóricas al contexto mientras que por escrito usamos signos de puntuación y pocos del resto de los elementos nombrados;
- fonología y grafía: la comunicación oral incorpora elisiones, contracciones, elisiones... que el lenguaje escrito no usa;
- morfología: la oralidad tiende a soluciones más formales que la escritura;
- sintaxis: en el discurso oral tendemos a usar estructuras más simples, anacolutos y frases sin terminar, elipsis y un orden de la frase muy variables, mientras que en la escritura aparecen estructuras más complejas y desarrolladas, una estructura de la frase más estable y no suelen darse de anacolutos o elipsis;
- léxico: la oralidad tiende a un léxico menos formal que el escrito, usa menos vocablos, acepta la repetición léxica en contraste con la escritura o incorpora tics lingüísticos.

Por último es importante y nos será de utilidad recordar también las dos funciones básicas de la lengua que describen Brown y Yule (1983): la función interactiva, cuyo objetivo principal es mantener las relaciones sociales y cuyo contenido no es tan importante; y la función transaccional, cuyo objetivo principal es dar información. La mayoría de situaciones en las que se escucha consisten en una mezcla de ambas.

2. ¿CUÁL ES EL PROPÓSITO EXACTO DE LA PRUEBA?

En nuestro caso será evaluar el nivel de comprensión oral en audiciones sin colaboración, es decir, audiciones en las que el oyente no ayuda en la construcción del discurso. De todos modos, en estas audiciones se incluirán textos interactivos imitando la realidad, en la que se dan situaciones en las que oímos conversaciones en las que no tomamos parte, como cuando oímos conversaciones de forma casual o cuando estamos presentes en debates pero no participamos en el grupo de discusión.

3. ¿QUÉ TIPO DE PRUEBA NECESITO?

Aquí debo preguntarme para qué hacemos la prueba exactamente. El objetivo general de la evaluación es conseguir información, pero necesitamos definir que tipo de información queremos y según esto podremos elegir la prueba más adecuada (Hugues, 1989: 8):

INFORMACIÓN DESEADA	PRUEBA ADECUADA
¿Cuál es el nivel general de habilidad lingüística de quien aprende?	Prueba de competencia (p. ej. DELE: www.cervantes.es)
¿Cuáles son los puntos débiles y fuertes de cada estudiante?	Prueba de diagnóstico (p. ej. DIALANG: www.dialang.org)
¿Cuál es el lugar más adecuado para un estudiante dentro de nuestro programa?	Prueba de clasificación (p. ej. AVE: http://ave.cervantes.es)
¿Cuánto ha aprendido un estudiante en un curso?	Prueba de aprovechamiento

4. ¿QUÉ TIPO DE EVALUACIÓN UTILIZARÉ?

Hemos visto los tipos de pruebas que podemos diseñar pero es también importante que hablemos de los enfoques de evaluación desde las que las construiremos.

Evaluación formativa vs. evaluación sumativa

Utilizamos una evaluación *formativa* cuando lo que pretendemos es controlar u obtener información sobre el progreso de nuestro alumnado a fin de modificar nuestros planes de enseñanza. Su idea principal es la de mejorar el aprendizaje. La evaluación *sumativa*, en cambio, mide lo que se ha conseguido al final de un curso con una calificación.

Evaluación directa vs. evaluación indirecta

Las pruebas pueden evaluar de forma directa o indirecta. La evaluación *directa* requiere del candidato que actúe en la habilidad que se desea medir. Si queremos evaluar cómo escribe, le pondremos a escribir, o si queremos saber

cómo es su pronunciación, lo pondremos a hablar. Para evaluar la comprensión auditiva y lectora, en cambio, solemos pedirle a los estudiantes que hagan otras cosas, además de escuchar y leer, para demostrarnos que han entendido lo oído o leído: necesitamos una evaluación *indirecta*. Por ejemplo, les pedimos que hagan evidente que han comprendido un texto marcando casillas, completando frases, respondiendo preguntas, etc. Una prueba indirecta clásica es un ejercicio de selección múltiple o de relleno de huecos, técnicas que veremos más adelante. La ventaja de la evaluación indirecta es que sus resultados son más generalizables ya que la muestra que se puede obtener es mucho mayor. La desventaja es que la relación entre la actuación y las habilidades que queremos evaluar tiende a ser débil e incierta.

Evaluación de puntos concretos vs. evaluación integrativa

La evaluación de *puntos concretos* o específicos se centra en el examen de elementos aislados y descontextualizados del sistema de la lengua, por ejemplo en ortografía: v/b, en morfosintaxis: fue/iba, en léxico: ir/venir, etc. En las pruebas de la evaluación integrativas se observa el uso de muchos elementos a la vez en una situación de uso: leer un texto, escribir una nota, etc.

Evaluación subjetiva vs. evaluación objetiva

Aquí la diferencia radica solo en los métodos de puntuación. La evaluación es *objetiva* si no se requiere el juicio de un evaluador, como en un ejercicio de selección múltiple, en el que una respuesta correcta equivale, por ejemplo, a un punto. La evaluación *subjetiva*, en cambio, es una valoración que realiza un examinador.

En la evaluación directa (una entrevista oral, escribir una carta) generalmente se otorgan calificaciones según un juicio o valor. 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 se puede ver la actuación de la lengua en conjunto, de forma compleja. Las indirectas y objetivas se basan en la atomización de la lengua a la que dividen en partes. El problema es que lengua y comunicación son más que la simple suma de sus partes.

No obstante, y para ser justos, toda evaluación debería ser lo más objetiva posible. Las consecuencias de las valoraciones personales que conllevan a decisiones subjetivas respecto a la selección de contenido y calidad de la actuación se deberían reducir tanto como fuera posible, sobre todo cuando se trata de una evaluación sumativa. La subjetividad de la evaluación se puede reducir para aumentar así su validez y fiabilidad siguiendo las siguientes pautas:

- Especificar claramente el contenido de la evaluación.
- Crear unos criterios y unas puntuaciones que se discutan entre todos los examinadores hasta ponerse de acuerdo sobre qué significan exactamente.

