



marcoELE. Revista de Didáctica Español Lengua Extranjera

ISSN: 1885-2211

redaccionmarcoele@gmail.com

MarcoELE

España

Casañ Núñez, Juan Carlos

**Desarrollo de una prueba de comprensión audiovisual**

marcoELE. Revista de Didáctica Español Lengua Extranjera, núm. 22, 2016, -Junio

MarcoELE

España

Disponible en: <https://www.redalyc.org/articulo.oa?id=92153185016>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

CASAÑ NÚÑEZ, JUAN CARLOS  
UNIVERSIDAD POLITÉCNICA DE VALENCIA  
DESARROLLO DE UNA PRUEBA DE COMPRENSIÓN AUDIOVISUAL

BIODATA

Doctorando en el programa en Lenguas y Tecnología de la Universidad Politécnica de Valencia, Máster Universitario en Enseñanza del Español como Lengua Extranjera por la Universidad Internacional Menéndez Pelayo-Instituto Cervantes, licenciado en Filología Hispánica por la Universidad de Valencia y en Filología Inglesa por la Universidad Jaume I (Castellón). Ha desarrollado su labor profesional en las áreas de la enseñanza del ELE y de la educación en España, Estados Unidos, Francia, Polonia y Portugal. Se puede consultar una selección de sus publicaciones en [Academia.edu](https://Academia.edu).

El autor puede ser contactado en la siguiente dirección de correo-e: [juacanue@posgrado.upv.es](mailto:juacanue@posgrado.upv.es)

RESUMEN

Este artículo forma parte de una investigación doctoral que estudia el uso de preguntas de comprensión audiovisual integradas en la imagen del vídeo como subtítulos y sincronizadas con los fragmentos de vídeo relevantes. Anteriormente se han publicado un marco teórico que describe esta técnica (Casañ Núñez, 2015b) y un ejemplo en una secuencia didáctica (Casañ Núñez, 2015a).

El presente trabajo detalla el proceso de planificación, diseño y experimentación de una prueba de comprensión audiovisual con dos variantes que será administrada junto con otros instrumentos en estudios cuasiexperimentales con grupos de control y tratamiento. Fundamentalmente, se pretende averiguar si la subtitulación de las preguntas facilita la comprensión, si aumenta el tiempo que los estudiantes miran en dirección a la pantalla y conocer la opinión del grupo de tratamiento sobre esta técnica.

En la fase de experimentación se efectuaron seis estudios. En el último estudio piloto participaron cuarenta y un estudiantes de ELE (veintidós en el grupo de control y diecinueve en el de tratamiento). Las observaciones de los informantes durante la administración de la prueba y su posterior corrección sugirieron que las indicaciones sobre la estructura del test, las presentaciones de los textos de entrada, la explicación sobre el funcionamiento de las preguntas subtituladas para el grupo experimental y la redacción de los ítems resultaron comprensibles. Los datos de las dos variantes del instrumento se sometieron a sendos análisis de facilidad, discriminación, fiabilidad y descriptivos. También se calcularon las correlaciones entre los test y dos tareas de un examen de comprensión auditiva. Los resultados mostraron que las dos versiones de la prueba estaban preparadas para ser administradas.

**PALABRAS CLAVE:** comprensión auditiva, comprensión audiovisual, evaluación de la comprensión audiovisual, subtitulación, vídeo

ABSTRACT

This paper is part of wider doctoral research exploring the use of audiovisual comprehension questions embedded in the image as subtitles and synchronized with the relevant fragments. A framework describing this technique (Casañ Núñez, 2015b) and an example in a pedagogical sequence (Casañ Núñez, 2015a) have already been published.

This present work details the planning, design and trialling of an audiovisual comprehension test with two variants to be administered together with other instruments in quasi-experimental studies with control and treatment groups. Basically, the aim is to investigate whether subtitled questions facilitate comprehension, whether students' time looking at the screen increases, and what the treatment group thinks about the technique.

In the trialling phase, six studies were conducted. Forty-one SFL students participated in the last pilot study (twenty-two in the control and nineteen in experimental group). The observations of the informants during the test administration and the subsequent correction suggested that the directions on the test structure, the presentation of the input texts, the explanation of how the subtitled questions work for the experimental group and the wording of the items were all understood. The data of each version of the instrument was subjected to analysis of facility, discrimination, reliability and descriptives. Correlations between the tests and two tasks from a listening exam were also calculated. Results showed that the two versions of the test were ready to be administered.

**KEY WORDS:** listening, audiovisual comprehension, testing audiovisual comprehension, subtitling, video

# ÍNDICE

1. Antecedentes	5
1.1. Concepto de comprensión auditiva / audiovisual	5
1.2. Procesos de comprensión auditiva / audiovisual	5
1.3. La dimensión visual es parte del mensaje	6
1.4. Investigaciones sobre el impacto de la imagen, la conducta visual y la valoración de la imagen	7
1.5. Complejidad de las actividades de comprensión audiovisual simultáneas al visionado	8
1.6. Preguntas de comprensión audiovisual integradas en el vídeo como subtítulos	8
1.7. La evaluación de la comprensión auditiva / audiovisual con vídeo	9
1.8. El estudio actual	10
2. Desarrollo del test	12
2.1. Planificación	12
2.1.1. Propósito y destinatarios	12
2.1.2. Necesidad	13
2.1.3. Ámbito de uso de la lengua meta	13
2.1.4. Delimitación del constructo	13
2.1.5. Longitud y duración del test	15
2.1.6. Nivel de los ítems	16
2.1.7. Tipo de ítems	16
2.1.8. Puntuación de los ítems	17
2.1.9. Otras consideraciones	17
2.2. Diseño	19
2.2.1. Especificaciones de los textos de entrada	19
2.2.2. Indicaciones para los examinandos	21
2.2.2.1. Presentación de la estructura general del test	21
2.2.2.2. Introducción de los textos de entrada	21
2.2.2.3. Explicación del funcionamiento de las preguntas subtítuladas	21
2.2.3. Primera formulación de los ítems	22
2.3. Experimentación	22
2.3.1. Estudio 1	23
2.3.2. Estudio 2	24
2.3.2.1. Objetivos	24
2.3.2.2. Diseño de investigación	24
2.3.2.3. Método	24
2.3.2.3.1. Participantes	24
2.3.2.3.2. Instrumento	25
2.3.2.3.3. Procedimiento	27
2.3.2.3.4. Análisis de datos	27
2.3.2.4. Resultados y discusión	27
2.3.3. Estudio 3	29
2.3.3.1. Objetivos	29
2.3.3.2. Diseño de investigación	29
2.3.3.3. Método	30
2.3.3.3.1. Participantes	30
2.3.3.3.2. Instrumento	30
2.3.3.3.3. Procedimiento	32
2.3.3.3.4. Análisis de datos	32
2.3.3.4. Resultados y discusión	33
2.3.4. Estudio 4	34
2.3.5. Estudio 5	36
2.3.5.1. Objetivos	36
2.3.5.2. Diseño de investigación	36
2.3.5.3. Método	37
2.3.5.3.1. Participantes	37

2.3.5.3.2. Instrumento	40
2.3.5.3.3. Procedimiento	42
2.3.5.3.4. Análisis de datos	43
2.3.5.4. Resultados y discusión	44
2.3.5.4.1. Claridad de las indicaciones y los ítems	44
2.3.5.4.2. Tiempo empleado	44
2.3.5.4.3. Análisis de facilidad	44
2.3.5.4.4. Análisis de discriminación	45
2.3.5.4.5. Análisis de fiabilidad interna	46
2.3.5.4.6. Análisis descriptivo	47
2.3.5.4.7. Correlaciones	48
2.3.5.4.8. Limitaciones	50
2.3.6. Estudio 6	51
2.3.6.1. Objetivos	52
2.3.6.2. Diseño de investigación	52
2.3.6.3. Método	53
2.3.6.3.1. Participantes	53
2.3.6.3.2. Instrumento	53
2.3.6.3.3. Procedimiento	53
2.3.6.3.4. Análisis de datos	54
2.3.6.4. Resultados y discusión	54
3. Conclusiones	56
Agradecimientos	58
Bibliografía	58
Anexos	65
Anexo 1	65
Anexo 2	66
Anexo 3	67
Anexo 4	68
Anexo 5	69

## 1. ANTECEDENTES

### 1.1. CONCEPTO DE COMPRENSIÓN AUDITIVA / AUDIOVISUAL

A diferencia de la expresión oral y escrita, que tienen productos observables, la comprensión auditiva sucede de manera interna, invisible a los ojos del observador. Incluso cuando el oyente responde con un gesto o una interjección que sugieren que ha comprendido, el interlocutor no puede tener la seguridad de que haya sido así (Lynch, 2009). Debido a este carácter invisible, resulta complejo analizar su naturaleza y llegar a una descripción definitiva.

En este artículo se concibe la comprensión auditiva como un proceso de interpretación de información auditiva y visual, como sugieren especialistas como Riley (1979), Martín Peris (1991), Rubin (1995) y Lynch (2012). Rubin (1995: 7) propone la siguiente definición: “an active process in which listeners select and interpret information which comes from auditory and visual cues in order to define what is going on and what the speakers are trying to express”. Asimismo, se aboga por una terminología que recoja el carácter multimodal de la mayor parte de las situaciones de comprensión auditiva, como apuntan autores como Riley (1979), Harris (2003) y Lynch (2012). De este modo, se sugiere destinar *comprensión auditiva* a la comprensión de información a través del canal oral (por ejemplo, una conversación telefónica), utilizar *comprensión audiovisual* para describir la comprensión de información simultáneamente por el canal oral y el visual (por ejemplo, la comprensión de material en vídeo), y emplear *comprensión auditiva / audiovisual* para hablar de forma general. En lo que resta de artículo se sigue esta propuesta, aunque somos conscientes de que el término *comprensión auditiva* predomina en la literatura.

### 1.2. PROCESOS DE COMPRENSIÓN AUDITIVA / AUDIOVISUAL

La comprensión auditiva / audiovisual puede parecer un proceso sencillo y natural: un interlocutor produce un enunciado e inmediatamente otro interlocutor puede comprenderlo. Sin embargo, desde un punto de vista psicolingüístico, resulta sorprendente que el oyente sea capaz de percibir e interpretar la información verbal y no verbal en tiempo real.

En la actualidad se distinguen dos sentidos en el procesamiento: de abajo arriba (también conocido como *bottom-up*, ascendente o sintético) y de arriba abajo (también llamado *top-down*, descendente o analítico). Nunan (1999: 200) describe el procesamiento de abajo arriba como “a process of decoding the sounds that one hears in a linear fashion, from the smallest meaningful units (or phonemes) to complete texts”. Lynch (2009: 162) define el procesamiento de arriba abajo como “the listener’s use of higher level information, such as expectations and topic knowledge, to construct an interpretation of an incoming message”. Estos procesos son interdependientes. Flowerdew y Miller (2005: 25) relatan que “experimental subjects are unable to identify truncated sounds in isolation from the words they form part of, whereas subjects are quite able to identify truncated words so long as they are presented with the surrounding context”. Para Field (2008: 133) “we could not identify the topic of an utterance without some minimal

decoding, and we could not appreciate the relevance of the utterance without some minimal use of context and co-text”.

Un aspecto clave en la actualidad es conocer cómo se relacionan los procesos *top-down* y *bottom-up* (Lynch, 2006; Field, 2008). Lynch (2010) describe un estudio realizado por Long en 1990 en el que se comparó los efectos del conocimiento del mundo y del nivel de comprensión auditiva de estudiantes americanos de español. Long llegó a la conclusión de que el nivel de comprensión auditiva juega un papel importante cuando los oyentes no poseen conocimientos relevantes sobre el tema, y un papel menos trascendente cuando disponen de estos conocimientos. Long también encontró evidencias de efectos distorsionadores de los esquemas mentales. Esto lleva a la autora a comentar que “it is clear that schemata can hurt, as well as help” (citado en Lynch, 2010: 78). Para Field (2008: 131) el contexto y el cotexto tienen dos funciones: “they enrich the raw meaning of the utterance and make it relevant to the current situation” y “they provide extra evidence that assists the decoding process”. También considera que los oyentes competentes y los poco competentes utilizan el contexto y el cotexto de forma diferenciada. Los oyentes competentes en L1 o L2 los utilizan para enriquecer su comprensión del mensaje mientras que los poco competentes los emplean sobre todo para compensar la incompreensión de partes del mensaje.

### 1.3. LA DIMENSIÓN VISUAL ES PARTE DEL MENSAJE

Varios autores han señalado la importancia de lo visual en la comprensión auditiva / audiovisual. Riley (1979: 84) considera que “such factors as proxemics, kinesics and deictics are all part of the message. They are not just a sort of gloss on the verbal component”. Para Brown y Yule (1983: 85) “the visual environment gives an enormously important extra dimension of information”. Lynch (2010: 80) afirma que “the power of the visual component of communication should not be underestimated”.

¿Qué aporta la imagen? Según Brown y Yule (1983), la dimensión visual proporciona información sobre los interlocutores, el contexto, y la relación entre los hablantes. En un análisis sobre el rol de lo visual en la comprensión auditiva / audiovisual, Riley (1979) identifica seis funciones comunicativas de lo visual:

- a) *Deictic*: referencias a elementos presentes en la situación comunicativa.
- b) *Interactional*: señales para expresar el cambio de turno entre los hablantes.
- c) *Modal*: la manera de comprometerse con lo que se dice (por ejemplo, decir algo con cara de enfadado o sonriendo).
- d) *Indexical*: proporciona información sobre uno mismo (por ejemplo, edad, tamaño o estado emocional)
- e) *Linguistic*: conductas no verbales que tienen un significado preciso.
- f) *Situational*: el impacto o relevancia del contexto espacio-temporal en la conducta comunicativa.

Para Whittaker (1995: 523) lo visual puede facilitar la comunicación de tres formas posibles: “by providing (1) cognitive cues to enhance shared understanding, (2) social cues for providing information about interpersonal communication and affect, or (3) conversation process cues to facilitate speaker change”. Según el autor, la información

sobre el grado de comprensión y el turno de palabra se puede comunicar eficazmente de manera verbal, pero no ocurre lo mismo con la información de carácter afectivo. Whittaker sugiere que podría deberse al carácter involuntario de gran parte de la información afectiva.

#### 1.4. INVESTIGACIONES SOBRE EL IMPACTO DE LA IMAGEN, LA CONDUCTA VISUAL Y LA VALORACIÓN DE LA IMAGEN

En los últimos años ha aumentado el número de investigaciones sobre el impacto de la imagen del vídeo en la comprensión de una LE, sobre todo, en el área de la evaluación de la comprensión auditiva / audiovisual. Estos estudios suelen utilizar un diseño cuasiexperimental en el que un grupo responde a un test de comprensión a partir de una grabación en vídeo y el otro grupo contesta al mismo test a partir de una versión solo audio del vídeo. Las investigaciones no son concluyentes y aportan resultados aparentemente contradictorios. Algunos estudios no hallan diferencias significativas entre efectuar un mismo test con una grabación audiovisual y hacerlo con una versión solo audio del mismo vídeo (Gruba, 1993; Coniam, 2001), otros encuentran que se obtienen mejores resultados con la versión solo audio (Suvorov, 2008) y los hay que revelan mejores puntuaciones medias con la versión audiovisual (Sueyoshi y Hardison, 2005).

Algunos autores han investigado cuál es el comportamiento visual de estudiantes de LE cuando realizan una actividad de comprensión de forma simultánea a la emisión del vídeo. Si asumimos que la comprensión auditiva / audiovisual comprende la interpretación de información auditiva, la media del tiempo de visualización de los participantes sugiere menos atención al vídeo de la que sería deseable: en el estudio de Ockey (2007), la media de observación es el 44,9 %<sup>1</sup> del tiempo, Wagner (2007) obtiene el 69 % y Wagner (2010) el 47,9%. Cabe añadir que estas investigaciones contabilizan el tiempo que los participantes miran en dirección a la pantalla, y que el tiempo de concentración en las imágenes podría ser diferente.

En relación con la valoración de la imagen por parte de los examinandos, es globalmente positiva en algunos estudios (Sueyoshi y Hardison, 2005; Wagner, 2010) y negativa en otros (Alderson, Clapham y Wall, 1995).

Estas divergencias pueden deberse a una gran diversidad de factores, entre ellos, la focalización de las preguntas, la complejidad de la tarea, la manera como se ha efectuado el visionado, el origen y tipo de texto, la calidad del *input*, la perceptibilidad del *input* visual, la mayor o menor solidaridad entre lo visual y lo oral en los textos audiovisuales, el estrés generado por el test, la influencia de la cámara de vídeo en la conducta visual...

---

<sup>1</sup> Ockey (2007) no proporciona la media, pero a partir del tiempo que cada participante mira hacia la pantalla se puede calcular que es el 44,9 %.

## 1.5. COMPLEJIDAD DE LAS ACTIVIDADES DE COMPRESIÓN AUDIOVISUAL SIMULTÁNEAS AL VISIONADO

Escuchar, leer y escribir simultáneamente es extremadamente difícil, sobre todo en una lengua extranjera (Underwood, 1989). Escuchar, ver, leer y escribir al mismo tiempo puede serlo todavía más. Por un lado, existe un conflicto de atención visual entre ver el vídeo y completar una actividad escrita de forma simultánea. Este conflicto contribuye a explicar el relativamente bajo grado de atención a la imagen en pruebas de comprensión con vídeo referidas anteriormente (44,9 % en Ockey, 2007; 69 % en Wagner, 2007; y 47,9 % en Wagner, 2010). Estos niveles de atención a la imagen podrían repercutir negativamente tanto en la comprensión del vídeo como en el desarrollo de la comprensión auditiva / audiovisual. Por otro lado, la memoria de trabajo tiene una capacidad muy limitada para procesar información nueva (Sweller, Ayres y Kalyuga, 2011) y esto debe tenerse en cuenta en el diseño de las actividades simultáneas al visionado. Como señalan Vandergrift y Goh (2012), prestar atención a la imagen, al audio, y a la tarea de forma simultánea puede llevar a la saturación de la memoria de trabajo. Para reducir el conflicto de atención visual y la carga cognitiva de la actividad, se recomienda que las actividades escritas impliquen poca lectura y escritura, y que requieran mantener pocos elementos activos en la memoria de trabajo. También se sugiere la utilización de la técnica del visionado con pausas propuesta por autores como Stempleski y Tomalin (2001). En tercer lugar, se aconseja la aproximación espacial del vídeo y la actividad o su integración dentro de una misma pantalla para reducir el tiempo necesario para desplazar la vista de un estímulo a otro (de la actividad al vídeo y viceversa).

## 1.6. PREGUNTAS DE COMPRESIÓN AUDIOVISUAL INTEGRADAS EN EL VÍDEO COMO SUBTÍTULOS

A partir de la revisión de la bibliografía, la propia experiencia docente con la subtitulación de preguntas de comprensión audiovisual, y las aportaciones del Prof. Dr. Rafael Arnold (Universidad de Rostock), la Prof.<sup>a</sup> Dra. Ana María Gimeno (Universidad Politécnica de Valencia), la Prof.<sup>a</sup> Dra. María Virginia González (Universidad de Valencia), el Prof. Dr. Francisco Javier Satorre (Universidad de Valencia), el Prof. Dr. Agustín Yagüe (EOI Drassanes, de Barcelona) y un profesor doctor que prefiere mantenerse en el anonimato, Casañ Núñez (2015b) redacta un marco teórico sobre el uso de preguntas de comprensión audiovisual integradas en el vídeo como subtítulos. Básicamente, el procedimiento consiste en que cada cuestión aparece en pantalla segundos antes de que comience el fragmento relevante del texto audiovisual y permanece visible hasta su finalización. Esta técnica novedosa podría aportar varios beneficios frente a visionados con los materiales de la actividad en papel. Entre ellos, podría reducir el conflicto de atención entre ver el vídeo y realizar la tarea, al aproximar espacial y temporalmente las preguntas y los fragmentos relevantes, y podría disminuir las necesidades cognitivas de la actividad, al no tener que prestar atención a varias preguntas de forma simultánea. El procedimiento no está exento de inconvenientes. Su principal limitación radica en que podría no preparar adecuadamente para visionados sin preguntas subtituladas. Para conocer más detalles sobre esta metodología se sugiere la consulta del marco teórico (Casañ Núñez, 2015b) y de un ejemplo en una secuencia didáctica para la enseñanza del ELE (Casañ Núñez, 2015a).