- Adoptar procedimientos normalizados respecto a la forma de realizar las evaluaciones. Cuanto más conocido y experimentado sea un procedimiento, mayor será el nivel de consenso.
- Exigir varias valoraciones siempre que sea posible.
- Comprobar la calidad de la evaluación (validez, fiabilidad) analizando los datos de evaluación.

Evaluación con referencia a la norma vs. evaluación con referencia a un criterio

La evaluación con *referencia a la norma* clasifica, valora o posiciona a los alumnos con relación a sus compañeros. Es decir, que se habla por ejemplo, de "eres el número 3" en relación a la clase; o "están entre el 2% de los más ricos del país", en relación a la demografía; o "está entre el 14% de los mejores de su promoción" en relación al grupo de alumnos que realizan la prueba, hacen una carrera, etc. Un uso habitual de la evaluación referida a la norma ocurre en las pruebas de clasificación para formar los grupos de los distintos niveles.

Como una reacción contra la referencia a la norma está la evaluación con *referencia a un criterio*, de este modo el alumno es evaluado meramente en función de su capacidad, sin tener en cuenta la capacidad de otras personas. Las escalas de nivel del *Marco Común de Referencia* ofrecen una descripción de un contínuum a partir del cual se puede clasificar con referencia a un criterio. Un ejemplos de este tipo de evaluación son los DELE (Diplomas de Español como Lengua Extranjera).

Evaluación realizada por otras personas vs. autoevaluación

La evaluación *realizada por otras personas* es aquella cuyas valoraciones son realizadas por el profesor o el examinador. En la *autoevaluación*, en cambio, las valoraciones respecto a la competencia lingüística las hace la misma persona que se evalúa.

El potencial más importante de la autoevaluación está en su utilización como herramienta para la motivación y para 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 *Portfolio*, del Consejo de Europa. Si te interesa saber más sobre él, puedes obtener información en:

www.eeooiinet.com/portfolio/triptico_portfolio.pdf

Aún teniendo en cuenta la importancia didáctica que la autoevaluación tiene, no parece ser una medida muy fiable por una serie de factores que influyen en los juicios de la gente sobre su propia actuación. Estos factores pueden ser conscientes y deliberados, como una supervaloración para intentar impresionar y conseguir un trabajo mejor; o una infravaloración para entrar en un grupo de lengua cuyo nivel resulte cómodo y sin retos. Inconscientemente pueden influir factores como diversos niveles de autoseguridad y percepción. Si un estudiante tiene un historial de logros académicos poco espectaculares, seguramente infravalore sus habilidades. Generalmente las mujeres estiman sus habilidades menos que los hombres y las personas mayores menos que los jóvenes. Cuanto

más homogéneos sean los examinados -en términos de edad, sexo, nacionalidad, y contexto educacional, por ejemplo- mayor será la similitud de sus juicios, claro que esto no significa que éstos serán más adecuados: simplemente compartirán los mismo prejuicios. (Hugues 1989)

5. LA PRUEBA DEBE SER VÁLIDA

La validez¹ es la propiedad más importante de un test. La idea es muy básica, ¿evalúa la prueba lo que se supone que debería evaluar? Se distinguen diversos tipos: *validez de contenido*, se da si la prueba evalúa exactamente los objetivos que había previsto; *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; *validez de predicción*, se da si la prueba tiene capacidad de predicción de las conductas que demostrarán los examinados.

La comprensión oral no puede examinarse de forma directa, así que siempre tendremos que darle a los examinados alguna tarea y medir su comprensión oral a partir de su actuación en dicha tarea. Al haber siempre otras destrezas implicadas, además de la comprensión oral, tendremos que tener cuidado en no acabar midiendo la capacidad de comprensión lectora, de expresión escrita o de memoria de un alumno en vez de su capacidad de comprender textos orales.

6. LA PRUEBA DEBE SER FIABLE

Decimos que una prueba es fiable² si los resultados que ofrece tienen estabilidad, es decir, si evaluando siempre a la misma persona en la misma situación obtuviéramos los mismos resultados. Distinguimos diversos tipos:

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

¹ Para una descripción detallada: Bachman 1990: 236-294, Bachman y Palmer 1996: 21-23, Alderson et al., 1995: 165-189

² Para una descripción detallada de fiabilidad: Bachman 1990: 160-233, Bachman y Palmer 1996: 19-20; Alderson et al., 1995: 87-89 y 127-144 para la fiabilidad de los examinadores.

7. LA PRUEBA DEBE TENER AUTENTICIDAD SITUACIONAL

La autenticidad exige que las tareas de una prueba guarden relación con actividades -en este caso, de audición- 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. Bachman y Palmer (1996: 23-25) designan este concepto como autenticidad situacional.

También es importante que los textos orales sean auténticos, es decir que los textos deberían tener las características del lenguaje hablado, pero "aunque es importante que los textos sean realistas, no todos necesitan ser grabados de la vida real. Un texto desarrollado a partir de un guión puede carecer de algunas características importantes del habla, pero aún así puede medir importantes destrezas de la comprensión oral." (Buck 2001: 165)

8. LA PRUEBA DEBE TENER AUTENTICIDAD INTERACCIONAL

Bachman y Palmer (1996: 25-26) definen interaccional (*interactiveness*) como el tipo de interacción entre la tarea de evaluación y el examinado, lo cual "supone tomar en consideración si las operaciones que se les exigen a los candidatos se corresponden con los procesos que ocurrirán de forma natural al interactuar con estos textos" (Bordón 2000: 81). Es decir, que una tarea será interaccional, o tendrá autenticidad interaccional si requiere del alumno el conocimiento, destrezas y habilidades que necesitaría una persona para actuar en una situación de uso de la lengua.

9. LA PRUEBA DEBE INTENTAR PROPORCIONAR UN IMPACTO POSITIVO³

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 "una 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 *washback positivo*, que sea beneficioso y no lo contrario.

³ "Autenticidad, interaccionalidad e impacto son tres cualidades que muchos especialistas en evaluación consideran como parte de la validez" (Bachman y Palmer 1996: 41)

Otro tipo de impacto a evitar es un tipo de efecto no deseado (*bias*) que puede aparecer cuando una prueba favorece a unos examinados pero desfavorece a otros. Esto puede pasar cuando un acento es más conocido para un grupo que para otro (Bachman y Palmer 1996: 29-35). En el caso de los acentos, Buck recomienda que si los examinados no están acostumbrados a un mismo tipo de acento, se utilice "un abanico de acentos lo más amplio posible. Así, si los oyentes tienen problemas con un hablante, igual entienden mejor a otro." (Buck 2001: 162)

10. LA PRUEBA DEBE SER VIABLE

La prueba debe ser práctica y posible en la situación en la que se administra (Bachman y Palmer 1996: 35-38). Hay que preguntarse: ¿con qué recursos cuento?, ¿de cuántas personas, espacio y equipos técnicos?, ¿qué restricciones tengo a causa de los medios e instalaciones?, ¿cuánto tiempo hay disponible (tanto para la administración de la prueba, como para su corrección y puntuación)?, ¿quién evalúa?, ¿cuál es la estructura de los cursos?

11. ¿QUÉ TIPO DE ESTUDIANTE SE PRESENTARÁ A LA PRUEBA?

Hay que tener en cuenta las características de las personas que harán la prueba (edad, sexo, lengua materna, estudios, etc.) cuyo conocimiento nos ayudará, por ejemplo, en la selección de temas de interés para las tareas.

12. ESPECIFICAR LO QUE EL EXAMINADO DEBERÍA SER CAPAZ DE HACER

Para ello parto de un eje que propone Buck (2001: 104) en el que se cruzan (a) las destrezas y subdestrezas de las que deberían ser capaces los examinados en un ámbito y nivel determinado⁴, y (b) la tareas que se desarrollan en situaciones de uso de la lengua en las que es necesario interpretar textos orales. Para ello dispongo del *Marco común europeo de referencia* (2002: 68-70) y de la descripción de destrezas y subdestrezas de la comprensión oral que resume Hugues (1989: 161-162) de las que tendré en cuenta no sólo las subdestrezas del lenguaje transaccional, sino también el del lenguaje interaccional puesto que aunque las actividades estén diseñadas para una audición *sin colaboración*, ya hemos apuntado que los textos orales sí deberían presentar muestras de estas dos funciones de la interacción oral.

Sin embargo, en las pruebas de aprovechamiento tenemos la ventaja de que también tenemos el apoyo de los índices del manual del curso, en el que se suele dar una relación del tipo de tareas de comprensión oral en las que se ha ido ejercitando al alumno y del tipo de textos que se han utilizado.

⁴ Una descripción de los ámbitos la puedo encontrar en el *Marco común europeo de referencia* (2002: 49-56)

13. DEFINIR EL TIPO DE TEXTOS ORALES

Es importante definir qué tipos de textos utilizaremos tanto por razones de *washback* como de validez de contenido. Según el nivel y las actividades - especificadas en el punto 15- que el alumno debe ser capaz de hacer, señalaremos qué tipos de textos se incluirán. Tendremos en cuenta:

- el *tipo de texto*: monólogos, diálogos (con dos o más participantes), anuncios, conferencias, etc.;
- la *estructura del texto*: descripción, exposición, argumentación, instrucciones, narración;
- la *longitud*: medida en segundos o minutos, y la *velocidad del habla* que suele medirse por palabra por minuto. (Hugues 1989: 162)

Buck (2001: 123) señala que los textos largos tenderán a utilizarse para requerir habilidades discursivas, pero que pueden cansar y deberían ser interesantes para ayudar a mantener la concentración. Si además son difíciles, deberían tener una estructura clara para que el examinado tenga algún punto de orientación cuando se pierda, además debería tener redundancia y reformulación de los puntos más importantes. De todos modos, cuando se trabaja con textos difíciles es aconsejable que sean cortos.

Por lo que respecta a la velocidad, hay muchas evidencias que indican que los textos con un habla más lenta son más fáciles; si queremos evaluar el proceso automático, necesitaremos textos rápidos. Las características de los textos pueden cambiar de una situación a otra, pero deben utilizar lenguaje oral realístico y con sonido de calidad. Además, cuando los textos incluyen discurso interactivo es muy importante asegurarse que se pueden distinguir fácilmente a los hablantes. En vídeo esto es fácil de evitar, pero en audio muchas voces pueden sonar similares y crear confusión en el oyente. Para evitar esto se suelen elegir una voz de un hombre y de una mujer, o si son del mismo sexo se suelen utilizar diversos acentos o voces que se distinguen fácilmente. (Buck 2001: 166)

14. CONSEGUIR LOS TEXTOS

Podemos decidir elaborar los textos orales que serán la base de las tareas de evaluación, pero esto es difícil y muy complejo⁵. Otra alternativa es tomarlos del manual que hemos utilizado en el curso, aunque esto tiene la desventaja de que el alumno puede conocerlos con antelación y con ello perderíamos un elemento fundamental de la comprensión oral: lo *efímero* del lenguaje hablado (Cassany, 1987: 31). La posibilidad más factible es tomarlos de otros manuales o materiales similares a los trabajados en el curso. Por otra parte Isabel Bermejo y Marta Higuera (2000: 141-146) presentan en su artículo "La comprensión

⁵ Para quienes decidan crear sus propios textos orales para la evaluación, Buck (2001: 154-193) ofrece una descripción detallada de cómo hacerlo.

auditiva en el aula de español como lengua extranjera: Recursos en Internet para la elaboración de actividades” muchas fuentes interesantes y de fácil acceso.

15. LOS TEXTOS EN VÍDEO

Estos son una réplica más realista de la comprensión oral y además proporcionan al examinado un punto de atención mientras está escuchando. En estos textos es importante asegurarse de que la información visual complementa a la auditiva y que no entra en conflicto con ella, y de que la realización con éxito de la tarea dependa del procesamiento del *input* auditivo y no pueda deducirse solo de la imagen. El vídeo tiene también desventajas: ver el vídeo, leer el texto, escuchar y escribir las respuestas todo al mismo tiempo puede ser complejo, por lo que muchos examinados dejan de mirar la imagen, lo cual inutiliza el propósito inicial (Alderson et al. 1995: 215).

Buck (ibídem) opina que la información visual tiene más importancia en el lenguaje interaccional, en el que el énfasis está en la relación entre los participantes, pero es menos importante en el lenguaje transaccional en el que el énfasis está en el contenido y en la mayoría de los casos las pruebas de evaluación están más interesadas en el lenguaje transaccional.