## 1.7. LA EVALUACIÓN DE LA COMPREENSIÓN AUDITIVA / AUDIOVISUAL CON VÍDEO

No hay consenso sobre la idoneidad del uso del vídeo en la evaluación de la comprensión auditiva / audiovisual. Entre las razones para su incorporación destacamos las siguientes: es consistente con una definición de la comprensión auditiva / audiovisual como proceso integrador de información auditiva y visual, como proponen autores como Martín Peris (1991); Rubin (1995) y Lynch (2012); está en consonancia con las prácticas pedagógicas en el aula (Gruba, 1997; Buck, 2001); incrementa la validez del test (Bejar, Douglas, Jamieson, Nissan, y Turner, 2000); Wagner, 2007, 2008), su autenticidad (Bejar et al., 2000; Alderson, 2005; Wagner, 2007, 2008) y su naturalidad (Alderson, 2005); y permite activar los esquemas mentales y proporciona un punto de atención mientras se está escuchando (Buck, 2001). Como argumentos en contra se han señalado que los estudiantes podrían estar demasiado ocupados con las preguntas para prestar atención a la imagen (Alderson, et al., 1995), los costes de producción y puesta en práctica (Bejar et al., 2000; Buck, 2001), y que las pruebas suelen medir la habilidad de comprender la lengua y no la información visual (Buck, 2001).

Los exámenes de certificación de nivel del español y del inglés como lenguas extranjeras más difundidos (los DELE, el *TOEFL* y los exámenes de la Universidad de Cambridge) no emplean grabaciones audiovisuales para evaluar la comprensión auditiva/audiovisual. El Servicio Internacional de Evaluación de la Lengua Española<sup>2</sup> (en adelante SIELE) y las pruebas de diagnóstico DIALANG<sup>3</sup> tampoco utilizan material en vídeo. Por el contrario, algunos exámenes de ELE de menor envergadura sí cuentan con un test de comprensión con vídeo. Cabe mencionar las pruebas de certificación de nivel intermedio y avanzado de ELE de las Escuelas Oficiales de Idiomas (en adelante EEOOI) de Cataluña<sup>4</sup>. Desafortunadamente, la información disponible sobre estos test es imprecisa. En relación con el certificado de nivel intermedio de ELE, en el documento *Descripción de la prueba y recomendaciones para los candidatos*<sup>5</sup> no se especifican de forma clara los temas del test de comprensión auditiva/audiovisual (“sobre temas generales [...] sobre temas de la vida cotidiana y de su interés”) ni el número y características de los textos de entrada (“un mínimo de tres textos de duración y tipología diferentes”) ni la cantidad de tareas e ítems (“preguntas de opción múltiple y de verdadero-falso”). Además, sorprende que en algunas regiones se evalúe la comprensión auditiva/audiovisual en ELE con vídeo y en otras con audio (por ejemplo, en las EEOOI de Valencia y Galicia), dado que las pruebas de certificación de las EEOOI están reconocidas en todo el estado español. Sobre este

---

<sup>2</sup> El SIELE es un nuevo servicio de evaluación y certificación del español promovido por el Instituto Cervantes, la Universidad Nacional Autónoma de México, la Universidad de Salamanca y la Universidad de Buenos Aires. La web del examen es: <https://siele.org/web/guest/home> (consulta: 01/03/2016)

<sup>3</sup> “El proyecto DIALANG es un sistema de evaluación pensado para alumnos de idiomas que quieren conseguir información de diagnóstico sobre su dominio de la lengua estudiada” (Consejo de Europa: 2002: 217). La versión actual data de 2006 y es accesible desde la siguiente dirección: <http://www.lancaster.ac.uk/researchenterprise/dialang/about.htm> (consulta: 01/03/2016)

<sup>4</sup> La información pública sobre estos exámenes se encuentra en la siguiente página web: <http://ensenyament.gencat.cat/ca/serveis-tramits/proves/proves-lliures-obtencio-titols/convocat-ordinaria-idiomes/mostres-proves/espanyol/> (consulta: 01/03/2016)

<sup>5</sup> El documento es accesible en la siguiente dirección: [http://ensenyament.gencat.cat/web/.content/home/serveis-tramits/proves/proves-lliures-obtencio-titols/convocat-ordinaria-idiomes/mostres\\_proves/espanyol/Esplnt\\_descr\\_recom.pdf](http://ensenyament.gencat.cat/web/.content/home/serveis-tramits/proves/proves-lliures-obtencio-titols/convocat-ordinaria-idiomes/mostres_proves/espanyol/Esplnt_descr_recom.pdf) (consulta: 01/03/2016)

asunto, Gozalo Ausín (2013: 2) señala que “las características de la organización y redacción de las pruebas son enormemente variopintas”.

En el ámbito autonómico, cabe destacar el Diploma de Lengua Española para Trabajadores Inmigrantes<sup>6</sup> (en adelante Diploma LETRA). Se trata del único examen de cierta envergadura que conocemos en ELE que contiene una prueba de comprensión audiovisual y que ha difundido parte de su proceso de desarrollo (v. Martín Leralta, 2011). El Diploma LETRA ha sido diseñado por el Grupo de Investigación en Lingüística aplicada a la enseñanza del español de la Universidad Nebrija, y “pretende evaluar un nivel inicial de dominio de la competencia comunicativa de los trabajadores inmigrantes, fundamentalmente en el ámbito administrativo-laboral” (Baralo Ottonelo y Martín Leralta [coords.], ca. 2011: 2). La prueba de comprensión audiovisual tiene una duración de veinticinco minutos, está constituida por tres tareas y evalúa tres capacidades.

Tabla 1. Tareas y capacidades evaluadas en el examen LETRA  
(Baralo Ottonelo y Martín Leralta [coords.], ca. 2011: 7 y 8).

TAREAS
Tarea 1. Visionado de un vídeo con una conversación informal en el ámbito personal o público. Tarea 2. Visionado de un vídeo con una conversación transaccional en el ámbito administrativo laboral. Tarea 3. Visionado de un vídeo con una conversación transaccional en el ámbito laboral / de empleo y trabajo.
CAPACIDADES EVALUADAS
El trabajador inmigrante puede en discursos sencillos: a) Comprender las ideas principales. b) Extraer datos concretos (p.e. número de teléfono, deletreo de un nombre, instrucciones sobre cómo llegar a un sitio, documentación). c) Inferir datos que no están presentes en los enunciados, a partir de recursos lingüísticos y no lingüísticos.

## 1.8. EL ESTUDIO ACTUAL

Este trabajo forma parte de una investigación conducente al título de doctor que estudia el uso de preguntas de comprensión audiovisual integradas en la imagen del vídeo como subtítulos y sincronizadas con los fragmentos de vídeo relevantes para responderlas. Anteriormente se ha difundido un marco teórico que describe esta metodología novedosa (Casañ Núñez, 2015b) y una aplicación de este marco al diseño de una secuencia didáctica (Casañ Núñez, 2015a).

Una de las estrategias contempladas para validar algunos aspectos del marco teórico consiste en llevar a cabo estudios cuasiexperimentales con grupos de control y tratamiento. En estos estudios se administran tres instrumentos: dos cuestionarios y un test de comprensión audiovisual. Uno de los cuestionarios (en adelante cuestionario previo) tiene, fundamentalmente, el propósito de conocer a la muestra, averiguar sus preferencias de aprendizaje en relación con la práctica de la comprensión auditiva /

---

<sup>6</sup> La página web oficial del examen es la siguiente: <http://www.diplomalettra.com/> (consulta: 01/03/2016).

audiovisual en clase y saber cómo practican esta actividad comunicativa fuera del aula. El otro cuestionario (en adelante cuestionario posterior) tiene el fin de averiguar la opinión del grupo de tratamiento sobre la subtitulación de preguntas de comprensión audiovisual y se administra después de completar cada una de las tareas del test. La prueba cuenta con dos versiones: una variante para el grupo de tratamiento, en adelante denominada experimental, y otra para el grupo de control, en adelante denominada tradicional. En ambas versiones, el test se entrega en papel. En la versión experimental, además, los enunciados de las preguntas de comprensión audiovisual aparecen integrados en la imagen a modo de subtítulos. Básicamente, se pretende averiguar si la subtitulación de las preguntas facilita la comprensión, si aumenta el tiempo que los estudiantes miran en dirección a la pantalla y conocer la opinión del grupo de tratamiento sobre esta técnica.

El presente trabajo describe el proceso de desarrollo del test de comprensión audiovisual en sus dos variantes. La elaboración de la prueba se articula en tres fases interrelacionadas: planificación, diseño y experimentación<sup>7</sup> (v. figura 1). En la primera fase se detallan las especificaciones del test. Se describen los siguientes aspectos: propósito y destinatarios, necesidad, ámbito de uso de la lengua meta, delimitación del constructo, longitud y duración del test, nivel de los ítems, tipo de ítems, puntuación de los ítems, lengua del test y de la respuesta, momento de entrega de las preguntas y estrategias adoptadas para reducir el conflicto de atención. En la fase de diseño se precisan las características de los textos de entrada, se explicita la información que reciben los examinandos en relación con la estructura general del test, los textos de entrada y las preguntas subtituladas, y se redacta el test. La etapa de experimentación consiste en la administración de la prueba para recabar información y analizarla. Esto incluye pruebas piloto y estudios con otros fines, como la estimación del nivel aproximado de los ítems.

El desarrollo del test sigue un proceso lineal y también recursivo. Es lineal en la medida en que las fases se suceden unas a otras; y es recursivo en tanto en cuanto los resultados y decisiones de una fase pueden modificar una fase anterior. Esto es especialmente evidente entre las fases 2 y 3, en las que se diseñan y pilotan sucesivas versiones de las tareas hasta conseguir una prueba satisfactoria.

## 1. PLANIFICACIÓN

Propósito y destinatarios  
Necesidad  
Ámbito de uso de la lengua meta  
Delimitación del constructo  
Longitud y duración del test  
Nivel de los ítems  
Tipo de ítems  
Puntuación de los ítems  
Otras consideraciones (lengua del test y de la respuesta, momento de entrega de las preguntas y estrategias adoptadas para reducir el conflicto de atención)

---

<sup>7</sup> Se trata de pasos que se suelen considerar en la bibliografía. Por ejemplo, Bachman y Palmer (1996) proponen las etapas de *design*, *operationalization* y *administration* y Bordón (2006) habla de *fase previa*, *fase de diseño*, *fase de experimentación* y *fase de recapitulación*.

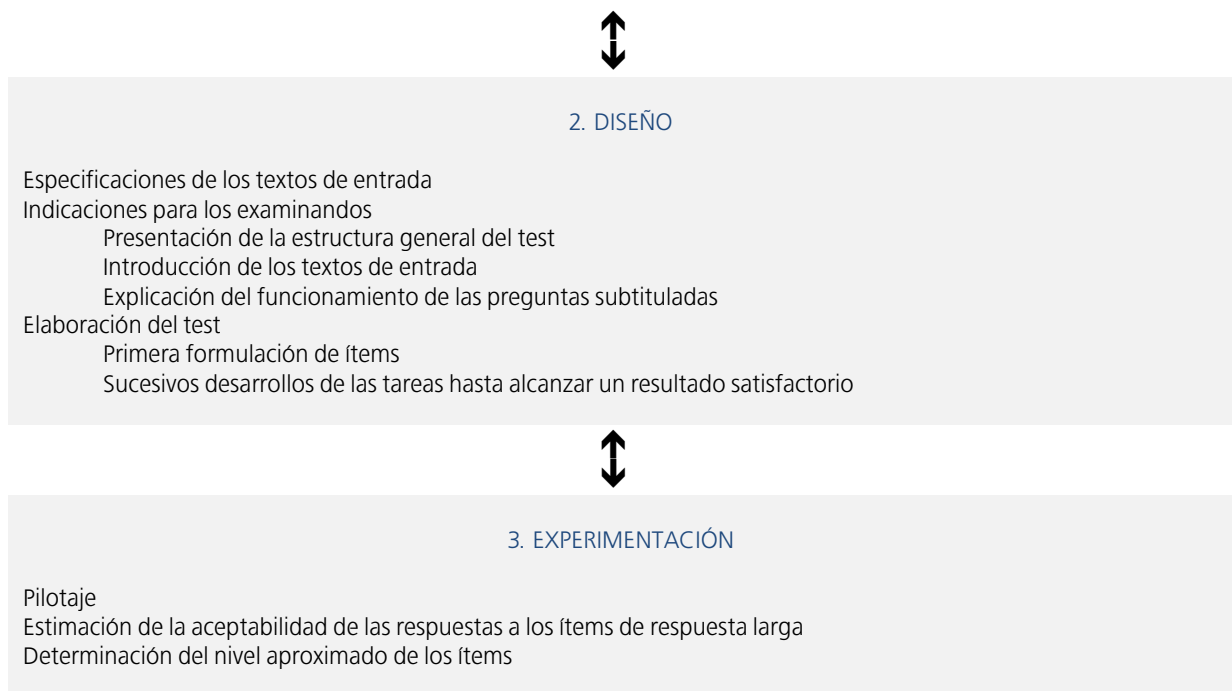


Figura 1. Esquema del desarrollo del test.

## 2. DESARROLLO DEL TEST

La exposición del desarrollo de la prueba sigue las tres fases presentadas en la figura 1 del § 1.8. con un matiz: por claridad, las sucesivas versiones de las tareas se describen en los estudios piloto de la fase de experimentación.

### 2.1. PLANIFICACIÓN

#### 2.1.1. PROPÓSITO Y DESTINATARIOS

El test se diseña con fines de investigación y para ser administrado junto con otros instrumentos (v. § 1.8.). Los destinatarios son estudiantes adultos de ELE capaces de ver cine subtitulado.

La prueba está dirigida a varios niveles de comprensión audiovisual en función de los objetivos de investigación y el nivel de los textos y los ítems. Con el propósito de comparar las puntuaciones del test en estudios cuasiexperimentales, conocer la opinión del grupo de tratamiento sobre esta técnica e investigar la conducta visual de los grupos está destinada a aprendices con un nivel B1+/B2. Asimismo, se pretende administrar los ítems basados en el texto más sencillo (la secuencia de *Ópera prima*) a personas con un nivel de comprensión A2 / B1 para estudiar de forma limitada los mismos aspectos. Por último, aunque en niveles superiores al B2 es previsible que el test resulte muy fácil y no se aprecien diferencias en las puntuaciones de las dos versiones de la prueba, se contempla realizar estudios para conocer la opinión del grupo de tratamiento y comparar la conducta visual de los grupos.

## 2. 1. 2. NECESIDAD

Es preciso elaborar el test por dos razones. En primer lugar, no hemos encontrado en la bibliografía pruebas de comprensión audiovisual para los mismos destinatarios y con el mismo ámbito de uso de la lengua meta que pudieran ser empleadas en la investigación. En segundo lugar, la subtitulación de preguntas de comprensión audiovisual es un procedimiento novedoso y desconocemos la existencia de aplicaciones en la didáctica y evaluación de lenguas extranjeras.

## 2. 1. 3. ÁMBITO DE USO DE LA LENGUA META

Bachman y Palmer (1996: 44) definen *target language use domain* como “a set of specific language use tasks that the test takers are likely to encounter outside of the test itself, and to which we want our inferences about language ability to generalize”. El ámbito de uso de la lengua meta (ULM) es la comprensión de conversaciones informales del ámbito personal<sup>8</sup> procedentes de comedias románticas cinematográficas españolas. Se trata, pues, de situaciones en las que no se puede participar como interlocutor, y en las que se puede observar a los interlocutores y el contexto situacional.

## 2. 1. 4. DELIMITACIÓN DEL CONSTRUCTO

Anteriormente se ha mencionado una serie de argumentos a favor de la inclusión del vídeo en la evaluación de la comprensión auditiva / audiovisual (v. § 1.7.). En este momento es oportuno añadir una razón relacionada con el constructo. Según Wagner (2008), el ULM debe determinar las características de las tareas del test. Por esta razón, el autor considera que si el ULM presenta una situación de comprensión en la que el oyente puede ver al interlocutor (y por lo tanto utilizar la información no verbal), el constructo debe incluir la dimensión visual. De lo contrario, el constructo estaría infrarrepresentado y se producirían amenazas a la validez del test. Si el ULM consistiera en situaciones de comprensión en las que no se puede ver al hablante, se debería prescindir de la dimensión visual para no generar varianza irrelevante de constructo<sup>9</sup>. En el estudio actual, el ULM justifica la presencia de la dimensión visual.

Para definir el constructo se ha consultado en primer lugar el *MCER*. Desgraciadamente, como señala Alderson (2005), el *MCER* no proporciona una descripción del constructo de la comprensión auditiva / audiovisual ni considera el tipo de operaciones mentales que un oyente realiza cuando escucha. También se han analizado los constructos de pruebas de comprensión auditiva y audiovisual. Como se ha comentado en el § 1.7., en el ámbito de los exámenes de ELE en España, el único examen de cierta envergadura que conocemos

---

<sup>8</sup> El ámbito personal “es el de la vida privada del individuo que se centra en su familia y en sus amigos, y en el que se realizan prácticas individuales tales como la lectura por placer, la escritura de un diario personal, la dedicación a un interés particular o a una afición, etcétera” (Consejo de Europa, 2002: 49).

<sup>9</sup> Messick (1996: 4) describe estos dos tipos de amenazas a la validez. “In the threat to validity known as *construct underrepresentation* (which jeopardizes authenticity), the assessment is deficient: the test is too narrow and fails to include important dimensions or facets of focal constructs”. “In the threat to validity known as *construct-irrelevant variance* (which jeopardizes directness), the assessment is too broad, containing excess reliable variance that is irrelevant to the interpreted construct”.

en ELE que contiene una prueba de comprensión audiovisual y que ha difundido parte de su proceso de desarrollo es el Diploma LETRA. En tercer lugar, se han examinado las taxonomías de habilidades<sup>10</sup> que componen la comprensión auditiva / audiovisual sugeridas por autores como Richards (1983), Rost (1990, 2002), Weir (1993) y Buck y Tatsuoka (1998). La intención principal de mayor parte de estos listados “is to aid in instruction or test-writing by identifying subskills that can be practiced or tested in isolation” (Rost, 2002: 120). Sin embargo, estas propuestas no están exentas de inconvenientes. Buck (2001: 51) señala que hay que tomar las taxonomías con cautela porque “those based on theory are only lists of what scholars think are likely to be important, whereas those based on research use complex statistical techniques that are still not fully understood”. El autor añade que las taxonomías no indican la importancia relativa de cada habilidad ni como utilizarlas en la construcción de test. Asimismo, sobre la base de que es posible identificar las habilidades en muchos niveles de detalle, Buck también sugiere que “they may not be something we have within us, but may just be something we do with language” (51).

Aunque definir el constructo del test en términos de microhabilidades plantea ciertas dudas, es necesario especificarlo. El test mide tres capacidades:

1. Extraer datos concretos importantes en conversaciones informales cara a cara pertenecientes a comedias románticas cinematográficas españolas.
2. Identificar las ideas generales en conversaciones informales cara a cara procedentes de comedias románticas cinematográficas españolas.
3. Reconocer sentimientos en conversaciones informales cara a cara de comedias románticas cinematográficas españolas.

La comprensión de información específica es una habilidad contemplada en varias taxonomías (Rost, 1990; Weir, 1993; Buck y Tatsuoka, 1998). Ocurre lo mismo con la comprensión de ideas generales (Richards, 1983; Rost, 1990; Weir, 1993). Ambas se recogen en el apartado de operaciones cognitivas de la parrilla de comprensión auditiva del *Dutch Common European Framework Construct Project*<sup>11</sup>. Asimismo, son evaluadas en test de comprensión auditiva que emplean textos dialógicos (DIALANG y los DELE) y en la prueba de comprensión audiovisual del Diploma LETRA. Cabe decir, que el ULM del examen LETRA guarda algunas similitudes con el ULM de este estudio: en ambos se trata de comprender conversaciones en las que se puede ver y escuchar a los interlocutores pero en las que no se puede intervenir. También hay diferencias importantes, entre otras, que el ULM del test de este trabajo se limita a conversaciones informales del ámbito personal, mientras que el ULM del examen LETRA incluye tanto conversaciones informales del ámbito personal como conversaciones transaccionales de otros ámbitos.

---

<sup>10</sup> “Skills can be thought of as the different cognitive processes involved in understanding language” (Richards y Burns, 2012: 19).

<sup>11</sup> El *Dutch Construct Project* ha desarrollado una parrilla para analizar test de comprensión auditiva y de lectura, con el propósito de relacionarlos con el *MCER*. La parrilla es accesible en la siguiente dirección: <http://www.lancaster.ac.uk/fss/projects/grid/> (consulta: 01/03/2016).

Se ha considerado que la capacidad de reconocer sentimientos juega un papel importante en el ULM, es decir, en la comprensión de conversaciones informales del ámbito personal de comedias románticas, y por ello, que era necesario medirla. Esta habilidad no aparece referenciada como tal en las taxonomías consultadas, pero está estrechamente relacionada con la capacidad de hacer inferencias señalada en varias propuestas (Richards, 1983; Rost, 1990; Weir, 1993; Buck y Tatsouka, 1998) y la habilidad de utilizar “facial, paralinguistic, and other cues to work out meaning” (Richards, 1983: 228). Además, forma parte del constructo de algunos test de comprensión auditiva del español y del inglés como lenguas extranjeras ampliamente extendidos: en el DELE de nivel C1 se mide la comprensión de sentimientos, actitudes y emociones en conversaciones informales y transaccionales, y en el *Cambridge English: First* (nivel B2) se evalúa la comprensión de sentimientos y actitudes en textos monológicos y dialógicos. La parrilla de comprensión auditiva del *Dutch Common European Framework Construct Project* no menciona específicamente la comprensión de sentimientos pero sí contempla dos capacidades próximas: la comprensión de la actitud y el estado de ánimo del hablante.

Aunque las dos primeras capacidades evaluadas se podrían asociar con la descodificación y la tercera con la inferencia, en la definición del constructo no se ha establecido tal relación porque, según Sperber y Wilson (1986, 2004), incluso el contenido comunicado de forma explícita a través de un enunciado va más allá de lo codificado lingüísticamente. Según estos autores, la información comunicada explícitamente se recupera a través de un proceso que combina la descodificación y la inferencia a través de la desambiguación, la asignación de referente y el enriquecimiento. La desambiguación consiste en seleccionar la acepción más adecuada de un término polisémico de acuerdo con el contexto. Escandell Vidal (1996) pone el ejemplo del enunciado “¿Dónde está el gato?”. Si tiene lugar dentro de una vivienda probablemente se referirá a un animal doméstico. En cambio, si se produce en medio de la carretera, es muy probable que se refiera al instrumento que sirve para levantar el coche. La asignación de referente supone la identificación de la entidad a la que se refiere el enunciado con la ayuda de la información contextual. Por ejemplo, al escuchar el enunciado “déjame el periódico” el oyente debe decidir cuál es el periódico al que se refiere el hablante. A través del enriquecimiento el oyente completa los términos expresados de forma imprecisa. Verbigracia: en el enunciado “las polainas y el piolet son demasiados caros” el oyente tiene que enriquecer el adjetivo “caro” de forma diferente en función de los referentes (polainas y piolet) y del contexto, ¿son caros para quién?, ¿son caros para qué?

### 2. 1. 5. LONGITUD Y DURACIÓN DEL TEST

Se diseña un test con siete ítems y dos textos de entrada. Esta extensión se considera adecuada por dos razones. En primer lugar, se considera suficiente para los objetivos de investigación (v. § 2. 1. 1.) y para medir las tres habilidades del constructo (v. § 2. 1. 4.) en un ámbito de uso de la lengua muy específico (v. § 2. 1. 3.)<sup>12</sup>. En segundo lugar, la

---

<sup>12</sup> Algunas pruebas de comprensión audiovisual consultadas también tienen un número de ítems reducido. El test de comprensión audiovisual del Diploma LETRA mide tres habilidades en varios ámbitos y el modelo disponible en internet está compuesto por tres tareas y once ítems (v. § 1. 7.). Wagner (2002) describe la elaboración de una prueba con vídeo que evalúa seis capacidades dentro del ámbito académico. La versión final de este test cuenta con trece ítems.

extensión del test facilita su administración en términos de tiempo y exigencias cognitivas junto con los otros dos instrumentos de investigación: el cuestionario previo y el cuestionario posterior (v. § 1.8.). En términos de tiempo, la concisión de los instrumentos posibilita que puedan completarse en una sola sesión y, en términos de exigencias cognitivas, la brevedad contribuye a que no se genere un cansancio excesivo que comprometa los resultados. Es posible separar la administración del cuestionario previo de la de los otros instrumentos, pero esto puede complicar la administración porque es necesario reunir a los informantes en dos ocasiones.

Los examinandos disponen de todo el tiempo que deseen para completar los cuestionarios y la prueba. En el estudio 5 se contabilizó el tiempo empleado por el grupo de control y los dos grupos experimentales para completar los instrumentos. El tiempo medio empleado por los grupos para rellenar una versión adaptada del cuestionario previo fue siete minutos y treinta segundos. La administración del test tradicional tuvo una duración de doce minutos y veintisiete segundos, y la del test experimental y el cuestionario posterior una media de diecinueve minutos y cincuenta y ocho segundos. Estos tiempos sirven de orientación para planificar los estudios cuasiexperimentales.

#### 2.1.6. NIVEL DE LOS ÍTEMS

La prueba está compuesta por siete ítems: tres miden la primera habilidad del constructo, dos la segunda y dos la tercera. Además de evaluar la primera capacidad, el ítem inicial tiene otros dos objetivos: servir como ejemplo del funcionamiento de las preguntas subtituladas en la variante experimental del test, por un lado, y generar confianza en los informantes de los grupos de control y tratamiento, por otro. Para cumplir estas dos últimas funciones, es conveniente que la pregunta inicial sea fácil o muy fácil.

Para estudiar algunos aspectos, es necesario que cada habilidad sea medida por al menos dos ítems de distinto nivel. Los resultados del estudio 6 sugieren que las tres preguntas que miden la primera habilidad son de dos niveles diferentes y que los ítems que evalúan las otras capacidades también lo son (v. § 2.3.6.4.). El nivel aproximado de las preguntas según el *MCER* se encuentra en la tabla 23.

#### 2.1.7. TIPO DE ÍTEMS

El test contempla tres tipos de ítems: elección múltiple, respuesta corta y respuesta de extensión media<sup>13</sup>. Bordón (2006) nombra algunas ventajas de las preguntas de respuesta preseleccionada, entre ellas, que se puede utilizar con textos muy variados (entre los que nombra los dialógicos), que se puede adaptar a varios niveles, que es fácil de corregir, que proporciona resultados fiables y que resulta aceptable para la mayor parte de los estudiantes. Este formato no está libre de inconvenientes. Ur (1999: 39) considera que el proceso de comprensión de la pregunta y las opciones escritas “may take more time and effort than the point ostensibly tested”, Hughes (1989) advierte que es difícil estimar el peso de las preguntas acertadas al azar en las puntuaciones de cada individuo. Bordón (2006) destaca que se trata de una tarea no auténtica, y varios autores señalan la

---

<sup>13</sup> La descripción de las preguntas (incluyendo el tipo de ítems) se encuentra en el § 2.3.5.3.2.

complejidad de redactar preguntas de elección múltiple para que no resulten ambiguas o con respuestas evidentes (Hughes, 1989; Alderson et al., 1995; Ur, 1999; Buck, 2001). Con el propósito de minimizar el impacto del azar, se ha seguido la recomendación de Alderson et al. (1995) de incluir un mínimo de cuatro opciones en cada pregunta. Al mismo tiempo, las alternativas se han limitado a cuatro opciones breves para reducir el impacto de la lectura en la memoria de trabajo y en la atención mientras se ve y escucha el vídeo. En términos de capacidades evaluadas, las preguntas de elección múltiple se emplean para medir la comprensión de datos concretos y la identificación de ideas generales. No se utilizan para medir la comprensión de sentimientos porque “dubious answers are particularly common in inferencing questions” (Alderson et al., 1995: 47).

Las preguntas de respuesta corta ofrecen algunas ventajas sobre las preguntas de elección múltiple: el azar tiene menos influencia en la respuesta, no están limitadas por la necesidad de distractores y son más fáciles de redactar (Pastor Villalba, 2009). Su mayor inconveniente consiste en determinar qué se considera una respuesta correcta (Buck, 2001). Como en el test solo se contempla el uso de ítems de respuesta única, esta dificultad no se plantea. En términos de habilidades evaluadas, estos ítems se destinan a la comprensión de datos concretos, lo que casa con la observación de Buck (2001) y Pastor Villalba (2009) de que el formato es muy adecuado para la comprensión de información comunicada explícitamente.

Las preguntas de respuesta larga permiten “ir más allá de la superficie del texto” (Pastor Villalba, 2009: 19) y, por este motivo, se han considerado especialmente apropiadas para medir la comprensión de sentimientos. Comparten con las preguntas de respuesta corta la facilidad de redacción y la dificultad de corrección. En relación con la cuestión de la corrección, Buck (2001: 140) identifica dos problemas potenciales: “firstly, determining what constitutes a reasonable interpretation of the text, and secondly, what constitutes a sufficient response to the question”. Las decisiones tomadas para evaluar las respuestas se describen en el § 2.1.8.

#### 2.1.8. PUNTUACIÓN DE LOS ÍTEMS

Todos los ítems se puntúan de forma dicotómica. En el caso de las preguntas de elección múltiple y los ítems de respuesta única, la calificación no plantea complicaciones porque solo existe una respuesta correcta. Para puntuar las dos preguntas de respuesta de extensión media, se utiliza la estrategia propuesta por Buck (2001: 141): administrar el test a un grupo de oyentes competentes y utilizar sus respuestas como base para estimar la aceptabilidad de las respuestas de los examinandos (v. § 2.3.1.).

#### 2.1.9. OTRAS CONSIDERACIONES

En este apartado se detallan algunas decisiones tomadas en relación con la lengua empleada en la redacción del test, el momento en que se entrega el test y la lengua de la respuesta. Asimismo, se indica cómo se ha intentado limitar la complejidad de escuchar, ver, leer y escribir simultáneamente.

Hughes (1989) señala que la comprensión de los ítems no debe suponer una dificultad para los examinandos y sugiere el uso de la lengua materna si todos los candidatos la

comparten. En el caso que nos ocupa, el test se redacta en la lengua meta porque no todos los participantes en la investigación tienen la misma lengua materna. Con el propósito de identificar e intentar resolver las dudas que pudiera plantear el test, antes del visionado se pide a los examinandos que lo lean y que, o bien confirmen la comprensión, o bien manifiesten sus dudas para intentar solucionarlas.

Como se acaba de mencionar, el test se entrega antes del visionado. Sherman (1997) sugiere que las preguntas tienen poca significación por sí solas, y que solo cobran sentido cuando se escucha el texto. No obstante, la entrega de las preguntas antes del visionado comporta algunos beneficios. Por una parte, se respeta las preferencias de los examinandos: los estudiantes valoran favorablemente poder leer las preguntas antes de la audición (Sherman, 1997; Buck, 2001). Por otra parte, posibilita la aclaración de dudas mencionada en el párrafo anterior. En el caso de la versión experimental del test, cabe reconocer que desconocemos en qué medida la coexistencia de los dos planos de trabajo (en papel y en pantalla) podría mantener el conflicto de atención visual que precisamente se quiere reducir con la subtitulación de las preguntas.

El test incluye ítems de respuesta abierta y es pertinente considerar en qué lengua o lenguas se pueden contestar. De acuerdo con Hughes (1989) y Buck (2001), la respuesta puede redactarse en la lengua materna o en la lengua meta. Para evitar que el nivel de producción escrita en la L2 condicione la respuesta a los ítems abiertos, en el test se ofrece la posibilidad de expresarse tanto en la L2 y como en una serie de lenguas comprensibles para el investigador o para personas que puedan ayudarlo. Estos idiomas se definen en función de las lenguas maternas de los estudiantes de ELE que vayan a participar en las pruebas y se especifican en el test. Cuando la respuesta se elabora en la L2, estamos de acuerdo con la sugerencia de Buck (2001: 126) de no penalizar errores si la respuesta es comprensible: "it seems reasonable not to penalise mistakes if the response is intelligible and clear. After all, we want to know whether test-takers have understood a text, not whether they can produce correct language".

En el §1.5. se ha descrito la complejidad de escuchar, ver, leer y escribir simultáneamente. Para limitar esta complejidad se intenta formular ítems breves y concisos. Asimismo, se sigue la recomendación de Alderson et al. (1995) de que los ítems sean independientes unos de otros, es decir, que la respuesta a un ítem no esté condicionada por la contestación a otro. Igualmente, se tienen en cuenta algunas sugerencias de Weir (1993: 112) relacionadas con las preguntas: "they should normally follow in the sequence in which the information appears in the discourse, and there are must be sufficient spacing between them".

## 2.2. DISEÑO

### 2.2.1. ESPECIFICACIONES DE LOS TEXTOS DE ENTRADA

El *Dutch Common European Framework Construct Project* ha elaborado una parrilla constituida por quince elementos para analizar los textos de entrada de la comprensión auditiva<sup>14</sup>. La parrilla también puede ser utilizada para planificar los test (Consejo de Europa, 2009). Para describir las características de los textos se han tenido en consideración estos quince parámetros. Además, se han incluido dos campos adicionales: calidad de la grabación y explotaciones didácticas de los textos para la enseñanza del ELE.

Tabla 2. Especificaciones de los textos de entrada

1. Fuente	Videotexto (fuente: <i>MCER</i> ) Secuencia cinematográfica de la comedia <i>Ópera prima</i> de Fernando Trueba en la que tiene lugar una conversación informal. Localización: de 00:00:00:00 a 00:02:36:12 <sup>15</sup> .	Videotexto (fuente: <i>MCER</i> ) Secuencia cinematográfica de la comedia <i>Los peores años de nuestra vida</i> de Emilio Martínez Lázaro en la que tiene lugar una conversación informal. Localización: de 00:11:05:00 a 00:12:43:16.
2. Autenticidad	Los textos tienen autenticidad de origen porque han sido producidos por y para hablantes de una L1 determinada. El discurso oral tiene la peculiaridad de que está planificado: es un discurso escrito para ser recitado por actores. También tiene otra característica: los diálogos se han redactado para parecer orales. Sin embargo, esta oralidad planificada no llega a ser totalmente genuina. Con relación a los géneros de ficción, Chaume (2001: 80) señala que: los recursos de coloquialidad encuentran su límite en la supresión de digresiones, redundancias, hipérbatos, anacolutos y dubitaciones que cansarían al espectador y que, probablemente, poco cooperarían a cumplir con las convenciones formales de longitud, inmediatez y relevancia lingüística establecidas en este género de textos.	
3. Tipo de discurso predominante (fuente: DIALANG)	Fático	Fático y argumentativo
4. Ámbito predominante (fuente: <i>MCER</i> )	Personal	Personal
5. Tema y contenido (fuente: <i>MCER</i> )	Identificación personal Casa y hogar Educación Relaciones con otras personas	Identificación personal Relaciones con otras personas Educación Otro: celebraciones (la Navidad)

<sup>14</sup> El enlace a la parrilla se indica en la nota número once.

<sup>15</sup> Para indicar la secuencia se ha utilizado un código de tiempo de ocho dígitos habitual en la grabación y edición de vídeo. Los tres primeros pares de dígitos corresponden a las horas, los minutos y los segundos, respectivamente, y el último par, a los fotogramas. Así, 00:02:36:12 indica el punto 2 minutos, 36 segundos y 12 fotogramas de la película *Ópera prima*.

6. Naturaleza del contenido	Principalmente contenido concreto	Contenido ligeramente abstracto
7. Duración del texto	2 min. y 36 s.	1 min. y 38 s.
8. Vocabulario	Principalmente vocabulario frecuente	Vocabulario bastante extenso
9. Gramática	Principalmente estructuras sencillas	Un número limitado de estructuras complejas.
10. Velocidad del texto (impresión global)	Normal	Rápida
11. Número de participantes	2	4
12. Acento de los participantes	Estándar	Estándar
13. Claridad de la articulación (impresión global)	La articulación es normal. El audio está grabado con sonido directo y se aprecia ligeramente el tránsito durante el diálogo.	La articulación es normal. El audio está grabado con sonido directo y se escucha ligeramente el tránsito en la parte del diálogo que transcurre en la calle.
14. Número de visionados	Uno. Buck (2001) analiza la cuestión del número de reproducciones y expresa su preferencia por una sola audición cuando sea posible. El autor considera que es lo más lógico desde el punto de vista de la autenticidad situacional e interaccional porque en casi todas situaciones fuera del aula los textos se escuchan sola una vez. En el caso que nos ocupa, el ámbito de uso de la lengua meta implica la comprensión de textos procedentes de largometrajes, y por lo tanto, textos que se suelen ver y escuchar solo una vez fuera del aula. Además, Buck (2001: 171) señala que el procesamiento de forma rápida y automática, por un lado, y “the ability to make inferences to bridge comprehension gaps”, por otro, son habilidades básicas de la comprensión que tienen que ser evaluadas.	
15. Nivel del texto	A2 / B1 (v. § 2.3.6.)	B2 / C1 (v. § 2.3.6.)
16. Calidad de la grabación	Se ha conservado la calidad de las ediciones en DVD de las películas en las dos versiones del test.	
17. Explotaciones didácticas de los textos	Casañ Núñez (2015a). Niveles A2-B2 en función de la selección de actividades.	Rabasco Macías (2007). Nivel B1. Santos Gargallo, I. y Santos Gargallo, A. (2001). Nivel intermedio-avanzado. Vizcaino Rogado (2007). Nivel intermedio.

## 2.2.2 INDICACIONES PARA LOS EXAMINANDOS

En este apartado se detallan las indicaciones que reciben los informantes durante la prueba. En todos los casos, al final de la explicación se pide a los informantes que confirmen su comprensión o expresen sus dudas para intentar resolverlas.

### 2.2.2.1 PRESENTACIÓN DE LA ESTRUCTURA GENERAL DEL TEST

Antes de introducir el primer texto de entrada, se expone la estructura general del test. Se menciona que la prueba está compuesta por dos secuencias cinematográficas de dos películas españolas y por dos tareas breves. Se añade que los vídeos se observarán una sola vez porque así sucede normalmente cuando se ve cine fuera del aula, y porque el objetivo es medir lo que se comprende después de un solo visionado. Además, en el caso del grupo experimental, se anticipa que después de cada tarea se completará un breve cuestionario.

### 2.2.2.2 INTRODUCCIÓN DE LOS TEXTOS DE ENTRADA

El hecho de que los examinandos se enfrenten a los textos sin haber recibido ningún tipo de conocimiento previo puede suponer una amenaza a la validez, ya que esa situación normalmente no es representativa del ámbito de uso de la lengua meta (Wagner, 2013). Para aproximar en cierta medida las condiciones de visionado a las habituales fuera del aula, antes de entregar cada tarea se proporciona a los estudiantes parte del conocimiento que podría tener el público sobre cada película.

Los textos se introducen sucintamente de forma oral. En relación con *Ópera prima*, se menciona que es una comedia española de 1980 dirigida por Fernando Trueba. Se añade que el director fue galardonado con un Óscar de Hollywood a la mejor película extranjera por *Belle Époque*. Asimismo, se explica el significado de *ópera prima* y se comenta que el título es un juego de palabras porque este fue su primer largometraje<sup>16</sup>. Por último, se señala que el texto es la primera escena de la película y que dura aproximadamente dos minutos.

En cuanto a *Los peores años de nuestra vida*, se menciona que es una comedia española de 1994 dirigida por Emilio Martínez-Lázaro. A continuación se nombra una de las producciones más conocidas del realizador. En el caso de la investigación posterior al estreno de *Ocho apellidos vascos* en marzo de 2014, se hace referencia a esta película. En el caso de estudios anteriores, se menciona *El otro lado de la cama*. Por último, se añade que el texto es una de las primeras escenas de la película y que dura aproximadamente dos minutos.

### 2.2.2.3 EXPLICACIÓN DEL FUNCIONAMIENTO DE LAS PREGUNTAS SUBTITULADAS

El grupo de tratamiento realiza la prueba en su variante experimental. Por este motivo, es necesario indicarles la existencia de las preguntas de comprensión subtituladas y especificar su temporalización. Con este fin se explica que el vídeo está encajado entre dos franjas

---

<sup>16</sup> El título tiene otros niveles de significado (la primera escena comienza en la boca de metro Ópera y transcurre en la Plaza de la Ópera de Madrid, y el protagonista se enamora de su prima), pero no se comentan para no revelar aspectos de la trama.

negras y que los subtítulos se sitúan en la franja inferior. Esta exposición oral se acompaña de un dibujo en la pizarra o de una imagen para mayor claridad (v. la figura 2). Asimismo, se comenta que las preguntas son visibles pocos segundos antes del fragmento relevante para responderlas y que desaparecen cuando finaliza dicho fragmento. A continuación se precisa que existe una excepción: la última cuestión de cada tarea se presenta durante los últimos segundos del vídeo y es de comprensión global. Esta explicación tiene lugar después de que los informantes hayan leído los ítems de la tarea 1 y antes de ver la primera secuencia.



Figura 2. Imagen para apoyar la explicación sobre las preguntas subtituladas.