16. ASEGURAR LA DEPENDENCIA DEL PASAJE

Es fundamental que el éxito de la realización de la tarea dependa de la comprensión del texto oral y de nada más, a esta noción se la llama *dependencia del pasaje* (Buck 2001: 126). Esta dependencia puede fallar por dos razones, la primera porque la respuesta correcta se pueda extraer por algún tipo de clave, por ejemplo, cuando en una selección múltiple la respuesta correcta es diferente a las demás (más larga); la segunda porque se pueda responder utilizando el conocimiento general del mundo.

17. COMPROBAR LA INFORMACIÓN NECESARIA

Una forma útil de resaltar lo que una tarea determinada está evaluando es identificar la *información necesaria* en el texto para poder hacer la tarea correctamente. Una vez identificada esta información podemos asumir que el conocimiento, destreza o habilidades que se requieren en la comprensión serán importantes para la realización de la tarea. Normalmente eso será lo que la tarea estará evaluando. (Buck 2001: 130)

18. LAS INSTRUCCIONES DE LAS TAREAS

Se recomienda darlas en la lengua materna de los examinados si es posible, y si no, asegurarse de que el lenguaje de las instrucciones es más fácil que el nivel

que la prueba intenta medir. Hay que usar siempre ejemplos claros y, de ser posible, preparar a los alumnos con ítems similares antes de que hagan la prueba (Buck 2001: 119) o usar técnicas con las que los alumnos ya estén familiarizados (Hugues 1989).

19. LAS PREGUNTAS DE COMPRENSIÓN DEL TEXTO

Probablemente la tarea más común que se utiliza en la evaluación de la comprensión oral sea presentar un texto oral y hacer una serie de preguntas al oyente para medir hasta qué punto ha entendido el contenido del texto. Shohamy e Inbar (1991) distinguen tres tipos de preguntas dependiendo del tipo de información al que van dirigidas:

- *preguntas globales*, que requieren del examinado sintetizar información o extraer conclusiones;
- *preguntas locales*, que requieren del examinado detalles concretos o entender palabras aisladas; y
- *preguntas triviales*, que requieren del examinado que entienda detalles concretos pero irrelevantes y sin relación con el tema principal.

Shohamy e Inbar encontraron que las preguntas globales son más difíciles que las concretas, y que las preguntas triviales se respondían de una manera poco predecible y no eran una herramienta significativa para la evaluación.

20. LAS PREGUNTAS Y LA INFORMACIÓN EN EL TEXTO

Las preguntas de comprensión deberían requerir la información en el mismo orden en que van apareciendo las respuestas en el texto, y una vez utilizada una información en una pregunta, no debería volver a usarse en otra. Además, y en especial en textos largos, es importante espaciar las preguntas para tener tiempo de ir respondiéndolas y asegurarnos de que un estudiante no estará respondiendo a una pregunta al mismo tiempo que necesita escuchar la información necesaria para responder a la siguiente. (Hugues 1989: 164)

21. PREGUNTAS QUE REQUIEREN INFERENCIAS

Para Buck la inferencia está en el corazón del procesamiento del lenguaje, y por lo tanto, es importante evaluar esta habilidad: "Los buenos tests de comprensión oral incluirán ítems de inferencia. Hacerlos es difícil y lleva tiempo, pero son demasiado importantes para excluirlos" (Buck 2001: 147). Muchos especialistas recomiendan hacer preguntas que vayan más allá del sentido literal del texto, el problema es que estas preguntas son difíciles de hacer ya que la respuesta no aparece de forma explícita en el texto, además es difícil decir con certeza que una inferencia es errónea, al fin y al cabo somos libres de deducir lo que queramos. Hay que recordar que las inferencias se basan en nuestro

conocimiento general del mundo y por lo tanto ese conocimiento debe ser compartido por todos los examinados. Muchos de estos problemas se pueden evitar si se usan ítems de selección múltiple con una respuesta preferible a las demás opciones. En estos casos no hay que olvidar indicarle al examinado que elija la *mejor* opción y no la opción *correcta*. (ibídem: 148)

El tipo de información que puede pedirse en las preguntas de inferencia suele ser: pedir la idea central del texto, preguntar por algo que no se ha dicho claramente pero que el hablante indica ya sea por las palabras que ha elegido o por el tono de la voz, preguntar por la implicación pragmática o la consecución lógica de lo que el hablante ha dicho.

22. TOMAR NOTAS DURANTE LA AUDICIÓN

No parece existir evidencia de una mejora significativa en la actuación de los examinados si se les autoriza a tomar notas pero las investigaciones de Hale y Courtney (1994) evidenciaron que si “se les anima a hacerlo se perjudica significativamente su actuación.”

23. FIJAR EL TIEMPO DE REALIZACIÓN DE LAS TAREAS

Hay que asegurarse de que el examinado tendrá suficiente tiempo para hacer todas las actividades y de que el tiempo que hemos planeado es viable con la situación en que se desarrolla la prueba. Nos puede ayudar tener en cuenta que para marcar una respuesta se suelen necesitar cinco segundos por ítem, pero si hay que escribir una frase para responder, se necesitará un minuto por ítem (Buck 2001: 121).

Los examinados deben saber cuánto tiempo tienen para responder, pero cuando un estudiante pasa más tiempo del esperado pensando una respuesta, suele ser porque no ha entendido el texto lo bastante bien como para responder. Si pide tiempo extra para pensar en las respuestas, hay que asumir que su comprensión escasea y encontrar alguna forma de convencerle para que siga adelante.

24. LA LENGUA DE LAS RESPUESTAS

Pueden darse en la lengua meta o en la lengua materna del examinado (Buck 2001: 125; Hugues, 1989: 165), de forma escrita o hablada. Lo más importante es determinar si la respuesta será evaluada por su corrección, adecuación, etc., y si esa evaluación afectará la puntuación. “En un test de comprensión oral parece razonable no penalizar los errores si la respuesta es inteligible y clara. Después de todo, queremos saber si los examinados han entendido el texto, no si pueden producir lenguaje correctamente” (Buck 2001: 126). Hugues (1989: 170) también insiste en que “en la evaluación de las destrezas receptivas no hay

razones para deducir puntos por errores de gramática u ortografía, siempre y cuando esté clara la intención de la respuesta correcta.”