### 2.2.3. PRIMERA FORMULACIÓN DE LOS ÍTEMS

La primera redacción de los ítems se desarrolló en dos etapas. En un primer momento, se identificaron los contenidos de las secuencias cinematográficas que podrían ser objeto de un test de comprensión audiovisual a partir de la experiencia con los textos en el aula de ELE y en trabajos anteriores (v. Casañ Núñez, 2009a, 2009b), la consulta de explotaciones didácticas de los textos (v. la tabla 2 en el § 2.2.1.), y la estrategia conocida como *mind mapping*. Esta técnica permite extraer las ideas principales y los detalles importantes de los textos, y consiste en escuchar/ver el texto una vez y tomar notas de lo que se considera más importante (Weir, 1993). El *mind mapping* resultó especialmente pertinente porque dos habilidades evaluadas consisten en la comprensión de ideas principales y de datos concretos relevantes. En una segunda fase, se elaboraron ítems teniendo en consideración las ideas identificadas y las especificaciones del test.

### 2.3. EXPERIMENTACIÓN

En esta sección se exponen los seis estudios que se efectuaron para alcanzar una versión satisfactoria del test por orden temporal. Algunos de estos estudios también tuvieron otros objetivos, entre ellos, desarrollar el cuestionario previo y el cuestionario posterior (v. § 1.8.). Dado que el presente artículo está focalizado en la construcción del test, solo se detallan los aspectos relacionados la prueba. Asimismo, por una cuestión de espacio y relevancia, se describen en profundidad los estudios 2, 3, 5 y 6 y se resumen los restantes. En la tabla 3 se mencionan los propósitos principales de los estudios en relación con el test.

Tabla 3. Objetivos fundamentales de los estudios con relación al test.

ESTUDIO	OBJETIVOS	PARTICIPANTES
1	Pilotar la primera versión de los ítems. Recoger respuestas de oyentes competentes a los ítems abiertos con el fin de utilizarlas como base para determinar qué se considerarían contestaciones aceptables.	Cinco profesores de ELE, una especialista en la enseñanza del ELE y tres estudiantes de un máster en formación de profesores de portugués y ELE.
2	Pilotar la segunda tarea del test en sus dos versiones.	Veinticinco estudiantes de ELE (quince en el grupo de control y diez en el de tratamiento).
3	Pilotar la primera tarea del test en sus dos versiones.	Treinta y dos estudiantes de ELE (dieciocho en el grupo de control y catorce en el de tratamiento).
4	1ª fase: Averiguar si la versión tradicional del test era factible para oyentes competentes. 2ª fase: Averiguar si la versión tradicional del test era factible para oyentes competentes y conocer el nivel aproximado de los ítems.	Dos especialistas en la enseñanza del ELE.  Cinco profesores de ELE.
5	Pilotaje del test en sus dos variantes.	Cuarenta y un estudiantes de ELE (veintidós en el grupo de control y diecinueve en el de tratamiento).
6	Determinar el nivel aproximado de los textos y de los ítems.	Cuatro profesores de ELE del Instituto Cervantes, cuatro profesores de ELE universitarios y un redactor de ítems.

### 2.3.1. ESTUDIO 1

En este estudio nueve personas especializadas o con formación en ELE respondieron a la primera versión de las preguntas de comprensión audiovisual (v. § 2.2.3.) en condiciones similares a las de la prueba en su versión tradicional, y completaron un cuestionario después de ver cada texto de entrada. La primera tarea tenía dos propósitos: por un lado, conocer el comportamiento de los ítems de elección múltiple y respuesta única, y por otro, recoger contestaciones de oyentes competentes a los ítems abiertos con el fin de utilizarlas como base para determinar qué se considerarían respuestas aceptables. Los cuestionarios preguntaban a los informantes si habían tenido dificultades para contestar y si eliminarían, modificarían o añadirían algún ítem; asimismo, contaban con un espacio para añadir cualquier comentario relacionado con el test.

Participaron los profesores de ELE Carmen González, Adriana Repila, Estrella Redondo, Vanessa Roca y Francisco Teruel (Centro de Idiomas de la Universidad de Valencia), la Prof<sup>a</sup>. Dr<sup>a</sup> Virginia González (Universidad de Valencia) y las estudiantes del máster en formación de profesores de portugués y ELE de la Universidad de Coímbra Inês Machado, Sandra Santos y Verónica Marôco. Debido a un imprevisto, las estudiantes solo tuvieron tiempo para responder a las preguntas y al cuestionario relacionados con la secuencia cinematográfica de *Ópera prima*. Las colaboraciones tuvieron lugar en los respectivos centros de trabajo o formación de los participantes, es decir, en la Universidad de Valencia y en la Universidad de Coímbra, en mayo de 2013. Los informantes contribuyeron a la mejora del test dado que identificaron posibles fuentes de confusión, sugirieron mejoras

en la redacción de las preguntas y nuevos ítems, y aportaron un conjunto de respuestas a las preguntas abiertas que serían empleadas como referente para corregirlas. Cabe reconocer que el estudio fue limitado por el reducido número de personas, especialmente en el caso de la secuencia de *Los peores años de nuestra vida*.

## 2.3.2. ESTUDIO 2

### 2.3.2.1. OBJETIVOS

Se pretendía experimentar todo el proceso de administración y corrección de la segunda tarea del test con los propósitos de averiguar si las indicaciones iniciales sobre la estructura de la prueba (v. § 2.3.2.3.3.), la presentación del segundo texto de entrada (v. § 2.2.2.2.), la explicación sobre el funcionamiento de las preguntas subtituladas (v. § 2.2.2.3.) y la redacción de los ítems de la segunda tarea resultaban claras (v. anexo 1). Además, se quería saber si las respuestas de oyentes competentes a la pregunta abierta obtenidas en el estudio 1 servían como base para determinar la aceptabilidad de las respuestas de los participantes al ítem abierto. Adicionalmente, se perseguía conocer si el índice de facilidad de cada ítem era plausible teniendo en cuenta la dificultad del ítem y el nivel de comprensión auditiva / audiovisual de los informantes, y se quería pilotar el cuestionario posterior (v. § 1.8., este aspecto se describirá en otro trabajo). Por último, se pretendía detectar problemas no anticipados.

### 2.3.2.2. DISEÑO DE INVESTIGACIÓN

Se hizo un ensayo parcial del futuro estudio cuasiexperimental con grupos de control y tratamiento no equivalentes (v. § 1.8. y 2.1.1.). El estudio fue parcial porque solo se pilotó la segunda tarea y porque no se administró el cuestionario previo (v. § 1.8.). Para estimar de forma indirecta y limitada el grado de semejanza de los grupos se compararon las puntuaciones en el examen de expresión e interacción orales de la asignatura con un test *t*. El grupo de control efectuó la prueba con los ítems tradicionales y el grupo de tratamiento con los experimentales. Este último grupo, además, completó el cuestionario posterior (v. § 1.8.). Después de la prueba, se corrigió el test con los informantes y se explotó el texto con fines didácticos.

### 2.3.2.3. MÉTODO

#### 2.3.2.3.1. PARTICIPANTES

Participaron veinticinco estudiantes de la asignatura de ELE de nivel B2 / B2+ *Español 5* ofrecida por la Facultad de Letras de la Universidad de Coímbra en el curso 2013/2014. Se trataba de una disciplina para estudiantes de tercer año de Lenguas Modernas y sus dos turnos eran impartidos por el autor de este trabajo. Este hecho favorecía que los estudiantes se tomaran la prueba en serio, requisito necesario para que los resultados del pilotaje fueran válidos. No se administró un pretest de comprensión audiovisual para determinar el nivel de los participantes porque se desconoce la existencia de pruebas de este tipo. Por los textos empleados en clase (audiciones y vídeos de manuales de ELE, canciones, cortometrajes, fragmentos de películas y documentales y un largometraje) y el tipo de tareas, el nivel de comprensión auditiva / audiovisual se estimaba en torno al B2+/C1.

Se consideró que la asignatura y los participantes eran adecuados por diversos motivos. Primero, el cine español formaba parte del programa y se explotaban regularmente cortometrajes y fragmentos de películas. Segundo, los informantes constituían una muestra representativa de los destinatarios del test en la medida que eran estudiantes de ELE adultos con un nivel de comprensión audiovisual en torno al B2+/C1 y estaban habituados a ver cine y televisión con subtítulos (V. § 2.1.1.).

Aleatoriamente, un turno fue designado grupo de control y el otro de tratamiento. El grupo de control estaba compuesto por trece estudiantes lusófonos y dos anglófonos, y el grupo experimental estaba formado por seis informantes lusófonos, tres anglófonos y uno hispanohablante. Los participantes no lusófonos eran alumnos de intercambio.

Para estimar de forma indirecta y limitada el grado de equivalencia de los grupos se contrastaron las notas de los informantes en el examen de expresión e interacción orales de la asignatura. La prueba tuvo lugar los días 9 y 11 de diciembre de 2013 (dos semanas después del test de comprensión audiovisual) y se puntuó de 0 a 20. Todos los análisis estadísticos se hicieron con la versión 21 de SPSS para Windows, excepto el tamaño del efecto, que se computó en una calculadora. Los descriptivos básicos se recogen en la tabla 4.

Tabla 4. Descriptivos básicos de la prueba de expresión e interacción orales del estudio 2.

GRUPO		N	MÍNIMO	MÁXIMO	MEDIA	ERROR TÍP. MEDIA	DESV. TÍP.
Control	Prueba	15	11	19	14.80	.545	2.111
Experimental	Prueba	10	13	19	16.00	.596	1.886

El grupo experimental obtuvo una media ligeramente superior al de control (16 frente a 14.8). Para averiguar si la diferencia era estadísticamente significativa se llevó a cabo un test *t* para muestras independientes. Esta prueba requiere que las muestras inferiores a treinta casos tengan distribución normal (Pestana y Gageiro, 2008). Esta exigencia se comprobó con el test de Shapiro-Wilk<sup>17</sup>. Los resultados de test de normalidad indicaron que tanto el grupo de control, *S-W* = .961, *g* = 15, *p* = .710, como el de tratamiento, *S-W* = .954, *g* = 10, *p* = .711, presentaban distribuciones que no diferían significativamente de las normales. El test *t* mostró que la diferencia entre las medias del examen no era significativa: *t* = -1.451, *g* = 23, *p* = .160, *r* = .289. Esto constituía un indicio de que había, al menos, cierto grado de equivalencia entre los grupos.

### 2.3.2.3.2. INSTRUMENTO

La segunda tarea de comprensión audiovisual tenía dos variantes: una tradicional para el grupo de control y otra experimental para el grupo de tratamiento. En ambos casos la tarea se entregaba en papel (v. anexo 1). En la versión experimental, además, los enunciados de los ítems aparecían integrados en la imagen como si fueran subtítulos (v. figura 3). Para inscribir los enunciados se aplicaron los criterios formulados por Casañ

<sup>17</sup> Se optó por la prueba de normalidad Shapiro-Wilk porque es la más adecuada para muestras inferiores a cincuenta casos (Sen y Srivastava, 1990).

Núñez (2015b: 16-24). El programa informático utilizado para editar la secuencia fue Adobe Premiere Pro CS5.5 para Windows. Las especificaciones del texto de entrada se encuentran en el § 2.2.1. y en la tabla 5 se resumen las características principales de los ítems experimentales. Los ítems tradicionales compartían con los experimentales las categorías *enunciado*, *focalización*, *tipoy nivel estimado por el autor*.

Tabla 5. Descripción de las preguntas subtituladas de la segunda tarea (secuencia de *Los peores años de nuestra vida*).

ENUNCIADO	FOCALIZACIÓN	TIPO	TEMPORALIZACIÓN	N.º DE LÍNEAS	FUENTE Y TAMAÑO	COLOR	NIVEL ESTIMADO POR EL AUTOR
1. ¿De qué temas habla el chico pelirrojo?	Extraer datos concretos	Elección múltiple	00:11:40:00 00:12:25:24	una	Times New Roman 36	Blanco (código hex. #FFFFFF)	B1+ / B2
2. ¿Cómo se llama la chica?	Identificar las ideas generales	Elección múltiple	00:12:39:01 00:12:43:24	una			A1 / A2
3. ¿Qué sentimientos puede tener el chico pelirrojo por la chica?	Reconocer sentimientos	Respuesta abierta	00:12:45:00 00:12:48:15	dos			B1+ / B2



Figura 3. Fotogramas de *Los peores años de nuestra vida* con preguntas de comprensión audiovisual subtituladas.

### 2.3.2.3.3. PROCEDIMIENTO

El test fue administrado por el investigador a los grupos de control y tratamiento el 27 de noviembre de 2013 en la misma aula durante los respectivos horarios lectivos. Para reproducir el vídeo se empleó el equipo técnico que se utilizaba habitualmente en esa clase: pantalla de proyección tripié, videoprojector portátil, ordenador portátil y dos altavoces. La administración fue análoga en ambos grupos, con excepción de las diferencias necesarias para llevar a cabo el estudio.

Primeramente, se expuso brevemente la estructura de la prueba. Se comentó que el test estaba formada por una secuencia cinematográfica y una tarea breves. También se indicó que la secuencia se observaría una sola vez porque así sucede normalmente cuando se ve cine fuera del aula, y porque el objetivo era medir lo que se comprendía después de una sola visualización. En el grupo experimental, además, se mencionó que después de la tarea se completaría un cuestionario. Antes de proseguir, se preguntó a los informantes si había dudas hasta el momento.

A continuación se añadió que la prueba se corregiría una vez finalizada y que el vídeo se explotaría con fines de aprendizaje. Luego se introdujo oralmente el texto de entrada (v. § 2.2.2.2.) y se pidió a los estudiantes que confirmaran la comprensión o expresaran sus dificultades. Después se repartió la prueba (v. anexo 1), se solicitó a los participantes que la leyeran y se preguntó si había algún contenido que precisara aclaración. En relación con el ítem de respuesta abierta, se consideró pertinente justificar la posibilidad de escribir en la L1 porque el objetivo de la prueba era evaluar la comprensión y no la producción. Además, en el caso del grupo experimental, se anunció la existencia de las preguntas subtituladas en el vídeo y se explicó su temporalización (v. § 2.2.2.3.). Sobre este último aspecto se pidió a los participantes que indicaran si había alguna duda. Después del visionado, se esperó a que todos los informantes hubieran completado el test para recogerlo. En el caso del grupo de tratamiento, se repartió y respondió al cuestionario posterior. Luego se corrigió el test con los participantes. Por último, se explotó el fragmento con fines didácticos.

Los ítems fueron puntuados por el autor de forma dicotómica. Para estimar la aceptabilidad del ítem abierto se utilizaron las respuestas de oyentes competentes recogidas en un estudio anterior (v. § 2.3.1.). El autor volvió a corregir los test dos semanas después con el propósito de comprobar la exactitud de las puntuaciones.

### 2.3.2.3.4. ANÁLISIS DE DATOS

Las puntuaciones del test se introdujeron en la versión 21 del programa SPSS para Windows. Una semana después se verificó que los datos insertados eran correctos. Posteriormente, se realizó un análisis de frecuencias para comprobar que no había valores anómalos ni valores perdidos. Por último, se calculó el índice de facilidad de cada ítem.

### 2.3.2.4. RESULTADOS Y DISCUSIÓN

El primer objetivo del estudio era averiguar si las indicaciones iniciales sobre la estructura de la prueba, la presentación del texto de entrada, la explicación sobre el funcionamiento

de las preguntas subtituladas y la redacción de los ítems resultaban claras para los informantes. Para abordar los tres primeros aspectos, se preguntó oralmente a los participantes. Como nadie manifestó tener dudas y algunos examinandos confirmaron la comprensión, se dedujo que no plantearon dificultades. Para averiguar si la redacción de los ítems era clara se articularon dos estrategias: preguntar a los estudiantes si entendían los ítems antes del visionado y corregir el test con los informantes. En la lectura previa de los ítems los participantes no expresaron tener problemas y algunos confirmaron su comprensión, lo que dio a entender que no planteaban problemas. Sin embargo, la corrección del test con los examinandos reveló que estudiantes de ambos grupos no habían interpretado “cena” como la última comida del día, sino que le habían atribuido un sentido posible en portugués (pero no en español) que convertía la primera opción en correcta. Para conocer la profundidad de este lapsus, se decidió contabilizar el número de personas que habían escogido esta alternativa. Se descubrió que el desliz había sido generalizado, puesto que el 86.67 % de los informantes del grupo control y el 70 % de los participantes del grupo de tratamiento eligieron esta opción. Para evitar futuras confusiones, se decidió sustituir “cena de Navidad” por “tiempo atmosférico” en las dos primeras alternativas del ítem 1 (v. anexo 2). En lo que respecta a las otras dos preguntas, hubo acuerdo sobre cuál era la respuesta correcta en el segundo ítem cerrado, y qué constituía una contestación aceptable en el tercer ítem.

Otro objetivo consistía en saber si las respuestas de oyentes competentes al ítem abierto obtenidas anteriormente (v. § 2.3.1.) servían como base para determinar la aceptabilidad de las respuestas de los participantes a esta pregunta. Las experiencias positivas en la corrección de la prueba con los informantes, por un lado, y de los veinticinco test, por otro, sugerían que conformaban un referente para evaluar las preguntas. Una fuente de evidencia adicional, en este caso indirecta, residía en el hecho de que el índice de facilidad de este ítem se encontraba dentro de los valores esperados (v. § siguiente).

En tercer lugar se pretendía conocer si el índice de facilidad de cada ítem era plausible. Se estimaba que los participantes tenían un nivel de comprensión auditiva / audiovisual B2+/C1 y que los ítems 1, 2 y 3 eran de nivel B1+/ B2, A1 / A2 y B1+/ B2, respectivamente. Por esta razón, se preveían índices de facilidad muy elevados, especialmente en el caso del ítem 2. Como se ha comentado, el ítem 1 resultó confuso. En consecuencia, su índice de facilidad fue muy bajo en ambos grupos. Los ítems 2 y 3 tuvieron índices de facilidad conforme a lo esperado. El ítem 2, de nivel A1/A2, no revistió ninguna dificultad para los estudiantes y su índice de facilidad fue del 100 % en ambos grupos. El ítem 3, de nivel B1+/B2, tuvo un índice de facilidad alto pero menor que el ítem 2: 73.3 % en el grupo de control y 80 % en el de tratamiento. El índice de facilidad de los ítems se encuentra en la tabla 6.

Tabla 6. Índice de facilidad de los ítems de la secuencia de *Los peores años de nuestra vida*.

GRUPO	ÍTEM	N	ÍNDICE DE FACILIDAD
Control	1	15	7 %
	2	15	100 %
	3	15	73.3 %
Experimental	1	10	20 %
	2	10	100 %
	3	10	80 %

En resumen, los resultados de esta prueba piloto con los ítems de la secuencia de *Los peores años de nuestra vida* sugirieron que las indicaciones iniciales sobre la estructura de la prueba, la presentación del texto de entrada y la explicación sobre el funcionamiento de las preguntas subtítuladas eran comprensibles, que las respuestas de oyentes competentes al ítem abierto obtenidas en el estudio 1 servían como base para valorar la pregunta de comprensión de sentimientos, que la redacción de dos de los tres ítems era clara, y que dos de los tres ítems tenían índices de facilidad plausibles. El estudio también reveló que existía una fuente de confusión en las opciones del ítem 1, lo que llevó a su reformulación. En el anexo 2 se puede consultar una versión genérica de esta tarea del test.

Cabe reconocer que los resultados fueron limitados. En primer lugar, era posible que la realización previa de la primera tarea hubiera tenido cierto impacto en la segunda. Además, esta influencia podría haber sido diferente en los grupos, dado que una de las funciones del primer ítem experimental de la tarea 1 era servir como ejemplo del funcionamiento de las preguntas subtítuladas para el grupo de tratamiento. En segundo lugar, participaron un número limitado de informantes en cada grupo (quince en el de control y diez en el experimental). Según Alderson et al. (1995: 75), para las pruebas piloto "large numbers are not necessary, but if there are at least 20, so much the better".

### 2.3.3. ESTUDIO 3

Esta prueba piloto imitó y completó al estudio 2 en tanto en cuanto se reprodujo en gran medida la investigación con los ítems del primer texto de entrada.

#### 2.3.3.1. OBJETIVOS

Se pretendía experimentar todo el proceso de administración y corrección de la primera tarea del test para averiguar si las indicaciones iniciales sobre la estructura de la prueba (v. § 2.3.3.3.3.), la presentación del primer texto de entrada (v. § 2.2.2.2.), la explicación sobre el funcionamiento de las preguntas subtítuladas (v. § 2.2.2.3.) y la redacción de los ítems de la primera tarea resultaban claras (v. anexo 3). Asimismo, se quería conocer si las respuestas de oyentes competentes a la pregunta abierta obtenidas en el estudio 1 servían como base para determinar la aceptabilidad de las respuestas de los participantes a este ítem abierto. Además, se perseguía comprobar si el índice de facilidad de cada ítem era plausible teniendo en cuenta la dificultad del ítem y el nivel de comprensión auditiva / audiovisual de los informantes, y se quería continuar el proceso de desarrollo de cuestionario posterior iniciado en el estudio 2 (este aspecto se desarrollará en otro trabajo). Por último, se pretendía detectar problemas no anticipados.

#### 2.3.3.2. DISEÑO DE INVESTIGACIÓN

Se hizo un ensayo parcial del futuro estudio cuasiexperimental con grupos de control y tratamiento no equivalentes (v. § 1.8. y 2.1.1.). El estudio fue parcial porque solo se pilotó la primera tarea y porque no se administró el cuestionario previo contemplado en el futuro cuasiexperimento (v. § 1.8.). Para estimar de forma indirecta y limitada el grado de semejanza de los grupos se compararon las puntuaciones en el examen de expresión e interacción orales de la asignatura con un test *t*. El grupo de control efectuó la prueba con

los ítems tradicionales y el grupo de tratamiento con los experimentales. Este último grupo, además, completó el cuestionario posterior (v. § 1.8.). En la siguiente clase se corrigió el test con los informantes.

### 2.3.3.3. MÉTODO

#### 2.3.3.3.1. PARTICIPANTES

Participaron treinta y dos estudiantes de la asignatura de ELE de nivel B2/B2+ *Español 5* ofrecida por la Facultad de Letras de la Universidad de Coimbra en el curso 2013/2014. Veinticinco de ellos habían colaborado en el primer estudio piloto y los otros siete eran alumnos de la misma asignatura. La descripción de los participantes y su idoneidad se especificaron en el § 2.3.2.3.1. Se mantuvo la distribución de turnos como grupos de control y tratamiento del primer estudio piloto. El grupo de control estaba formado dieciséis informantes lusófonos y dos anglófonos, y el grupo experimental estaba compuesto por diez estudiantes lusófonos, tres anglófonos y uno hispanohablante. Los participantes no lusófonos eran alumnos de intercambio.

Al igual que en el estudio 2, para estimar de forma indirecta y limitada el grado de equivalencia de los grupos se contrastaron las notas de los informantes en el examen de expresión e interacción orales de la asignatura. La prueba tuvo lugar los días 9 y 11 de diciembre de 2013 (una semana después del test de comprensión audiovisual) y se puntuó de 0 a 20. Todos los análisis estadísticos se efectuaron con la versión 21 de SPSS para Windows, a excepción del tamaño del efecto, que se computó en una calculadora. Los descriptivos principales se recogen en la tabla 7.

Tabla 7. Descriptivos básicos de la prueba de expresión e interacción orales del estudio 3.

GRUPO		N	MÍNIMO	MÁXIMO	MEDIA	ERROR TÍP. MEDIA	DESV. TÍP.
Control	Prueba	18	11	19	14.83	.459	1.948
Experimental	Prueba	14	13	19	15.50	.489	1.829

El grupo de tratamiento obtuvo una media ligeramente superior al de control (15.5 frente a 14.83). Para averiguar si esta diferencia era estadísticamente significativa se realizó un test *t* para muestras independientes. Primeramente, se efectuó el test de Shapiro-Wilk para comprobar que las muestras cumplieran el requisito de normalidad del test *t*. Los resultados indicaron que tanto el grupo de control, *S-W* = .957, *gl* = 18, *p* = .548, como el de tratamiento, *S-W* = .951, *gl* = 14, *p* = .582, presentaban distribuciones que no diferían significativamente de las normales. El test *t* mostró que no había diferencias significativas en las puntuaciones medias de los grupos en el examen: *t* = -.986, *gl* = 30, *p* = .332, *r* = .177. Este hecho sugería que había, al menos, cierto grado de equivalencia entre los grupos.

#### 2.3.3.3.2. INSTRUMENTO

La primera tarea de comprensión audiovisual tenía dos versiones: una tradicional para el grupo de control y otra experimental para el grupo de tratamiento. En ambas variantes la tarea se entregaba escrita en un folio (v. anexo 3). En la versión experimental, además, los

enunciados de los ítems aparecían integrados en la imagen como si fueran subtítulos (v. figura 4). Para inscribir los enunciados se aplicaron los criterios formulados por Casañ Núñez (2015b: 16-24). El programa informático utilizado para editar la secuencia fue Adobe Premiere Pro CS5.5 para Windows. La descripción del texto de entrada se encuentra en el § 2.2.1. En la tabla 8 se especifican las características principales de los ítems de la versión experimental. Los ítems tradicionales compartían con los experimentales las categorías *enunciado*, *focalización*, *tipo* y *nivel estimado por el autor*.

Tabla 8. Descripción de las preguntas subtituladas de la primera tarea (secuencia de *Ópera prima*).

ENUNCIADO	FOCALIZACIÓN	TIPO	TEMPORALIZACIÓN	N.º DE LÍNEAS	FUENTE Y TAMAÑO	COLOR	NIVEL ESTIMADO POR EL AUTOR
1. ¿Cómo se llama la chica?	Extraer datos concretos	Elección múltiple	00:01:10:00 00:01:37:00	una	Times New Roman 36	Blanco (código hex. #FFFFFF)	A1
2. ¿De qué hablan?	Identificar las ideas generales	Elección múltiple	00:01:44:00 00:02:02:02	una			A2 / B1
3. ¿Cuál es el teléfono de la chica?	Extraer datos concretos	Respuesta única	00:02:04:04 00:02:22:00	una			A2 / B1
4. ¿Qué sentimientos puede tener el chico por la chica?	Reconocer sentimientos	Respuesta abierta	00:02:27:08 00:02:36:12	dos			A2 / B1

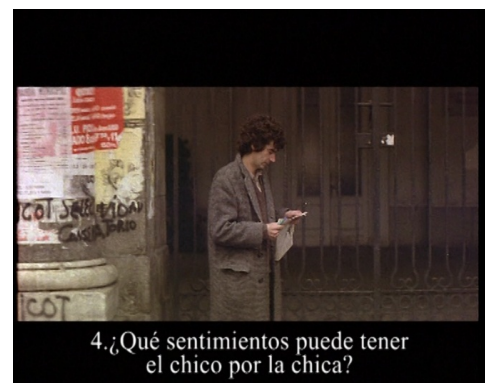


Figura 4. Fotogramas de *Ópera prima* con preguntas de comprensión audiovisual subtituladas.

### 2.3.3.3. PROCEDIMIENTO

La administración guarda muchas similitudes con la del primer estudio parcial. Aunque la gran mayoría de los estudiantes había participado en el primer estudio, se repitieron todos los detalles porque había nuevos informantes en ambos grupos.

La tarea fue administrada por el investigador a los grupos de control y tratamiento el 4 de diciembre de 2013 en la misma aula durante los respectivos horarios lectivos. Para reproducir el vídeo se empleó el equipo técnico que se utilizaba habitualmente en esa clase: pantalla de proyección tripié, videoprojector portátil, ordenador portátil y dos altavoces. Cabe mencionar que se trataba de la misma sala empleada en el primer estudio piloto. El procedimiento fue análogo en ambos grupos, con excepción de las diferencias imprescindibles para efectuar el estudio.

En primer lugar, se describió sucintamente la estructura de la prueba. Se indicó que el test se componía de una secuencia cinematográfica y una tarea breves. También se mencionó que la secuencia se observaría una sola vez porque así sucede normalmente cuando se ve cine fuera del aula, y porque el objetivo era medir lo que se comprendía después de una sola visualización. En el grupo de tratamiento, además, se comentó que después de la tarea se completaría un cuestionario. Antes de continuar, se preguntó a los participantes si había dudas hasta el momento.

Luego se introdujo oralmente el texto de entrada (v. § 2.2.2.2.) y se pidió a los estudiantes que confirmaran la comprensión o expresaran sus dificultades. Después se repartió la prueba (v. anexo 3), se solicitó a los participantes que la examinaran y se preguntó si había algún contenido que precisara aclaración. En relación con el ítem de respuesta larga, se consideró relevante justificar la posibilidad de escribir en la L1 porque la finalidad de la prueba era evaluar la comprensión y no la producción. Además, en el caso del grupo de tratamiento, se informó de la existencia de las preguntas integradas en la imagen a modo de subtítulos y se explicó su temporalización (v. § 2.2.2.3.). Sobre este último aspecto se pidió a los participantes que indicaran si tenían alguna duda. Después del visionado, se esperó a que todos los estudiantes hubieran completado el test para recogerlo. En el caso del grupo experimental, se repartió y respondió al cuestionario posterior. El test se corrigió en la siguiente sesión de clase.

Los ítems fueron puntuados por el autor de forma dicotómica. Para estimar la aceptabilidad del ítem de respuesta larga, se utilizaron las respuestas de oyentes expertos recogidas en un estudio anterior (v. § 2.3.1.). El autor volvió a corregir los test dos semanas después con el propósito de comprobar la exactitud de las puntuaciones.

### 2.3.3.4. ANÁLISIS DE DATOS

Las puntuaciones del test se introdujeron en la versión 21 del programa SPSS para Windows. Una semana después se verificó que los datos insertados eran correctos. Posteriormente, se realizó un análisis de frecuencias para comprobar que no había valores anómalos ni valores perdidos. Por último, se calculó el índice de facilidad de cada ítem.

### 2.3.3.4. RESULTADOS Y DISCUSIÓN

El primer objetivo del estudio era averiguar si las indicaciones iniciales sobre la estructura del test, la presentación del texto de entrada, la explicación sobre el funcionamiento de las preguntas subtituladas y la redacción de los ítems resultaban claras para los participantes. Para abordar los tres primeros aspectos, se preguntó directamente a los informantes durante la administración. Se consideró que no plantearon dificultades porque nadie manifestó tener dudas y algunos participantes confirmaron su comprensión. Con el propósito de conocer si la redacción de los ítems era clara se contemplaron dos estrategias: preguntar a los participantes si entendían los ítems antes del visionado y corregir el test con los informantes. En la lectura previa los estudiantes no expresaron tener problemas y algunos confirmaron su comprensión, lo que apuntaba a que eran comprensibles. Esta idea se vio reforzada con la corrección del test, ya que hubo acuerdo sobre cuáles eran las opciones acertadas en los ítems cerrados y el ítem de respuesta única, y qué se consideraba aceptable en el caso de la pregunta abierta.

Un segundo objetivo consistía en investigar si las respuestas de oyentes competentes al ítem abierto obtenidas en el estudio 1 permitían determinar la aceptabilidad de las respuestas de los participantes a esta pregunta. Las experiencias positivas en la corrección de la prueba con los estudiantes, por un lado, y de los treinta y dos test, por otro, indicaban que constitúan un referente para evaluar las preguntas. Otro indicio se hallaba en el índice de facilidad del ítem, ya que este se encontraba dentro de los valores esperados (v. § siguiente).

En tercer lugar se pretendía examinar si los índices de facilidad de los ítems eran plausibles. Se estimaba que los participantes tenían un nivel de comprensión auditiva / audiovisual B2<sup>+</sup> / C1 y que los ítems 1, 2, 3 y 4 eran de nivel A1, A2 / B1, A2 / B1 y A2 / B1, respectivamente. Por este motivo, se esperaban índices de facilidad muy elevados, especialmente en el ítem 1. Todos los ítems tuvieron índices de facilidad conforme a lo esperado. La pregunta 1, de nivel A1, no supuso ninguna dificultad para los estudiantes y fue respondida correctamente por todos. El resto de ítems tuvieron índices de facilidad superiores al 83 %. El índice de facilidad de los ítems se encuentra en la tabla 9.

Tabla 9. Índice de facilidad de los ítems de la secuencia de *Ópera prima*.

GRUPO	ÍTEM	N	ÍNDICE DE FACILIDAD
Control	1	18	100 %
	2	18	100 %
	3	18	89 %
	4	18	83.3 %
Experimental	1	14	100 %
	2	14	86 %
	3	14	86 %
	4	14	92.9 %

En suma, los resultados de esta prueba piloto con los ítems de la secuencia de *Ópera prima* sugirieron que las indicaciones iniciales sobre la estructura del test, la presentación del texto de entrada y la explicación sobre el funcionamiento de las preguntas subtituladas resultaron comprensibles para los informantes, que las respuestas de oyentes

competentes al ítem abierto obtenidas en el estudio 1 servían como base para valorar la pregunta de comprensión de sentimientos, que la redacción de los ítems era clara, y que todos los ítems tenían índices de facilidad plausibles. También cabe señalar que el estudio fue limitado porque participaron pocos informantes en cada grupo (dieciocho en el de control y catorce en el de tratamiento).

#### 2.3.4. ESTUDIO 4

Gracias a la investigación previa fue posible construir una versión completa del test. Las especificaciones de los textos de entrada se encuentran en el § 2.2.1., la descripción de los ítems en los § 2.3.3.3.2. (tarea 1, secuencia de *Ópera prima*) y 2.3.2.3.2. (tarea 2, secuencia de *Los peores años de nuestra vida*) y los ítems en los anexos 4 (tarea 1) y 2 (tarea 2).

El estudio 4 se llevó a cabo en dos fases. La primera parte tenía los objetivos de averiguar si la variante tradicional del test era factible para oyentes competentes y de mejorar un cuestionario para profesores de ELE sobre la subtitulación de preguntas de comprensión audiovisual que sería administrado en la segunda fase (este último aspecto formará parte de otro trabajo). Para conocer si el test era realizable, dos oyentes competentes efectuaron el test en su versión tradicional bajo las condiciones de la prueba. Colaboraron dos especialistas en la enseñanza del ELE: la Prof<sup>a</sup>. Dr<sup>a</sup> Ana María Gimeno (Universidad Politécnica de Valencia) y la Prof<sup>a</sup>. Dr<sup>a</sup> Virginia González (Universidad de Valencia). El estudio tuvo lugar el 14 de enero de 2014 en la Universidad Politécnica de Valencia. La corrección del test reveló que las profesoras no tuvieron ninguna dificultad para completar la prueba, lo que sugirió que era factible para hablantes competentes. Esta primera parte de la investigación fue muy limitada a causa del reducido número de participantes.

La segunda parte del estudio compartió con la primera el propósito de comprobar si la prueba era factible para oyentes competentes. Además, se pretendía averiguar el nivel aproximado de cada ítem de la versión tradicional. Por “aproximado” nos referíamos a una imprecisión de hasta dos niveles del *MCER*. Este grado de exactitud se consideraba suficiente para los objetivos de investigación con los que se diseñaba el test. En tercer lugar, se quería saber la opinión de profesores de ELE sobre la subtitulación de preguntas de comprensión audiovisual (este último punto se desarrollará en un futuro artículo). Con el fin de investigar si la prueba era realizable, cinco profesores de ELE nativos efectuaron el test en su versión tradicional en situación de examen. Para determinar el nivel aproximado de cada pregunta, se entregó una nueva copia del test a los informantes y se les pidió que indicaran el nivel de cada ítem según el *MCER* teniendo en consideración las condiciones de la prueba. Se añadió que, si había dudas entre dos niveles, se sugirieran ambos. Participaron los siguientes docentes del Centro de Idiomas de la Universidad de Valencia: Jaume Brines, Carmen González, Adriana Repila, Vanessa Roca y María José Solomando. El estudio se realizó el 17 de enero de 2014 en el centro de trabajo de los informantes. Los profesores respondieron a todos los ítems adecuadamente, lo que estuvo en consonancia con los resultados obtenidos en la primera fase de la investigación y reforzó la idea de que la prueba era factible para oyentes competentes. Las observaciones de los profesores sobre el nivel de los ítems se recogen en la tabla 10.

Tabla 10. Nivel de los ítems indicado por cinco profesores de ELE del Centro de Idiomas de la Universidad de Valencia.

TEXTO DE ENTRADA	ÍTEM	N	NIVELES INDICADOS
Secuencia de <i>Ópera prima</i>	1	5	A1 tres veces, A1/A2 dos veces
	2	5	A2 tres veces, A2/B1, B1
	3	5	A2 tres veces, A1, B1
	4	5	B1 tres veces, A1/A2, A2/B1
Secuencia de <i>Los peores años de nuestra vida</i>	5	5	B2 dos veces, A2 / B1, B1, B2 / C 1
	6	5	A1 tres veces, A1 / A2, A2
	7	5	B2 dos veces, B1, B1 / B2, B2/C 1

En los siete ítems se identificó una opción mayoritaria, es decir, un nivel indicado por al menos tres de los cinco informantes. Además, en todas las preguntas, excepto en la tercera, existía un nivel sugerido por dos de los cinco participantes. Este número representaba el 40 %. Debido a ello, consideramos que lo más apropiado era tener en cuenta tanto la opción mayoritaria como el nivel sugerido por casi la mitad de los docentes. De acuerdo con este razonamiento, el nivel aproximado de los ítems era el siguiente:

Tabla 11. Nivel aproximado de los ítems.

TEXTO DE ENTRADA	ÍTEM	NIVEL ESTIMADO <sup>a</sup>
Secuencia de <i>Ópera prima</i>	1	A1 / A2
	2	A2 / B1
	3	A2
	4	B1 / A2
Secuencia de <i>Los peores años de nuestra vida</i>	5	B2 / B1
	6	A1 / A2
	7	B2 / B1

a. Cuando se duda entre dos niveles, se sugieren ambos. En estos casos, el primero corresponde al nivel más probable.

En los estudios 2 y 3 el autor ofreció una estimación del nivel de los ítems. Si se compara aquellas sugerencias con las obtenidas en este estudio, se observa un grado de correspondencia relativamente alto, ya que en todos los ítems coincide uno de los niveles propuestos y en tres casos la concordancia es total.

En resumen, fue posible señalar el nivel aproximado de los ítems con una imprecisión de hasta dos niveles del *MCER*. No obstante, cabe señalar dos limitaciones. En primer lugar, las estimaciones se basaron en las sugerencias de solo cinco profesores de ELE. En segundo lugar, el grado de conformidad entre los informantes fue variable. El mayor acuerdo se encontró en el ítem 1, donde tres docentes indicaron el nivel A1 y dos sugirieron A1/A2; y la mayor discrepancia se observó en el ítem 5, donde se mencionaron cuatro niveles distintos. Estas divergencias se atribuyeron principalmente a la propia complejidad de determinar la dificultad de los ítems. Posteriormente se efectuó un nuevo estudio para determinar el nivel aproximado de los ítems y los textos (v. § 2.3.6.).

## 2.3.5. ESTUDIO 5

### 2.3.5.1. OBJETIVOS

El propósito era pilotar<sup>18</sup> los tres instrumentos de investigación de los futuros estudios cuasiexperimentales: los cuestionarios y el test de comprensión audiovisual en sus dos variantes (v. § 1.8.). Como este artículo describe el desarrollo del test, los aspectos relacionados con los cuestionarios se tratarán en otro trabajo. Los objetivos específicos del estudio eran los siguientes:

1. averiguar si la presentación de la estructura general del test (v. § 2.2.2.1.), la introducción de los textos de entrada (v. § 2.2.2.2.), la explicación sobre el funcionamiento de las preguntas subtituladas (v. § 2.2.2.3.) y la redacción de los ítems resultaban claras (v. anexo 5);
2. contabilizar el tiempo empleado por el grupo de control para terminar el test tradicional y por el grupo experimental para completar la prueba y el cuestionario posterior;
3. saber si los índices de facilidad de los ítems eran plausibles;
4. conocer si los índices de discriminación de los ítems eran apropiados;
5. analizar la fiabilidad interna de los test;
6. realizar un análisis descriptivo de las pruebas;
7. investigar si las dos variantes del test estaban correlacionadas con dos tareas de un test de comprensión auditiva.

### 2.3.5.2. DISEÑO DE INVESTIGACIÓN

Se hizo un ensayo del futuro estudio cuasiexperimental con grupos de control y tratamiento no equivalentes. En una clase previa a la prueba se administró el cuestionario previo (v. § 1.8.). El grupo de control efectuó el test de comprensión audiovisual en su versión tradicional y el grupo de tratamiento en su variante experimental. Este último grupo, además, rellenó el cuestionario posterior (v. § 1.8.). Los test se corrigieron con los informantes.

Para estimar de forma indirecta y aproximada el grado de equivalencia de los grupos en relación con la comprensión audiovisual, se administraron dos tareas de un test de comprensión auditiva y se compararon los resultados con la prueba *U* de Mann-Whitney. Para abordar el primer objetivo de investigación se recurrió a las observaciones de los participantes durante la administración de la prueba y su posterior corrección. El tiempo utilizado para completar las pruebas fue contabilizado por el investigador con un reloj.