25. ¿CUÁNTAS VECES ESCUCHAR LA AUDICIÓN?

Según las investigaciones de Berne (1995), y de Cervantes y Gainer (1992), parece evidente que poner la audición una segunda vez hace la tarea más fácil. Desde la perspectiva de la autenticidad situacional e interaccional del lenguaje, escuchar la audición solo una vez parece más lógico porque en el mundo real solemos oír el texto solo una vez y la habilidad de procesar el lenguaje de forma rápida y automática es una subdestreza básica de la comprensión oral que necesita ser evaluada. Asimismo, cuando los oyentes no comprenden algo tienen que hacer deducciones para construir un puente por encima del vacío de comprensión y esta es otra subdestreza de la comprensión oral muy importante (Buck 2001: 171).

Sin embargo, en la mayoría de situaciones de uso de la lengua no es vital que el oyente entienda exactamente todo lo que se está diciendo, y en caso de ser importante que entienda de forma tan precisa, suele tener la oportunidad de pedir aclaraciones o de negociar el significado de alguna forma. De hecho, los textos orales que necesitan ser comprendidos con detalle como el anuncio de un retraso en un tren, o de un cambio en una puerta de embarque, se repite varias veces. La situación de evaluación es poco natural al pedir que el oyente comprenda con mayor grado de precisión de lo que es normal, por ello no parece ser algo tan poco natural poner dos veces una audición (ibídem). Un compromiso puede ser escuchar dos veces cuando la tarea requiere una audición intensiva, de búsqueda de información concreta; y solo una vez cuando las tareas requieren una comprensión extensiva o de comprensión de la idea general.

26. ¿MOSTRAR LAS PREGUNTAS DE COMPRENSIÓN ANTES DE LA AUDICIÓN?

Parece que para los examinados es positivo psicológicamente poder ver las preguntas antes de escuchar el texto y que en cambio la incerteza de no saber por qué se está escuchando tiene un efecto negativo (Buck 2001: 137). Sherman (1997) descubrió en sus investigaciones que la mejora más fuerte en la actuación se daba cuando se mostraban las preguntas después de que los oyentes hubiesen escuchado el pasaje una vez pero antes de escucharlo una segunda. Hugues (1989: 165) también recomienda enseñar antes de la audición las preguntas e incluso darles tiempo para familiarizarse con ellas, excepto en la técnica de *tomar notas* en que las preguntas se dan después de la audición y que veremos a continuación en el punto 27.

27. LAS TÉCNICAS QUE PUEDO UTILIZAR

Todas las tareas tienen sus puntos fuertes y débiles, y tienden a exigir habilidades diferentes dándole ventaja unas veces a unos examinados y otras a otros. “Si usamos una variedad de diferentes tipos de tareas, el test tendrá más probabilidades de ofrecer una evaluación equilibrada y será por tanto un test más justo” (Brindley 1998). Hay muchas técnicas que podemos utilizar para medir la comprensión oral como son: la selección múltiple, los ítems dicotómicos (verdadero/falso), las preguntas de respuesta corta, las de respuesta larga, el relleno de huecos, la transferencia de información, tomar notas, la transcripción, etc. (Pastor 2005: 51-53)⁶.

SELECCIÓN MÚLTIPLE

Ventajas:

- Altamente fiable.
- Respuesta cerrada: puntuación rápida y objetiva.
- Permite responder, en el mismo tiempo, más ítems que otras técnicas.
- Permite la evaluación de destrezas receptivas como la comprensión oral.

Desventajas:

- Es tremendamente difícil escribir ítems eficaces.
- Evalúa solo la capacidad de reconocer, lo cual no asegura la capacidad de uso.
- El azar puede tener un efecto considerable.
- No todo puede evaluarse con esta técnica ya que hay estructuras que no ofrecen alternativas plausibles.
- El *washback* puede ser negativo si se abusa de su uso.

Observaciones:

- Es importante que haya un mínimo de 4 opciones para bajar el posible efecto del azar.
- Los distractores deben ser plausibles y de una longitud similar a la opción correcta.
- Una opción que no atrae no sirve para nada.
- Las opciones deben ser breves para que la memoria a corto plazo las pueda registrar.
- Al elegir las alternativas, podemos buscar información similar a la opción correcta o intentarlo con palabras de sonido similar.
- Algo que parezca plausible para una persona que no haya oído el texto también suele funcionar.
- Hay que asegurarse de que la forma de la pregunta no deja adivinar la respuesta correcta: comprobar que todas las opciones están en consonancia

⁶ Para una descripción detallada de estas técnicas con sus ventajas y desventajas ver Hugues 1989: 165-170.

con la pregunta o frase incompleta, que no es más larga o diferente de algún modo, etc.

- Comprobar que tanto la pregunta como las opciones no son ambiguas, que ninguna de las opciones incorrectas puede considerarse correcta, que la respuesta no se pueda extraer de conocimiento general o por sentido común.

ÍTEMS DICOTÓMICOS (SÍ/NO – VERDADERO/ FALSO)

Ventajas:

- Respuesta cerrada: puntuación rápida y objetiva.
- Permite resolver muchos en poco tiempo.
- Permite la evaluación de destrezas receptivas como la comprensión oral.
- Ítems con preguntas fáciles de hacer.

Desventajas:

- El azar tiene un 50% de posibilidades.

Observaciones

- Se suele incluir una tercera opción: “no aparece en el texto” o se pide que se argumente la elección para disminuir el efecto del azar. Pero esto último añade la habilidad de escribir y además las respuestas son entonces más difíciles de medir. (Hugues 1989: 79)
- Cuántos más ítems se introduzcan, más se reduce el posible efecto del azar.
- Buck (2001: 147) apunta también que otro problema es que unos intentan adivinar las respuestas que no conocen pero otros no. Él recomienda decirle a los examinados que intenten adivinar en esos casos para que todos sigan la misma estrategia. Además, según él, cuando se adivina no se deja todo al azar, los examinados tienen a menudo alguna comprensión en el punto en el que están siendo evaluados y se inclinan a favor de una respuesta o de otra, tal vez sin saber por qué, debido a esa comprensión parcial que también es información relevantes para el constructo que estamos midiendo.