---

<sup>18</sup> En el área de lingüística aplicada el pilotaje se ha definido como “a dress rehearsal” de los instrumentos y procedimientos investigación para asegurar su calidad (Dörnyei, 2007: 75).

Con el propósito de dar respuesta a los restantes objetivos de investigación, los datos de las dos variantes del test se sometieron a sendos análisis de facilidad, discriminación, fiabilidad y descriptivos. Asimismo, se hicieron estudios de correlación entre los test y dos tareas de un examen de comprensión auditiva.

### 2.3.5.3. MÉTODO

#### 2.3.5.3.1. PARTICIPANTES

Participaron cuarenta y un estudiantes de la asignatura de ELE de nivel A2+ *Español II* ofrecida por la Facultad de Letras de la Universidad de Coímbra en el curso 2013/2014. *Español II* tenía tres turnos y todos ellos eran impartidos por el autor de este trabajo. Dado que dos clases tenían menos personas matriculadas, se decidió que completaran la misma versión del test. Aleatoriamente, los turnos menos numerosos fueron designados grupos de control y el turno restante grupo experimental. El grupo de control estaba formado por veintidós personas y el de tratamiento por diecinueve. Como se mencionó en el estudio 2, según Alderson et al. (1995: 75), para las pruebas piloto “large numbers are not necessary, but if there are at least 20, so much the better”. Además de compartir la asignatura y el docente, los informantes tenían otras características en común. Todos los informantes tenían entre dieciocho y veinticuatro años excepto dos personas: en el grupo de control un participante tenía entre veinticinco y veintinueve años, y en el grupo experimental un estudiante tenía entre cuarenta y cinco y cuarenta y nueve años. Asimismo, todos los participantes eran lusófonos excepto un informante del grupo de tratamiento de origen ucraniano. Igualmente, todos estudiaban licenciaturas de ciencias humanas y sociales (v. tabla 12). *Grosso modo*, los integrantes de cada grupo también llevaban estudiando español un tiempo similar (v. tabla 13) y habían pasado un tiempo análogo en países hispanohablantes (v. tabla 14).

Tabla 12. Estudios de los participantes

GRUPO			FRECUENCIA	PORCENTAJE
Control	Válidos	Estudios Europeos	8	36.4
		Portugués	1	4.5
		Turismo, Ocio y Patrimonio	12	54.5
		Lenguas Modernas	1	4.5
		Total	22	100.0
Experimental	Válidos	Estudios Clásicos	1	5.3
		Estudios Europeos	3	15.8
		Periodismo	4	21.1
		Portugués	4	21.1
		Psicología	3	15.8
		Turismo, Ocio y Patrimonio	4	21.1
		Total	19	100.0

Tabla 13. Tiempo que llevaban estudiando español los informantes

GRUPO			FRECUENCIA	PORCENTAJE
Control	Válidos	Menos de 24 semanas	14	63.6
		Entre 24 y 52 semanas	1	4.5
		Entre 104 y 156 semanas	3	13.6
		Total	18	81.8
		Perdidos <sup>a</sup>	Sistema	4
	Total	22	100.0	
Experimental	Válidos	Menos de 24 semanas	11	57.9
		Entre 24 y 52 semanas	3	15.8
		Entre 104 y 156 semanas	3	15.8
		Más de 156 semanas	1	5.3
		Total	18	94.7
	Perdidos <sup>a</sup>	Sistema	1	5.3
	Total	19	100.0	

a. Los valores perdidos corresponden a informantes que no completaron el cuestionario previo.

Tabla 14. Tiempo pasado en países hispanohablantes por los participantes

GRUPO			FRECUENCIA	PORCENTAJE
Control	Válidos	No han visitado países hispanohablantes	4	18.2
		Menos de una semana	4	18.2
		De 1 a 3 semanas	6	27.3
		De 30 a 36 semanas	1	4.5
		Más de 37 semanas	3	13.6
		Total	18	81.8
		Perdidos <sup>a</sup>	Sistema	4
Total	22	100.0		
Experimental	Válidos	No han visitado países hispanohablantes	3	15.8
		Menos de una semana	3	15.8
		De 1 a 3 semanas	8	42.1
		De 30 a 36 semanas	2	10.5
		Más de 37 semanas	2	10.5
		Total	18	94.7
	Perdidos <sup>a</sup>	Sistema	1	5.3
	Total	19	100.0	

a. Los valores perdidos corresponden a informantes que no completaron el cuestionario previo.

No se administró un pretest para determinar el nivel de comprensión audiovisual en ELE a los participantes porque se desconocía la existencia de pruebas de este tipo. Por los textos auditivos y audiovisuales empleados en clase durante el curso (audiciones de manuales de ELE, canciones, cortometrajes y fragmentos de películas y documentales) y el tipo de tareas, el nivel de comprensión auditiva / audiovisual de la mayor parte de los estudiantes se estimaba en torno al B1<sup>+</sup>. Esta discrepancia entre el nivel global de la asignatura y el nivel de comprensión tenía su origen en la proximidad entre las lenguas y culturas española y portuguesa. Como señala Trindade Natel (2002: 831), el conocimiento previo del estudiante lusófono “funciona como un medio facilitador de la comprensión” oral y escrita en español.

Se consideró que la asignatura y los participantes eran adecuados por dos razones. En primer lugar, el cine español formaba parte del programa y se explotaban regularmente cortometrajes y fragmentos de películas. En segundo lugar, los informantes constituían una muestra representativa de la población, ya que eran estudiantes de ELE adultos con un nivel de comprensión audiovisual en torno al B1<sup>+</sup> y estaban habituados a ver cine y televisión subtítulos (v. § 2. 1. 1.).

Para estimar de forma indirecta y aproximada el grado de equivalencia de los grupos en relación con la comprensión audiovisual, se contrastaron los resultados en dos tareas de un test de comprensión auditiva. Estas tareas procedían de un modelo de examen del DELE B1 diseñado por Hidalgo de la Torre (coord.) (2013: 52 y 53). Concretamente, se seleccionaron las dos primeras actividades del modelo uno porque tanto el cuadernillo del candidato como los textos orales estaban disponibles para descarga en internet<sup>19</sup>. En total había doce ítems que medían la extracción de ideas principales o de datos concretos (dos de las tres habilidades del constructo del test de comprensión audiovisual). Los textos de entrada eran seis mensajes breves (de aprox. cincuenta palabras) en un contestador telefónico y un texto informativo extenso (de aprox. cuatrocientas palabras). Desgraciadamente, las tareas presentaban dos problemas: en el mensaje cuatro de la tarea uno no era Mauro quien llamaba a David, sino al contrario, y en el segundo texto oral no se efectuaba la pausa de treinta segundos para leer las preguntas que se anunciaba en las instrucciones. La errata en el ítem se corrigió en la versión entregada a los informantes. En cuanto a la ausencia de la pausa, el investigador detuvo la reproducción en el momento oportuno durante medio minuto. No se utilizó el modelo de examen del DELE B1 disponible en la web del Cervantes porque los textos orales solo estaban disponibles para su reproducción en continuo y la calidad del audio se juzgó insuficiente para su reproducción en una clase de unos ochenta metros cuadrados. La administración tuvo lugar el 19 de mayo de 2014, ocho semanas después del test de comprensión audiovisual en clase. La prueba se presentó como una actividad de clase en situación de examen para conocer el nivel de comprensión auditiva. Esta circunstancia favorecía que los estudiantes completaran las dos tareas lo mejor posible. Seis participantes del grupo de control faltaron aquel día y no realizaron la prueba. Los análisis estadísticos se realizaron con la versión 21 de SPSS para Windows, con excepción del tamaño del efecto, que se computó en una calculadora. Los grupos obtuvieron puntuaciones similares y altas (9.56 y 9.68 sobre 12), lo que sugería, por un lado, que había pocas diferencias entre los grupos, y por otro, que las tareas de nivel B1 resultaron accesibles (v. la tabla 15). Para comparar los resultados se descartó la posibilidad de utilizar un test *t* para muestras independientes porque no se verificaba el requisito de normalidad en la distribución del grupo experimental ( $S-W = .872$ ,  $gI = 19$ ,  $p = .016$ ) y no se podía invocar el teorema del límite central por el reducido tamaño de la muestra. En su lugar se hizo la prueba *U* de Mann-Whitney para dos muestras independientes. Los resultados indicaron que no había diferencias estadísticamente significativas en las puntuaciones de las tareas en función del grupo ( $U = 146.5$ ,  $z = -.185$ ,  $p = .853$ ,  $r = -$

---

<sup>19</sup> El enlace es el siguiente:

<http://www.octaedro.com/OCTart.asp?libro=53006&id=es&txt=Prepara%20y%20practica%20el%20DELE%20B1> (consulta: 01/03/2016)

.031). Esto sugería que existía cierto grado de equivalencia con respecto a la comprensión audiovisual, dada la relación entre la comprensión audiovisual y la auditiva.

Cabe reconocer varias limitaciones. En primer lugar, no participaron todos los informantes. En segundo lugar, los textos auditivos utilizados no eran representativos del ámbito de uso de la lengua meta del test de comprensión audiovisual. En tercer lugar, no se administró el test completo, sino solo las dos primeras tareas disponibles para descarga. Y en último lugar, hubo un lapso temporal de aproximadamente dos meses entre las pruebas de comprensión audiovisual y auditiva, por lo que los participantes podían haber mejorado o empeorado su español a ritmos diferentes en ese tiempo. Esta amenaza a la validez interna se conoce con el nombre de maduración (Creswell, 2012) y se consideró parcialmente controlada porque los informantes cursaban la asignatura de *Español II* con el mismo docente.

Tabla 15. Descriptivos básicos de dos tareas de un test de comprensión auditiva para preparar el DELE B1

GRUPO		N	MÍNIMO	MÁXIMO	MEDIA	ERROR TÍPICO	DESV. TÍP.
Control	Test	16 <sup>a</sup>	5	12	9.566	.4652	1.861
Experimental	Test	19	7	12	9.684	.4192	1.827

a. El grupo de control estaba constituido por veintidós personas. Solo se dispone de datos de dieciséis participantes porque seis informantes no efectuaron las tareas del test.

### 2.3.5.3.2. INSTRUMENTO

El test contó con dos variantes: una tradicional para el grupo de control y otra experimental para el grupo de tratamiento. En ambas versiones, los ítems se entregaban en papel (v. anexo 5). En la versión experimental, además, los enunciados de las preguntas de comprensión audiovisual aparecían integrados en la imagen a modo de subtítulos (v. figura 5). Para inscribir los enunciados se aplicaron los criterios formulados por Casañ Núñez (2015b: 16-24). A diferencia de los estudios 2 y 3, en esta ocasión se utilizó una fuente sin remates porque es habitual en textos breves, transmite alegría y seguridad (Martínez Castillo, 2009) y es más fácil de leer en una pantalla (Eguaras, 2014). El programa informático utilizado para editar las secuencias fue Adobe Premiere Pro CS5.5 para Windows. La descripción del texto de entrada se encuentra en el § 2.2.1. En la tabla 16 se describen las características principales de los ítems de la versión experimental. Los ítems tradicionales compartían con los experimentales las categorías *enunciado*, *focalización*, *tipo* y *nivel estimado*.

Tabla 16. Descripción de los ítems experimentales del test

ENUNCIADO	FOCALIZACIÓN	TIPO	TEMPORALIZACIÓN	Nº DE LÍNEAS	FUENTE Y TAMAÑO	COLOR	NIVEL ESTIMADO <sup>a</sup>
1. ¿Cómo se llama la chica?	Extraer datos concretos	Elección múltiple	00:01:10:00 00:01:37:09	una	Arial Narrow 36	Blanco (código hex. #FFFFFF)	A1 / A2
2. ¿De qué hablan?	Identificar las ideas generales	Elección múltiple	00:01:44:14 00:02:02:02	una			A2 / B1
3. ¿Cuál es el teléfono de la chica?	Extraer datos concretos	Respuesta única	00:02:04:04 00:02:22:00	una			A2
4. ¿Qué sentimientos puede tener el chico por la chica?	Reconocer sentimientos	Respuesta abierta	00:02:27:08 00:02:36:12	dos			B1 / A2
5. ¿De qué temas habla el chico pelirrojo?	Extraer datos concretos	Elección múltiple	00:11:40:00 00:12:25:24	una			B2 / B1
6. ¿Cómo se llama la chica?	Identificar las ideas generales	Elección múltiple	00:12:39:04 00:12:43:24	una			A1 / A2
7. ¿Qué sentimientos puede tener el chico pelirrojo por la chica?	Reconocer sentimientos	Respuesta abierta	00:12:45:00 00:12:48:15	dos			B2 / B1

a. Estas orientaciones proceden de un estudio anterior (v. § 2.3.4). Posteriormente fue posible determinar con mayor exactitud el nivel de los ítems (v. § 2.3.6.4).



Figura 5. Fotogramas de *Ópera prima* y *Los peores años de nuestra vida* con preguntas de comprensión audiovisual subtituladas.

### 2.3.5.3.3. PROCEDIMIENTO

El test fue administrado por el investigador a los tres turnos de *Español II* en la misma clase el 26 de marzo de 2014. El turno de 14:00 a 16:00 actuó como grupo de control y los turnos de 16:00 a 18:00 y 18:00 a 20:00 como grupos de tratamiento. La prueba se planteó como una actividad de clase en situación de examen para conocer el nivel de comprensión audiovisual. El hecho de que el investigador fuera también el docente de la

asignatura y de que los test se recogieran favoreció que los estudiantes se tomaran la prueba en serio, condición necesaria para que los resultados del pilotaje fueran válidos.

Para reproducir el vídeo se empleó el equipo técnico que se empleaba regularmente en esa clase: pantalla de proyección tripié, videoproector portátil, ordenador portátil y dos altavoces. La administración fue análoga en los grupos de control y tratamiento, salvo en aquellos aspectos en que tenía que ser necesariamente diferente.

Justo antes de iniciar la administración se anotó la hora de inicio. A continuación se describió la estructura general del test (v. § 2.2.2.1.) y se preguntó a los informantes si había alguna duda. Acto seguido, se comentó que la prueba se corregiría una vez terminada. Luego se introdujo oralmente la secuencia de *Ópera prima* (v. § 2.2.2.2.) y se pidió a los estudiantes que confirmaran la comprensión o expresaran sus dificultades. Después se repartió la tarea 1 del test (v. anexo 5), se pidió a los participantes que la leyeran con atención y se preguntó si había alguna dificultad. En lo que respecta al ítem de respuesta larga, se consideró pertinente justificar la posibilidad de escribir en la L1 porque el objetivo de la prueba era evaluar la comprensión y no la producción. Además, en el caso de los grupos de tratamiento, se informó de la existencia de las preguntas integradas en la imagen a modo de subtítulos y se precisó su funcionamiento (v. § 2.2.2.3.). Sobre este último aspecto se preguntó a los participantes si existía alguna duda. Después del visionado, se aguardó a que todos los informantes hubieran terminado el test para recogerlo. En el caso de los grupos de tratamiento, se repartió el cuestionario posterior relacionado con la tarea 1 (v. § 1.8.). De nuevo, se esperó a que todos los participantes lo completaran para recogerlo.

Para administrar la segunda tarea, se reprodujo la secuencia de administración de la primera, con excepción de las partes que resultaban redundantes, es decir, la justificación de la escritura en la L1 y la explicación sobre las preguntas subtituladas. Después se registró la hora de finalización. Por último, se corrigió el test con los participantes.

Los ítems fueron calificados de forma dicotómica por el autor. Para estimar la aceptabilidad de las preguntas de respuesta larga, se utilizaron las respuestas de oyentes competentes recogidas en un estudio anterior (v. § 2.3.1.) y se aprovechó la experiencia de corrección de los estudios 2 y 3. El autor volvió a corregir los test dos semanas después con el propósito de comprobar la exactitud de las puntuaciones.

#### 2.3.5.3.4. ANÁLISIS DE DATOS

Los datos de las dos versiones del test se introdujeron en la versión 21 del programa SPSS para Windows. Una semana después se verificó que la información insertada era correcta. A continuación se realizó un análisis de frecuencias para comprobar que no había valores anómalos ni valores perdidos. Posteriormente, los datos de los test se sometieron a sendos análisis de facilidad, discriminación, fiabilidad y descriptivos. Adicionalmente, se efectuaron estudios de correlación entre los test y dos tareas de un test de comprensión auditiva. Todas las operaciones se hicieron en SPSS salvo aquellas que no son calculadas automáticamente por el programa, como sucede, entre otros, con los estadísticos de asimetría y curtosis estandarizados. En esos casos, se recurrió a una calculadora.

## 2.3.5.4. RESULTADOS Y DISCUSIÓN

### 2.3.5.4.1. CLARIDAD DE LAS INDICACIONES Y LOS ÍTEMS

El primer objetivo del estudio era averiguar si la estructura general del test (v. § 2.2.2.1.), las presentaciones de los textos de entrada (v. § 2.2.2.2.), el funcionamiento de las preguntas subtituladas (v. § 2.2.2.3.) y la redacción de los ítems (v. anexo 5) resultaban claros para los informantes. Para responder a estas cuestiones, se preguntó directamente a los participantes durante la administración. Además, en el caso de los ítems, se corrigió la prueba con los estudiantes. En relación con la estructura de la prueba y la introducción de los textos de entrada, informantes de los tres grupos confirmaron su comprensión y nadie mostró tener dudas. De forma análoga, en los grupos experimentales hubo participantes que manifestaron comprender el funcionamiento de las preguntas subtituladas y nadie expresó tener dificultades. Con respecto a los ítems, los resultados son similares: en la lectura previa algunas personas confirmaron su inteligibilidad y nadie exteriorizó ninguna complicación; y en la corrección del test hubo acuerdo sobre las respuestas a los ítems cerrados y el ítem de respuesta única, y sobre qué podría considerarse aceptable en el caso de las preguntas abiertas. De la revisión de la prueba también son dignos de mención otros aspectos: varios estudiantes comentaron que la secuencia de *Los peores años de nuestra vida* era claramente más complicada que la de *Ópera prima* y que esto se debía a que el chico pelirrojo hablaba muy deprisa. Asimismo, a propósito de la complejidad de los textos, surgió la cuestión de la dificultad de los ítems. Sobre este aspecto se registraron diversidad de opiniones pero los tres grupos observaron que las preguntas uno y seis eran las más sencillas. En resumen, las observaciones de los informantes sugirieron que las indicaciones sobre la estructura del test, las presentaciones de los textos de entrada, la explicación sobre el funcionamiento de las preguntas subtituladas y la redacción de los ítems resultaron claras.

### 2.3.5.4.2. TIEMPO EMPLEADO

La administración del test tradicional tuvo una duración de doce minutos y veintisiete segundos. Para completar el test experimental y el cuestionario posterior fueron necesarios veinte minutos y catorce segundos en el primer grupo experimental y diecinueve minutos y cuarenta y dos segundos en el segundo. Los examinandos disponen del tiempo que deseen para completar el test y los cuestionarios (v. § 2.1.5.), pero estos tiempos sirven de referencia para planificar futuros estudios.

### 2.3.5.4.3. ANÁLISIS DE FACILIDAD

Los índices de facilidad de los ítems de las dos versiones del test presentaban valores razonables teniendo en cuenta las estimaciones del nivel de comprensión audiovisual de los participantes (B1+) y de los ítems. Los ítems de un nivel muy inferior a los participantes (ítems 1 y 6) tenían índices de facilidad muy altos (superiores al 91 % en ambos grupos), los de nivel inferior (ítems 2, 3 y 4) presentan niveles altos o medio-altos (entre el 64 y el 82% en el grupo de control y entre el 63 y el 74 % en el de tratamiento), y aquellos ítems de nivel superior a los informantes (ítems 5 y 6) tenían índices de dificultad medios (del 50 y el 59 % en el grupo de control y del 47 y el 58 % en el de tratamiento). Según Alderson et al. (1995) los índices de facilidad en torno al 50 % son los que tienen mayor potencial

de discriminación porque producen la mayor variación de las puntuaciones. Green (2013) sugiere que los ítems con valores entre 20 y 80 % todavía son informativos si son capaces de discriminar y contribuyen a la validez interna. La autora añade que los valores apropiados dependen de los objetivos del test, los destinatarios y el uso de los índices de facilidad. En las dos versiones del test era pertinente que el ítem inicial fuera fácil o muy fácil, ya que, además de medir una habilidad del constructo, cumplía otras dos funciones: crear confianza en los grupos de control y tratamiento y servir como ejemplo de pregunta subtitulada para el grupo experimental (v. § 2.1.6.). También era lógico que hubiera gran variabilidad en los índices de facilidad porque había ítems con distintos niveles de dificultad (del A1/A2 al B1/B2). Por último, los ítems con índices de facilidad muy altos no dejaban de ser apropiados para comparar las puntuaciones de los grupos de control y tratamiento, conocer la opinión del grupo experimental sobre esta técnica e investigar la conducta visual de los grupos (v. § 2.1.1.). Por ejemplo, si los ítems con índices de facilidad muy altos en la versión tradicional tuvieran índices medios o bajos en la versión experimental, se podría deducir que la técnica docente dificulta la comprensión en lugar de facilitarla.