RESPUESTAS CORTAS

Ventajas:

- El azar influye menos en las respuestas.
- No está restringida por la necesidad de distractores.
- Los ítems son más fáciles de escribir.
- Muy adecuadas para evaluar la comprensión de información dada de forma explícita.

Desventajas:

- Las respuestas necesitan más tiempo para ser respondidas por lo que se reduce el número de ítems.
- El examinado tiene que producir lengua para responder.
- Se necesita más tiempo para puntuar y para desarrollar criterios de evaluación.
- Cuando la respuesta no aparece de forma explícita, es necesario determinar qué constituye una interpretación razonable y una respuesta suficiente.

Observaciones:

- Para atenuar las desventajas hay que intentar requerir respuestas cortas, de respuesta única (existe solo una respuesta posible) y que se pueda encontrar en el texto.
- Es muy útil intentar limitar las respuestas a tres palabras.
- Además del formato pregunta-respuesta, pueden presentarse de forma que haya que completar una frase.

RESPUESTAS LARGAS

Ventajas:

- Podemos ir más allá de la superficie del texto.
- Ítems con preguntas fáciles de hacer.

Desventajas:

- Puntuación subjetiva y costosa en tiempo.
- Necesitamos desarrollar una escala de criterios de puntuación: hay que decidir qué respuestas pueden darse y qué se considera suficiente.
- Los evaluadores necesitan practicar la escala hasta aplicarla de forma consistente.

RELLENO DE HUECOS

- Funciona bien si las palabras que faltan pueden encontrarse en el texto oral y los ítems están bien contextualizados.
- Es una alternativa cuando la respuesta única no es posible.
- No se piden habilidades significativas de producción.
- Muy fiable si se elabora una buena clave.

TRANSFERENCIA DE INFORMACIÓN

Está técnica implica actividades como marcar la ruta en un mapa, ponerle un nombre oído a una imagen, etc. Su ventaja es que minimiza las exigencias de otras habilidades.

TOMAR NOTAS

Esta técnica es útil para la comprensión oral de clases magistrales o de conferencias. Los examinados toman nota durante la audición y después responden a los ítems que no habían visto anteriormente.

TRANSCRIPCIÓN

Se le pide a los candidatos que transcriban números (al oír un número de teléfono, por ejemplo) y palabras que se deletrean (alguien dando su nombre al reservar un billete). Para una puntuación fiable hay que exigir una respuesta totalmente correcta al ítem para conceder el punto (Hugues 1989: 169)

28. PRE-TESTAR LAS TAREAS

Hay muchas cosas que pueden no funcionar en una tarea, por ejemplo, puede no haber una respuesta correcta o puede haber dos, o puede ser posible responder sin haber entendido el pasaje. La mejor forma de evitar estos problemas es darles la tarea a personas con el mismo perfil que lo estudiantes a los que va dirigida la prueba o a colegas para que la hagan y pedirles su opinión. Esto no nos dará información sobre la dificultad de la tarea, ni si está midiendo el constructo correcto. Para ello, es necesario pre-testar los ítems con alumnos de perfil similar a los que serán examinados y entonces analizar los resultados con procedimientos estándar de análisis de ítems como por ejemplo *el coeficiente de dificultad* o *el índice de discriminación* (Alderson et al. 1995: 81-86). Este pre-testado nos mostrará si las tareas tienen el nivel de dificultad correctos para poder discriminar bien entre examinados con diferentes habilidades en el constructo. La dificultad de una tarea se puede variar bien modificando el texto o bien modificando la tarea.

Cuando no es posible pre-testarlas podemos ir registrando las respuestas a medida que puntuamos los ítems: apuntamos qué respuestas son definitivamente correctas y cuáles definitivamente incorrectas. Marcamos aquellas que están en la frontera entre lo uno y lo otro. Al ir puntuando acabaremos comprendiendo mejor cómo está interactuando el examinado con la tarea y qué está midiendo la pregunta, en ese momento probablemente desarrollemos una mejor idea de qué casos fronterizos podemos aceptar como correctos y cuáles no. Entonces volvemos atrás y repuntamos las respuestas fronterizas. Esta no es una forma ideal de puntuación, pero en mundo real de recursos limitados, no siempre es posible pre-testar todo lo que nos gustaría. (Buck 2001: 141)

29. ¿QUÉ HACE LA PRUEBA MÁS FÁCIL O MÁS DIFÍCIL?

Buck (2001: 149-151) hace un listado tomando como fuente investigaciones de Anderson y Lynch (1988), Brown (1995), Freedle y Kostin (1996, 1999), Buck et al. (1997), Buck y Tatsuoka (1998) para definir cuáles son las características que hacen los textos y las tareas más fáciles o difíciles.

Los TEXTOS más fáciles se caracterizan porque el habla es más lenta, las pausas entre las unidades que forman una idea más largas, los patrones de entonación son familiares, el vocabulario es de alta frecuencia (común), la gramática sencilla, las ideas se unen en cláusulas coordinadas, hay referencias pronominales simples, las ideas aparecen enunciadas de forma explícita, se da más redundancia, los sucesos están descritos de forma lineal o cronológica, se explicita el punto principal antes de que aparezcan los ejemplos, los temas son conocidos, hay poca gente que distinguir, se localiza fácilmente a la persona u objetos protagonistas, los temas son concretos y no abstractos, la información visual o de otro tipo complementa al contenido y no entra en conflicto con él.

TEXTOS MÁS FÁCILES	TEXTOS MÁS DIFÍCILES
CARACTERÍSTICAS LINGÜÍSTICAS	
Habla más lenta.	Habla más rápida.
Pausas largas entre unidades que forman una idea.	Pausas cortas o ninguna entre unidades de ideas.
Pronunciación más conocida.	Pronunciación con la que no se está familiarizado.
Patrones de entonación naturales.	Patrones de entonación artificiales o inusuales.
Vocabulario de alta frecuencia (palabras comunes).	Vocabulario de baja frecuencia.
Gramática sencilla.	Gramática más compleja.
Ideas reunidas en cláusulas coordinadas.	Ideas organizadas en cláusulas subordinadas.
Referencias pronominales simples.	Referencias pronominales complejas.
EXPLICITAD	
Con ideas enunciadas de forma explícita.	Con ideas menos explícitas.
Con más redundancia (siempre y cuando el oyente se de cuenta de que son informaciones redundantes).	Con menos redundancia.