Tabla 17. Índice de facilidad de los ítems.

TEST	ÍTEM	N	ÍNDICE DE FACILIDAD	NIVEL ESTIMADO <sup>a</sup>
Tradicional	1	22	100 %	A1 / A2
	2	22	81.82 %	A2 / B1
	3	22	72.73 %	A2
	4	22	63.64 %	B1 / A2
	5	22	50 %	B2 / B1
	6	22	90.91 %	A1 / A2
	7	22	59.09 %	B2 / B1
Experimental	1	19	100 %	A1 / A2
	2	19	73.68 %	A2 / B1
	3	19	63.16 %	A2
	4	19	68.42 %	B1 / A2
	5	19	47.37 %	B2 / B1
	6	19	94.74 %	A1 / A2
	7	19	57.89 %	B2 / B1

a. Estas orientaciones proceden de un estudio anterior (v. § 2.3.4.). Posteriormente fue posible determinar con mayor exactitud el nivel de los ítems (v. § 2.3.6.4).

#### 2.3.5.4.4. ANÁLISIS DE DISCRIMINACIÓN

El cuarto objetivo de la investigación era conocer si los índices de discriminación de los ítems eran apropiados. Este índice se puede hallar de diversas maneras (Alderson et al., 1995). Para obtenerlo se calculó la correlación elemento-total corregida (en adelante CETC), como sugiere Green (2013).

Tabla 18. Índice de discriminación de los ítems.

TEST	ÍTEM	CORRELACIÓN ELEMENTO-TOTAL CORREGIDA
Tradicional	1	.000
	2	.459
	3	.343
	4	.533
	5	.445
	6	.378
	7	.285
Experimental	1	.000
	2	.541
	3	.444
	4	.574
	5	.502
	6	.301
	7	.421

Green (2013) expone que un valor de + .3 se considera normalmente aceptable y que para Henning (1987, citado en Green, 2013) .25 sería el mínimo admisible. En las dos versiones del test los ítems 2-7 presentaban un CETC superior a .3 o al menos .25, por lo que tenían valores aceptables o admisibles. En relación con la pregunta inicial, era lógico que su CETC fuera cero porque su índice de facilidad era 100 %. También era aceptable que fuera así por sus objetivos particulares (v. § 2.1.6. y 2.3.5.4.3.).

#### 2.3.5.4.5. ANÁLISIS DE FIABILIDAD INTERNA

Los participantes dispusieron de todo el tiempo que desearon para completar la prueba (v. § 2.1.5.). Esto constituye una condición necesaria para estimar la fiabilidad interna (Alderson et al., 1995). Para determinar la fiabilidad se calculó el coeficiente de alfa de Cronbach del test y el coeficiente de alfa de Cronbach si se elimina cada uno de los ítems. Además, se halló la media de las correlaciones inter-elementos.

La fiabilidad interna de la versión tradicional del test fue  $\alpha = .650$  y de la versión experimental  $\alpha = .705$ . Field (2009) y Pallant (2011) indican que los valores iguales o superiores a .7 suelen considerarse aceptables. Estos expertos en estadística también señalan que el número de elementos influye en el coeficiente de alfa. Field (2009) advierte que es posible obtener un valor de alfa elevado en una escala porque hay muchos elementos y no porque la escala es fiable. Pallant (2011: 97) menciona que “with short scales (e.g. scales with fewer than ten items) it is common to find quite low Cronbach values (e.g. .5)”. Para Suhr y Shay (2009) el nivel adecuado de fiabilidad interna depende del propósito del instrumento. Estos autores precisan que los valores iguales o mayores a .6 son admisibles para instrumentos con fines de investigación. Teniendo en cuenta el reducido número de ítems y los objetivos de las dos variantes de la prueba, se interpretó que los coeficientes de alfa obtenidos eran positivos.

También se calculó el coeficiente alfa de Cronbach si se elimina el elemento (v. tabla 19). Si el instrumento es fiable, ningún ítem debe reducir substancialmente el alfa (Field, 2009). Todos los elementos contribuían positivamente, excepto el ítem 1, que tenía un ligero impacto negativo: prescindir de esta pregunta aumentaba el  $\alpha$  a .669 en el grupo

de control y a .725 en el de tratamiento. Como la incidencia en el alfa era pequeña y el ítem, además de medir una habilidad del constructo, tenía otras funciones (v. § 2.1.6.), se mantuvo en la prueba.

Tabla 19. Alfa de Cronbach si se elimina el elemento.

TEST	ÍTEM	ALFA DE CRONBACH SI SE ELIMINA EL ELEMENTO
Tradicional	1	.669
	2	.585
	3	.620
	4	.550
	5	.585
	6	.616
	7	.645
Experimental	1	.725
	2	.636
	3	.665
	4	.625
	5	.647
	6	.700
	7	.673

Pallant (2011) recomienda calcular la media de las correlaciones inter-elementos cuando la escala tiene pocos elementos. El cómputo se hizo en SPSS. El programa prescindió del ítem 1 en el análisis porque tenía varianza cero. Las medias de las correlaciones inter-elementos de los ítems 2-7 fueron .262 en el grupo de control y .306 en el experimental. Según Briggs y Cheek (1986: 115) "the optimal level of homogeneity occurs when the mean inter-item correlation is in the .2 to .4 range". Por lo tanto, los valores obtenidos eran adecuados.

#### 2.3.5.4.6. ANÁLISIS DESCRIPTIVO

Tabla 20. Descriptivos principales del test.

TEST	N	MINIMO	MÁXIMO	MEDIA	MEDIA (%)	ERROR TÍP. MEDIA	MEDIANA	DESV. TÍP.	ASIMETRÍA	ERROR TÍP. ASIMETRÍA	CURTOSIS	ERROR TÍP. CURTOSIS
Tradicional	22	1	7	5.182	74.029	.3521	5.500	1.652	-1.016	.491	.731	.953
Experimental	19	2	7	5.053	72.186	.4081	5.000	1.779	-.618	.524	-.814	1.014

Las medias elevadas sugerían que las dos versiones de la prueba resultaron fáciles para los participantes. Esto era lo esperado porque se había estimado que cinco de los siete ítems eran de un nivel inferior al de los informantes. Las desviaciones típicas eran relativamente pequeñas en comparación con las medias. Cuando esto ocurre, las puntuaciones del test están próximas a la media y el valor del promedio aritmético es representativo de los resultados (Field, 2009).

El hecho de que los estadísticos de asimetría fueran negativos indicaba que las puntuaciones de los examinandos se acumulan en la parte derecha de la distribución. Se

denomina negativa porque “the scores tail off toward the left end of the graph” (Alderson et al., 1995: 93). Para interpretar el estadístico de asimetría lo estandarizamos dividiéndolo por su error típico. Si los valores están comprendidos entre  $\pm 1.96$  la distribución se considera simétrica con un nivel de confianza del 95 % (Rial Boubeta y Varela Mallou, 2008). En el grupo de control se obtuvo -2.069 y en el de tratamiento -1.17.

El estadístico de curtosis del test tradicional fue positivo, lo que sugería que existía una tendencia leptocúrtica, es decir, que había una proporción considerable de casos en la zona central. En cambio, en el caso del test experimental, el estadístico era negativo. Esto apuntaba a que había una tendencia platicúrtica, esto es, que las puntuaciones estaban más dispersas. Para interpretar el estadístico de asimetría lo estandarizamos dividiéndolo por su error típico. Si los valores están comprendidos entre  $\pm 1.96$  la distribución se considera normal (Rial Boubeta y Varela Mallou, 2008). En el grupo de control se obtuvo .767 y en el de tratamiento -.803, por lo que se podía asumir que las distribuciones eran mesocúrticas.

#### 2.3.5.4.7. CORRELACIONES

Era esperable que hubiera una asociación positiva entre las dos versiones del test de comprensión audiovisual y las dos tareas de un test de comprensión auditiva administradas a los informantes (v. § 2.3.5.3.1.). Para observar gráficamente la relación entre las variables en cada grupo se construyeron sendos diagramas de dispersión (v. las figuras 6 y 7).

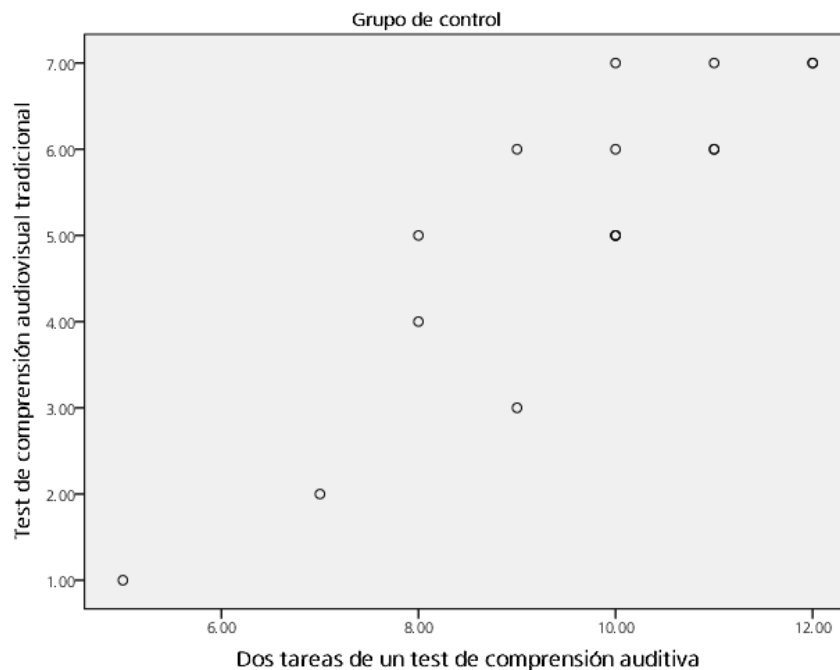


Figura 6. Diagrama de dispersión del test de comprensión audiovisual tradicional y dos tareas de un test de comprensión auditiva diseñado por Hidalgo de la Torre (coord.) (2013: 52 y 53).

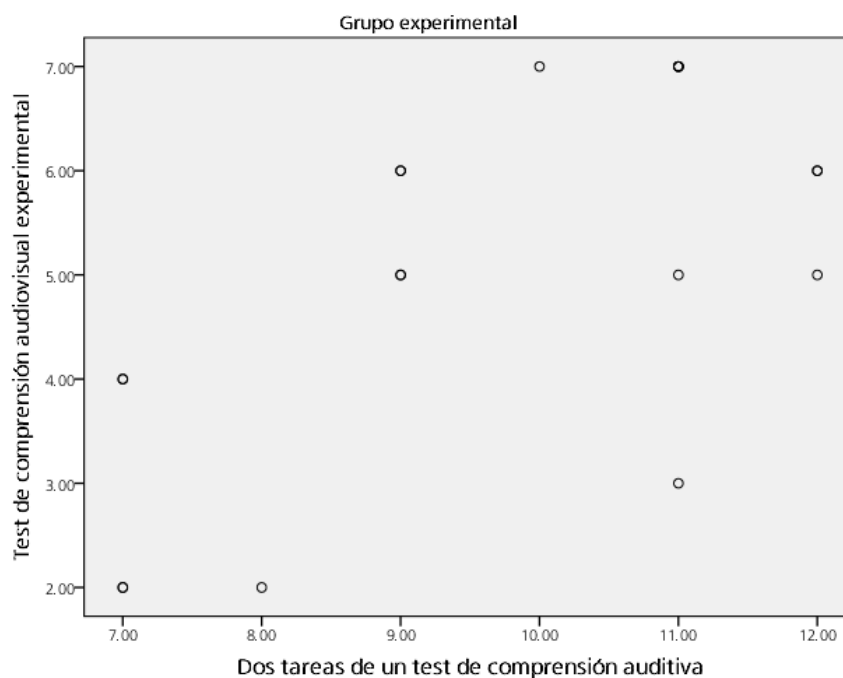


Figura 7. Diagrama de dispersión del test de comprensión audiovisual experimental y dos tareas de un test de comprensión auditiva diseñado por Hidalgo de la Torre (coord.) (2013: 52 y 53).

Los gráficos mostraban que existía cierta relación entre los resultados, dado que los puntos tendían a avanzar de la parte inferior izquierda a la parte superior derecha. Esta correlación era más fuerte y evidente en el grupo de control que en el experimental.

A continuación calculamos el coeficiente de correlación. Se descartó la posibilidad de hallar el coeficiente de Pearson porque las distribuciones del test de comprensión audiovisual no eran normales ni en el grupo de control ( $S-W: = .888$ ,  $gl = 22$ ,  $p = .017$ ) ni en el de tratamiento ( $S-W: = .874$ ,  $gl = 19$ ,  $p = .017$ ). Además, la distribución del test de comprensión auditiva del grupo experimental tampoco lo era ( $S-W: = .872$ ,  $gl = 19$ ,  $p = .016$ ). En su lugar, se computó el coeficiente de Spearman (v. tabla 21).

Las dos tareas del test de comprensión auditiva estaban relacionadas significativamente tanto con la versión tradicional del test de comprensión audiovisual ( $r_s = .833$ ) como con la versión experimental ( $r_s = .592$ ) (ambos  $ps < .01$ ). Dado que el valor de los coeficientes era superior a .5, se trataba de correlaciones fuertes (Cohen, 1988). Finalmente, calculamos los coeficientes de determinación para conocer la proporción de varianza en los rangos que compartían las variables. En el caso del grupo de control  $R^2_s = .694$  y en el caso del grupo de tratamiento  $R^2_s = .350$ .

Tabla 21. Correlaciones entre la prueba de comprensión audiovisual y dos tareas de un test de comprensión auditiva diseñado por Hidalgo de la Torre (coord.) (2013: 52 y 53).

	GRUPO		COMPRESIÓN AUDIOVISUAL	COMPRESIÓN AUDITIVA	
Rho de Spearman	Control	Comprensión audiovisual	Coeficiente de correlación	1.000	.833**
			Sig. (bilateral)	.	.000
			N	22	16 <sup>a</sup>
		Comprensión auditiva	Coeficiente de correlación	.833**	1.000
		Sig. (bilateral)	.000	.	
		N	16 <sup>a</sup>	16 <sup>a</sup>	
	Experimental	Comprensión audiovisual	Coeficiente de correlación	1.000	.592**
			Sig. (bilateral)	.	.008
		N	19	19	
Comprensión auditiva		Coeficiente de correlación	.592**	1.000	
	Sig. (bilateral)	.008	.		
	N	19	19		

\*\* . La correlación es significativa al nivel .01 (bilateral).

a. No se dispone de datos de los veintidós participantes porque seis informantes no hicieron la prueba de comprensión auditiva.

#### 2.3.5.4.8. LIMITACIONES

El estudio presentaba varias limitaciones. En primer lugar, no se controló la amenaza a la validez interna conocida como difusión de tratamientos<sup>20</sup> en las pruebas de comprensión audiovisual y de comprensión auditiva, dado que los grupos preestablecidos efectuaron las pruebas de forma consecutiva. Esto se debió a que los informantes tenían horarios lectivos dispares y no era viable realizar los test de forma simultánea. Aunque no se controló esta amenaza, existían dos circunstancias que reducían la probabilidad de que se hubiera producido el intercambio de información. La primera residía en el hecho de que había poco tiempo para que se llevase a cabo esta interacción porque las pruebas se hicieron de forma consecutiva y los informantes tenían diferentes horarios de clase. El segundo motivo era que los test no tenían repercusión en la nota final y, en principio, no existía un interés manifiesto por conocer su contenido.

En segundo lugar, la selección de los participantes constituía potencialmente una amenaza a la validez interna porque los grupos estaban preestablecidos. No obstante, esta posible fuente de invalidez se consideró parcialmente controlada porque la designación de los turnos de la asignatura de *Español II* como grupos de control y tratamiento fue aleatoria, porque todos los informantes guardaban ciertas similitudes (estudiantes de la misma universidad y asignatura de ELE, mismo docente, etc. v. § 2.3.5.3.1.) y porque no se obtuvieron diferencias significativas en la prueba *U* de Mann-

<sup>20</sup> Esto "ocurre cuando los sujetos de un grupo aprenden algo sobre el tratamiento o las condiciones de otro grupo" (Sans Martín, 2011: 27).

Whitney que comparaba las puntuaciones de los grupos en las dos tareas del test de comprensión auditiva. Cabe añadir que esta posible amenaza no afectaba al desarrollo individual de los test, solo a las comparaciones que pudieran hacerse entre los grupos.

En tercer lugar, la estrategia contemplada para estimar de forma indirecta y aproximada el grado de equivalencia de los grupos en relación con la comprensión audiovisual presentaba varios inconvenientes (v. el final del § 2.3.5.3.1.). De nuevo, cabe mencionar que esta limitación no repercutía en el desarrollo de las pruebas, solo en las comparaciones que se efectuaran entre los grupos.

Por último, hubiera sido conveniente realizar un análisis factorial para investigar el constructo subyacente del test. Sin embargo, el número de casos era insuficiente. A este respecto hay autores que recomiendan que la proporción entre el número de participantes y el número de ítems sea como mínimo de cinco a uno (Streiner, 1994, citado en O'Rourke y Hatcher, 2013) y otros que sugieren que haya al menos trescientos casos (Tabachnik y Fidell, 2007).

### 2.3.6. ESTUDIO 6

Entre los propósitos del estudio 4 (v. § 2.3.4.) se encontraba estimar el nivel aproximado de los ítems a partir de las valoraciones de profesores de ELE. Fue posible sugerir la dificultad de las preguntas pero se observaron distintos grados de convergencia en las aportaciones de los docentes. En aquel momento se consideró que las diferencias se debían fundamentalmente a la propia complejidad de determinar el nivel de ítems de comprensión audiovisual. Posteriormente, con la consulta del *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Consejo de Europa, 2009) y a la asistencia a cursos sobre la evaluación del ELE<sup>21</sup>, corroboramos el argumento de la complejidad y nos percatamos de dos motivos adicionales. En relación con el primer aspecto, el Consejo de Europa (2009: 37) señala que "listening is the most difficult skill to work on and so should be treated last. Several case studies piloting the Manual showed a considerably higher level of rater agreement and a lower spread of scores for production samples than for reception items". Las razones adicionales fueron de dos índoles. En primer lugar, el diseño de investigación era manifiestamente mejorable. Entre los puntos débiles, destacamos que no se hicieron actividades de familiarización con las escalas de comprensión auditiva y audiovisual del *MCER* ni se discutió el nivel de cada ítem hasta llegar a un consenso. En segundo lugar, las escalas de comprensión auditiva del *MCER* no son perfectas (Figueras, 2014). Alderson, Figueras, Kuijper, Nold, Takala y Tardieu (2006: 9) identifican cuatro tipos de problemas en la formulación de los descriptores del *MCER* de comprensión lectora y auditiva:

---

<sup>21</sup> Nos referimos a los cursos de acreditación de examinadores de los DELE ofrecidos por el Instituto Cervantes, y a la actividad formativa *Desarrollo y gestión de exámenes vinculados al Marco común europeo de referencia* organizada por la Universidad Internacional Menéndez Pelayo (en adelante UIMP) y el Instituto Cervantes en julio y agosto de 2014.

1. Inconsistencies, where a feature might be mentioned at one level but not at another, where the same feature might occur at two different levels, or where at the same level a feature might be described differently in different scales.
2. Terminology problems: synonymy or not?
3. Lack of definition, where terms might be given, but are not defined.
4. Gaps, where a concept or feature needed for test specification or construct definition is simply missing.

### 2.3.6.1. OBJETIVOS

Con este estudio se perseguía estimar el nivel aproximado de los ítems de la versión tradicional y el de los textos audiovisuales. "Aproximado" se entendía como una inexactitud de hasta dos niveles del *MCER*. Este grado de imprecisión se consideraba suficiente para los propósitos con los que se diseñaba el test.

### 2.3.6.2. DISEÑO DE INVESTIGACIÓN

Se pretendía que el nivel de los textos de entrada y de los ítems fuera determinado por un panel de jueces. Con este fin, se aplicaron de forma parcial algunos principios del *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Consejo de Europa, 2009)<sup>22</sup>. La investigación se planificó en cinco fases:

1. Reunión de un grupo de profesionales del ELE familiarizados con el *MCER*, los niveles de referencia y las escalas ilustrativas.
2. Repaso de las escalas de comprensión auditiva y audiovisual más pertinentes para el test por parte de los especialistas al inicio de la reunión.
3. Exposición de las instrucciones que se proporcionan en la versión tradicional de la prueba (v. § 2.2.2.).
4. Valoración individual del nivel de los ítems y de los textos por los especialistas.
5. Discusión del nivel de los ítems y los textos hasta alcanzar el máximo consenso posible.