ORGANIZACIÓN	
Sucesos descritos de forma lineal o cronológica.	Sin una estructura lineal.
Con los puntos principales explícitos claramente antes de los ejemplos.	Con ejemplos ilustrativos antes de decir de qué trata el texto.
CONTENIDO	
Temas más familiares.	Temas más desconocidos.
Con poca gente o cosas a distinguir.	Muchas cosas o personas a distinguir.
Se distingue fácilmente a los protagonistas u objetos más importantes.	Resulta difícil distinguir quiénes o qué son los protagonistas.
Textos con relaciones fijas entre los elementos.	Textos con relaciones cambiantes entre los elementos.
Contenidos concretos.	Contenidos abstractos.
CONTEXTO	
Textos con información visual o de otro tipo que complementa el contenido.	Textos con información visual o de otro tipo que entra en conflicto con el contenido.

Por lo que respecta a las TAREAS, estas son más fáciles cuando: se deja escuchar el texto dos veces, se dan las preguntas antes de la audición, hay menos información que procesar, la información se extrae de un sólo lugar (es decir, no está esparcida por varias partes del texto), hay que recordar información tal y como se da en el texto (pedir un resumen es mucho más difícil), se pide seleccionar información (es más difícil separar hechos de opiniones), se pide información relevante al tema principal (es más difícil comprender detalles irrelevantes), se pide una respuesta inmediata (si es diferida es más difícil).

TAREAS MÁS FÁCILES	TAREAS MÁS DIFÍCILES
Dejar escuchar el texto dos veces.	Escuchar el texto una sola vez.
Dar las preguntas antes de la audición.	Dar las preguntas después de la audición.
Hay que procesar menos información.	Hay que procesar más información.
Extraer información localizada en un solo lugar del texto.	Extraer una información esparcida por el texto.
Recordar información exacta del contenido.	Pedir un resumen o lo esencial de un pasaje.
Seleccionar información.	Separar hechos de opiniones.

Pedir información relevante al tema principal.	Pedir detalles irrelevantes.
Pedir una respuesta inmediata.	Pedir una respuesta diferida.

Buck (2001: 152). también describe las características que hacen más fáciles o difíciles LAS PREGUNTAS Y LOS ÍTEMS según la investigación de Freedle y Kostin (1996, 1999) sobre ítems de selección múltiple de comprensión oral. De este estudio se desprende claramente que el tema afecta a la dificultad, así como a la estructura retórica, pero que los dos determinantes más importantes de dificultad son la colocación de la información necesaria y el grado de solapamiento léxico. Cuando la información necesaria aparece al principio del texto o cuando se repite, los ítems tienden a ser más fáciles. El solapamiento léxico se da cuando las palabras usadas en el pasaje se encuentran en la pregunta o en las opciones. Un solapamiento léxico entre la opción correcta y el texto, especialmente en la llamada *información necesaria* es la mejor forma de predecir ítems fáciles. De modo similar, el solapamiento léxico entre el texto y las opciones incorrectas es la mejor forma de predecir ítems difíciles. Presumiblemente esto se debe a que los examinados tienden a seleccionar opciones que contienen palabras que reconocen del pasaje.

PREGUNTAS MÁS FÁCILES

PREGUNTAS MÁS DIFÍCILES

CUANDO REQUIEREN LA COMPRESIÓN DE ENUNCIADOS EXPLÍCITOS

- | | |
|--|--|
| <ul style="list-style-type: none"> - La información necesaria se repite. - Las opciones correctas contienen más palabras en la información necesaria que las incorrectas. - Las opciones correctas contienen más palabras en sus enunciados que las incorrectas. - Las opciones correctas contienen gramática menos compleja que las incorrectas. - Se dan más cambios de temas en el pasaje. - El tema no es académico. | <ul style="list-style-type: none"> - El pasaje tenía bastante más información antes - La pregunta contenía referentes (como pronombres). |
|--|--|

CUANDO REQUIEREN HACER INFERENCIAS

- | | |
|--|--|
| <ul style="list-style-type: none"> - La información necesaria se coloca al principio del pasaje. - La opción correcta contenía más palabras en su enunciado que las opciones incorrectas. - El tema era sobre arte o ciencias sociales. | <ul style="list-style-type: none"> - La información necesaria se ponía en medio del texto. - La estructura discursiva del pasaje implicaba una comparación. - La información necesaria no se repetía. |
|--|--|

CUANDO REQUIEREN LA IDENTIFICACIÓN DE LA IDEA PRINCIPAL

- | | |
|--|---|
| <ul style="list-style-type: none">- El tema no era académico.- La opción correcta incluía más palabras en el enunciado que la incorrecta. | <ul style="list-style-type: none">- La estructura retórica del pasaje es de problema-solución |
|--|---|

30. ESTABLECER LOS CRITERIOS DE NIVELES DE ACTUACIÓN

¿Qué puntuación será suficiente para considerar que el examinado ha realizado la tarea con éxito? ¿El 50%?, ¿el 70%? Recordemos primero que hemos elegido una evaluación *con relación a un criterio*, lo cual parece ser más complicado que si simplemente comparásemos los resultados entre los demás candidatos (*con relación a la norma*). En opinión de Hugues (1989: 141) la mejor forma de proceder es utilizando las propias tareas para definir el nivel. Es decir, si decimos que en un nivel A2 una persona “comprende instrucciones sencillas relativas a cómo ir de un lugar a otro, tanto a pie como en transporte público” (MCER, 2000: 70) y elegimos una tarea en la que el candidato debe hacer eso, debería pasar la prueba si es capaz de ello. El candidato debería obtener un 100% de ítems correctos en esta tarea, pero como la actuación humana no siempre es perfecta, continua Hugues, podemos bajar a un 80% el número de respuestas correctas para considerar que una tarea ha sido resuelta con éxito.

31. ESTABLECER LOS CRITERIOS DE EVALUACIÓN

Es importante que las tareas sean puntuables, es decir que debe ser posible decir, en base al pasaje escuchado, qué respuestas son correctas y cuáles incorrectas. Si se requiere que se responda con una frase, es importante disponer de criterios para la puntuación claramente determinados y aplicados de forma consistente. Es fundamental que los examinados sepan qué constituye una respuesta suficiente, hay que decirles qué y cuánto se espera en una respuesta. Es también importante que sepan cuántos puntos o qué proporción del total vale cada tarea para que puedan organizarse su tiempo y esfuerzo (Buck 2001: 122).

32. LAS ESPECIFICACIONES DEL EXAMEN

Escribir las especificaciones del examen para que sepamos cómo y para qué es la prueba tanto los diseñadores de la prueba, los evaluadores y los estudiantes. Buck (2001: 199) ofrece un modelo para la prueba de comprensión oral. Esta es una adaptación de las especificaciones para los diseñadores de la prueba:

ESPECIFICACIONES PARA LOS DISEÑADORES DE LAS PRUEBAS

PROPÓSITO DE LA PRUEBA: aprovechamiento

PERFIL DE LOS CANDIDATOS: edad, nivel, L1...

SECCIONES DE LA PRUEBA: cuántas, longitud

SITUACIONES DE USO DE ELE

TIPO DE TEXTOS: fuentes, temas, funciones, naturaleza del lenguaje oral, nivel de formalidad, instrucciones para seleccionarlos, cómo se presentarán (grabados, en vídeo, en vivo)

COMPETENCIA LINGÜÍSTICA: vocabulario, gramática, estructura discursiva, temas, global, detalle, inferencia.

ELEMENTOS LINGÜÍSTICOS: listado de estructuras, vocabulario, nociones, funciones, actos de habla

TIPO DE TAREAS: integrativas, auténticas, objetivas

NÚMERO DE ÍTEMS POR SECCIÓN: puntuación, ponderación

TÉCNICAS: selección múltiple, respuesta corta, relleno de huecos, etc.

INSTRUCCIONES PARA EL CANDIDATO Y EJEMPLOS DE ÍTEMS

CRITERIOS PARA LA PUNTUACIÓN: qué respuesta se considera satisfactoria y cuál no.

33. LA REALIZACIÓN DE LA PRUEBA

Por último y aunque parezca una obviedad, no hay que olvidar comprobar antes de administrar la prueba si ¿funciona el equipo técnico debidamente?, ¿conocen los profesores evaluadores cómo funciona el equipo?, ¿es necesario poner carteles fuera del aula para asegurar silencio?

BIBLIOGRAFÍA CITADA

- Alderson, J. Ch., C. Clapham, D. Wall (1995): *Exámenes de idiomas*. Madrid: CUP, 1998
- Alderson, J. Ch. y T. Linch (1988): *Listening*. Oxford: OUP
- Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*. Oxford: OUP, 2001
- Bachman, L. F. y A. S. Palmer (1996): *Language Testing in Practice*. Oxford: OUP, 2002
- Bermejo, I. y M. Higuera (2000): La comprensión auditiva en el aula de español como lengua extranjera: Recursos en Internet para la elaboración de actividades. *Carabela*, nº 49, pp. 141-146. Madrid: SGEL
- Berne, J. E. (1995) : How does varying pre-listening activities affect second language listening comprehension ? *Hispania*, 78, 316-329
- Bordón, T. (2000): La evaluación de la comprensión auditiva en aprendices de ELE. *Carabela*, nº 49, pp. 77-101, Madrid: SGEL
- Brindley, G. (1998): Assessing Listening Abilities. *Annual Review of Applied Linguistics*, 18, 171-91.
- BROWN, G. (1995): Dimensions of difficulty in listening comprehension. In D. J. Mendelsohn y J. Rubin (Eds.), *A guide for the teaching of second language listening*. San Diego: Dominie press.
- Brown, G. y G. Yule (1983): *Teaching the spoken language*. Cambridge: CUP
- Buck, G. (2001): *Assessing Listening*. Cambridge: CUP, 2002
- Buck, G. y K. Tatsuoka (1998): Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119-157.
- Buck, G., K. Tatsuoka, I. Kostin y M. Phelps (1997): The sub-skills of listening: Rule-space analysis of a multiple-choice test of second language listening comprehension. In A. Huhta, V. Kohonen, L. Kurki-Suonio y S. Luoma (Eds.), *Current developments and alternatives in language assessment*. Tampere: University of Jyväskylä.
- Cassany, D. (1987): *Descriure escriure*. Barcelona: Empúries, 1998
- Cassany, D. (1993): *Reparar la escritura*. Barcelona: Graó, 2000

Cervantes, R. y G. Gainer (1992): The effects of syntactic simplification and repetition on listening comprehension. *TESOL Quarterly*, 26, 118-126.

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>)

Freedlie, R. e I. Kostin (1996): *The prediction of TOEFL listening comprehension item difficulty for minitalk passages: implications for construct validity*. Princeton, NJ: ETC Research Report RR 96-29.

Freedlie, R. e I. Kostin (1999): Does text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16(1), 2-32.

Hale G. A., y R. Courtney (1994): The effects of note-taking on listening comprehension in the Test of English as a Foreign Language. *Language Testing*, 11(1), 29-48.

Hugues, A. (1989): *Testing for Language Teachers*. Cambridge: CUP, 2003

López García, Á. (2002): *Comprensión oral del español*. Madrid: ArcoLibros

Lynch, T. (1988): Theoretical perspective listening. *Annual Review of Applied Linguistics*, 18, 3-19.

Pastor, C. (2005): "La Cenicienta de las destrezas: la comprensión oral y su entrenamiento en el aula". *Unterricht Spanisch*, nº 8, pp. 8-14 y 51-53. Seelze: Friederich Verlag

Poch, D. (2000): Sensación física y realidad mental: de la onda sonora al significado de los enunciados. *Carabela*, nº 49, pp. 5-15, Madrid: SGEL

Sherman, J. (1997): The effect of question preview in listening comprehension tests. *Language Testing*, 14(2), 185-213.

Shohamy, E. y O. Inbar (1991): Validation of listening comprehension tests: the effect of text and question types. *Language Testing*, 8(1), 23-40.