El último paso no se realizó por falta de tiempo. En su lugar, el autor estimó el nivel de los ítems a partir del examen de las respuestas individuales de los profesionales.

---

<sup>22</sup> Este manual propone una metodología rigurosa para vincular exámenes con el *MCER*. Se trata de un proceso complejo: "linking an examination to the *CEFR* is a complex process involving many steps, which all require a professional approach" (Consejo de Europa, 2009: 90). La puesta en práctica de todos estos protocolos no estaba dentro de nuestras posibilidades en términos de conocimientos y de recursos. A este respecto, cabe añadir que el manual tiene un enfoque inclusivo y sugiere "the application of principles of best practice even in situations in which resources and expertise are limited" (9).

### 2.3.6.3. MÉTODO

#### 2.3.6.3.1. PARTICIPANTES

Colaboraran nueve personas: cuatro profesores de ELE del Instituto Cervantes, cuatro profesores de ELE universitarios y un redactor de ítems. Seis decidieron participar de forma anónima y tres revelar su identidad. Los informantes que se inclinaron por revelar su nombre fueron Valentín Cózar (Instituto Cervantes de Lisboa), Francisco del Moral (Universidad de Verona) y Carmen San Julián (Trinity College Dublin). Siete participantes eran españoles y dos polacos. Todos provenían de distintos centros, tenían experiencia en trabajar con el *MCER* y fueron reclutados en el curso *Desarrollo y gestión de exámenes vinculados al MCER* organizado por la UIMP y el Instituto Cervantes en Santander entre el 28 de julio y el 1 de agosto de 2014. Se trataba de una actividad formativa dirigida a profesores con una sólida trayectoria profesional y formativa, lo que constituía un claro indicio de la idoneidad de los participantes. En este curso de veinticinco horas presenciales se describieron y realizaron tareas necesarias para la formación de panelistas, entre otros, actividades de familiarización con las escalas ilustrativas del *MCER* y actividades de ilustración con muestras estandarizadas. Se considera que esto contribuyó a la formación de los jueces de este estudio. Por último, cabe añadir que el investigador no conocía directamente a ninguno de los participantes antes del inicio de la formación y que todos colaboraron desinteresadamente.

#### 2.3.6.3.2. INSTRUMENTO

Se utilizó la variante del test tradicional. Las especificaciones de la prueba se encuentran en el § 2.1., la descripción de los textos de entrada en la § 2.2.1., las características de los ítems en el § 2.3.5.3.2. y el test en los anexos 4 (tarea 1) y 2 (tarea 2).

#### 2.3.6.3.3. PROCEDIMIENTO

El primer paso consistió en reunir a un grupo de profesionales del ELE familiarizados con el *MCER*, los niveles de referencia y las escalas ilustrativas. Como se ha mencionado anteriormente (v. § 2.3.6.3.1.), los especialistas fueron reclutados entre los asistentes a un curso sobre el diseño de exámenes vinculados al *MCER*. El primer día de clase el investigador y también participante en la formación contactó con la dirección del curso para exponer los objetivos del estudio. También comentó su propósito a algunos participantes y recibió respuestas alentadoras. El último día de clase la dirección del curso brindó al autor la posibilidad de dirigirse a los asistentes a la formación. El investigador se presentó como lector de la Universidad de Coímbra y doctorando en la Universidad Politécnica de Valencia, y explicó los objetivos generales del test y la necesidad de reunir a un grupo de profesionales para estimar el nivel aproximado de los ítems. Se precisó que no se efectuaría la fase de discusión del nivel por el poco tiempo disponible y que la estimación del nivel se haría a partir de las respuestas individuales. Los especialistas manifestaron que colaborarían de buena voluntad. Se acordó que el momento más conveniente para la mayoría era ese mismo día, 1 de agosto de 2014, a las 16:30. Varias personas justificaron su imposibilidad de participar por compromisos previos y se entendió perfectamente.

La reunión tuvo lugar en la clase donde se había celebrado el curso de formación. Para reproducir las secuencias cinematográficas, se utilizó el equipo que se había utilizado durante la actividad formativa (videoprojector, altavoces, etc.) con excepción del ordenador, propiedad del investigador. Primero de todo, el autor agradeció a los informantes su participación. Acto seguido, comentó que las respuestas eran confidenciales y que solo se utilizarían con fines de investigación. También señaló que se podía participar de forma anónima o revelando la identidad. A continuación, los especialistas repasaron las escalas de comprensión auditiva y audiovisual más pertinentes para el test. Con este fin, se repartieron copias de las escalas del *MCER comprensión auditiva en general, comprender conversaciones entre hablantes nativos, identificación de las claves e inferencia (oral y escrita) y ver televisión y cine* (Consejo de Europa, 2002: 69, 73 y 74). Los informantes dedicaron un cuarto de hora a examinar las escalas individualmente.

Luego el autor anunció que les entregaría una copia en papel del test y que les pediría que anotaran el nivel de cada ítem y de los textos según el *MCER* en estas hojas. El investigador añadió que, si había dudas entre dos niveles, se sugirieran ambos. Después se precisaron las indicaciones iniciales que se proporcionan en la versión tradicional del test (v. § 2.2.2.1.) y se repartió el test (v. anexos 4 y 2). Antes de ver cada secuencia, se introdujeron los textos como está contemplado en la administración (v. § 2.2.2.2.). Primero se visionó la secuencia de *Ópera prima* y después la de *Los peores años de nuestra vida*. Los especialistas estimaron individualmente el nivel de los ítems y de los textos y registraron sus respuestas en las hojas de examen. En ambos casos se efectuaron cuantos visionados fueron solicitados por los profesionales.

Hubiera sido deseable discutir el nivel de los ítems y los textos hasta llegar al máximo consenso posible. Para ello, se hubiera adaptado a nuestras posibilidades el protocolo del *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Consejo de Europa, 2009). Como se ha comentado, la discusión no se llevó a cabo por el poco tiempo disponible. En su lugar, se recogieron las contestaciones de los profesionales y el autor estimó la dificultad de los ítems a partir de los niveles sugeridos por los informantes.

#### 2.3.6.3.4. ANÁLISIS DE DATOS

Se calcularon las frecuencias absolutas del nivel de cada ítem. Este cómputo se hizo manualmente por el reducido número de participantes. Para comprobar la exactitud de los resultados, el autor realizó la operación en dos ocasiones.

#### 2.3.6.4. RESULTADOS Y DISCUSIÓN

Las observaciones de los profesionales se recogen en la tabla 22. En las preguntas de comprensión de sentimientos (último ítem de los tareas) dos informantes no indicaron el nivel. Uno de ellos mencionó que consideraba que los ítems “no tienen que ver directamente con la comprensión lingüística”. Respetamos este punto de vista pero no lo compartimos. Las respuestas de oyentes competentes obtenidas en un estudio anterior (v. § 2.3.1.) sugirieron que se trataba de ítems que se contestaban por acumulación de

indicios verbales y no verbales. En el primer caso, el chico decía frases como “estás muy bien” y “estás formidable”, le pedía el número de teléfono, quería quedar con ella para cenar, se quedó pensativo mirando su número de teléfono... En el segundo caso, el chico se quedó estupefacto cuando la vio, intentó agradecerla cambiando de opinión sobre la Navidad, le molestó que el otro chico no le hubiera dicho que la había visto otro día... Esto es congruente con la definición de comprensión auditiva / audiovisual defendida en este trabajo (v. § 1.1.).

Tabla 22. Nivel de los textos y los ítems sugeridos por nueve profesionales del ELE.

TEXTO DE ENTRADA	ÍTEM	N	NIVELES INDICADOS
Secuencia de <i>Ópera prima</i>		9	A2 tres veces, B1 tres veces, A1, A1 / A2, A2 / B1
	1	9	A1 siete veces, A2 dos veces
	2	9	A2 cuatro veces, B1 tres veces, A1, B1 / B2
	3	9	A2 cinco veces, A1 dos veces, A1 <sup>+</sup> , B1 <sup>+</sup>
	4	9	A1 dos veces, B2 dos veces, A2, A2 / B1, C1, dos profesionales no indican el nivel
Secuencia de <i>Los peores años de nuestra vida</i>		9	B2 cinco veces, C1 cuatro veces
	1	9	B2 cinco veces, B1, B1 <sup>+</sup> , B2 / B2 <sup>+</sup> , B2 / C1
	2	9	A1 seis veces, A2 tres veces
	3	9	B2 tres veces, A1, A2 / B1, B2 / C1, C1, dos profesionales no indican el nivel

Con este estudio se perseguía estimar el nivel aproximado de los ítems de la versión tradicional y el de los textos audiovisuales. En relación con los textos de entrada, en ambos casos existía un nivel sugerido por cinco de los nueve especialistas y otro indicado por cuatro de ellos. Como la distancia entre las dos opciones era de solo una unidad, consideramos que lo más apropiado era tener en cuenta ambos niveles. En cuanto a los ítems, se identificó un nivel predominante en todos los casos, excepto en los ítems 2, 4 y 7. La pregunta 2 no planteó complicaciones porque, en conjunto, A2 y B1 habían sido indicados por ocho de los nueve profesionales, y este grado de precisión era suficiente para los objetivos del test. Además, estos niveles fueron los más frecuentes en el estudio 4 (v. § 2.3.4.). En cambio, los ítems de comprensión de sentimientos 4 y 7 resultaron más problemáticos. En el ítem 4 se apreció una disparidad de respuestas que impedía determinar el nivel. Como solución, recurrimos a la propuesta del estudio 4. En el ítem 7 se observaban discrepancias pero destacaba el nivel B2, sugerido por cuatro de los nueve informantes. Dado que este mismo grado de dificultad había sido señalado por cuatro de los cinco profesores de ELE en el estudio 4, consideramos que se trataba del nivel más probable. De acuerdo con estos razonamientos, el nivel aproximado de los textos y los ítems era el siguiente:

Tabla 23. Nivel aproximado de los textos y los ítems.

TEXTO DE ENTRADA	ÍTEM	NIVEL ESTIMADO <sup>a</sup>
Secuencia de <i>Ópera prima</i>	1	A1
	2	A2 / B1
	3	A2
	4	A2 / B1
Secuencia de <i>Los peores años de nuestra vida</i>		B2 / C1
	5	B2
	6	A1
	7	B2

a. Cuando se duda entre dos niveles, se sugieren ambos. En estos casos, el primero corresponde al nivel más probable.

Esta investigación permitió sugerir el nivel aproximado de los textos y los ítems con un grado de indefinición de hasta dos niveles del *MCER*, tal como se perseguía. Además, en cuatro preguntas se consiguió un nivel de precisión mayor que en el estudio 4 (v. § 2.3.4.). No obstante, el estudio presentaba varias limitaciones. Estas emergían, fundamentalmente, de la aplicación parcial e incompleta de la metodología del *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Consejo de Europa, 2009). Se puede destacar la ausencia de actividades de familiarización con las escalas de comprensión auditiva y audiovisual del *MCER*, y de un taller de ilustración con ítems de comprensión auditiva / audiovisual estandarizados antes de especificar el nivel de los ítems y los textos. En estos casos, la amplia experiencia de los profesionales con el *MCER* y su reciente participación en el curso de veinticinco horas sobre la vinculación de exámenes al *MCER* pudieron atenuar en cierta medida estas lagunas. Por añadidura, por una cuestión de tiempo, no se pudo discutir el nivel de cada ítem hasta llegar a un consenso. Asimismo, el número de informantes se redujo a nueve.

### 3. CONCLUSIONES

Este trabajo forma parte de una investigación conducente al título de doctor que explora la utilización de preguntas de comprensión audiovisual integradas en la imagen del vídeo como subtítulos y sincronizadas con los fragmentos de vídeo relevantes para responderlas. Esta técnica podría aportar varios beneficios frente a visionados con los materiales de la actividad en papel. Entre ellos, podría reducir el conflicto de atención entre ver el vídeo y realizar la tarea, al aproximar espacial y temporalmente las preguntas y los fragmentos relevantes, y podría disminuir las necesidades cognitivas de la actividad, al no tener que prestar atención a varias preguntas de forma simultánea. Anteriormente se han publicado un marco teórico que describe esta técnica (Casañ Núñez, 2015b) y un ejemplo en una secuencia didáctica (Casañ Núñez, 2015a).

En este artículo se ha detallado el proceso de planificación, diseño y experimentación de un test de comprensión audiovisual con dos variantes: una tradicional y otra experimental. En ambas versiones, el test se entrega a los examinados en papel. En la experimental, además, los enunciados de las preguntas de comprensión audiovisual

aparecen integrados en la imagen a modo de subtítulos. El test está diseñado para administrarse junto a otros instrumentos en estudios cuasiexperimentales con grupos de control y tratamiento. Esencialmente, se pretende averiguar si la subtitulación de las preguntas facilita la comprensión, si aumenta el tiempo que los estudiantes miran en dirección a la pantalla y conocer la opinión del grupo de tratamiento sobre esta técnica.

En la fase de experimentación se efectuaron seis estudios. En el estudio 5 se pilotaron las dos versiones del instrumento. Participaron cuarenta y un estudiantes de ELE (veintidós en el grupo de control y diecinueve en el de tratamiento). Las observaciones de los informantes durante la administración de la prueba y su posterior corrección sugirieron que las indicaciones sobre la estructura del test, las presentaciones de los textos, la explicación sobre el funcionamiento de las preguntas subtituladas para el grupo experimental y la redacción de los ítems resultaron comprensibles. Los datos de las dos variantes de la prueba se sometieron a sendos análisis de facilidad, discriminación, fiabilidad y descriptivos. También se calcularon las correlaciones entre los test y dos tareas de un examen de comprensión auditiva. Los resultados obtenidos mostraron que las dos versiones de la prueba estaban preparadas para ser administradas. Este estudio contó con varias limitaciones (v. § 2.3.5.4.8). La principal fue que no se controló la amenaza a la validez interna conocida como difusión de tratamientos, puesto que los grupos efectuaron las pruebas de forma consecutiva. No obstante, se dieron dos circunstancias que reducían la probabilidad de que se produjera el intercambio de información. En primer lugar, había poco tiempo para que tuviera lugar la interacción porque los test se administraron de forma consecutiva y los informantes tenían horarios de clase distintos. En segundo lugar, las pruebas no tenían repercusión en la nota de la asignatura y, en principio, no existía un interés manifiesto por conocer su contenido.

El estudio 6 tenía el propósito de determinar el nivel aproximado de los textos y de los ítems en relación con el *MCER*. Para ello, se configuró un panel de nueve jueces y se aplicaron de forma parcial algunos principios *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Consejo de Europa, 2009). La investigación permitió determinar el nivel aproximado de los textos y los ítems con un grado de indefinición de hasta dos niveles del *MCER*. No obstante, el estudio presentó varias limitaciones. Estas procedían, fundamentalmente, de la aplicación parcial de la metodología del manual (v. § 2.3.6.4). Las lagunas pudieron ser parcialmente suplidas por la amplia experiencia de los profesionales con el *MCER* y su reciente participación en un curso sobre la vinculación de exámenes al *MCER*.

En el momento de envío de este artículo podemos anunciar que el instrumento ha sido utilizado en estudios cuasiexperimentales con grupos de control y tratamiento en la Universidad de Valencia y en la Universidad de Rostock.

## AGRADECIMIENTOS

Quiero expresar mi gratitud a las personas que han participado en los estudios: a Ana María Gimeno (Universidad Politécnica de Valencia), a María Virginia González (Universidad de Valencia), a Jaume Brines, Carmen González, Estrella Redondo, Adriana Repila, Vanessa Roca, María José Solomando y Francisco Teruel (Centro de Idiomas de la Universidad de Valencia), a Inês Machado, Sandra Santos y Verónica Marôco (estudiantes del máster en formación de profesores de portugués y ELE de la Universidad de Coimbra en el curso 2013/2014), a los estudiantes de las asignaturas *Español 5* y *Español II* ofrecidas por la Universidad de Coimbra en 2013/2014, y a los profesionales de la enseñanza y/o evaluación que colaboraron en el estudio 6, entre ellos (cito a los que indicaron su nombre), Valentín Cózar (Instituto Cervantes de Lisboa), Francisco del Moral (Universidad de Verona) y Carmen San Julián (Trinity College Dublin). Asimismo, deseo dar las gracias a los profesores que me ayudaron a comprender el funcionamiento del programa SPSS: a Jesús Miguel Jornet (Universidad de Valencia), a José Manuel da Silva (Universidad de Coimbra) y especialmente a Zélia Fernandes (Universidad Nova de Lisboa). También quiero manifestar mi gratitud a Gustavo Casañ (Universidad Jaume I) por la lectura atenta del artículo y a Diana Silver (Universidad de Coimbra) por la revisión del resumen en inglés. Por último, agradezco a *marcoELE* las sugerencias recibidas para mejorar este trabajo.

## BIBLIOGRAFÍA

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: the interface between learning and assessment*. Nueva York: Continuum.

Alderson, J.C., Clapham, C. y Wall, D. (1995). *Language test construction and evaluation*. Cambridge: CUP.

Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S. y Tardieu, C. (2006). "Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of The Dutch CEFR Construct Project", en *Language Assessment Quarterly* vol. 3 núm. 1, págs. 3-30.

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

Baralo Ottonelo, M. y Martín Leralta, S. (coords.) (ca. 2011). *LETRA-Diploma de lengua española para trabajadores inmigrantes. Guía del profesor*. Madrid: Universidad Nebrija: [http://www.diplomalettra.com/examen\\_letra/pdf/LETRA\\_guia\\_para\\_el\\_profesor.pdf](http://www.diplomalettra.com/examen_letra/pdf/LETRA_guia_para_el_profesor.pdf) (consulta: 01/03/2016)

Bejar, I., Douglas, D., Jamieson, J., Nissan, S. y Turner, J. (2000). *TOEFL 2000 Listening Framework: A Working Paper*. Princeton: Educational Testing Service: <http://ets.org/Media/Research/pdf/RM-00-07.pdf> (consulta: 01/03/2016)

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: bases y procedimientos*. Madrid: Arco/Libros

Briggs, S. R. y Cheek, J. M. (1986). "The role of factor analysis in the development and evaluation of personality scales", en *Journal of Personality* vol. 54 núm. 1, págs. 106-148.

Brown, G. y Yule, G. (1983). *Teaching the Spoken Language: An approach based on the analysis of conversational English*. Cambridge: CUP.

Buck, G. (2001). *Assessing Listening*. Cambridge: CUP.

Buck, G. y Tatsuoka, K. (1998). "Application of the rule-space procedure to language testing: Examining attributes of a free response listening test", en *Language Testing* vol. 15 núm 2, págs 119-157.

Casañ Núñez, J.C. (2009a). "Descripciones", en *Azulejo para el español* núm. 2, págs. 15-18:

<https://sede.educacion.gob.es/publiventa/detalle.action?cod=14146>

(consulta: 03/02/2016)

Casañ Núñez, J.C. (2009b). "Didáctica de las grabaciones audiovisuales para desarrollar la comprensión oral en el aula de lenguas extranjeras", en *marcoELE* núm. 9, 142 págs. (tesina de máster):

<http://marcoele.com/suplementos/didactica-de-las-grabaciones-audiovisuales>

(consulta: 03/02/2016)

Casañ Núñez, J. C. (2015a). "Subtitulación de preguntas de comprensión audiovisual: ejemplificación en una secuencia de *Ópera prima* de Fernando Trueba", en *Foroele* núm. 11, págs. 45-56:

<https://ojs.uv.es/index.php/foroele/article/view/7095>

(consulta: 03/02/2016)

Casañ Núñez, J. C. (2015b). "Un marco teórico sobre el uso de preguntas de comprensión audiovisual integradas en el vídeo como subtítulos: un estudio mixto", en *marcoELE* núm. 20, 45 págs. :

<http://marcoele.com/compreesion-audiovisual-y-subtitulos/>

(consulta: 03/02/2016)

Chaume Varela, F. (2001) "La pretendida oralidad de los textos audiovisuales y sus implicaciones en traducción", en Agost Canós, R. y Chaume Varela, F. (eds.). *La traducción de los medios audiovisuales*. Castellón: Universidad Jaume I, págs. 77-88.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates.

Coniam, D. (2001). "The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: a case study", en *System* núm. 29 págs. 1-14.

Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Trad. del Instituto Cervantes. Madrid: Anaya.

Consejo de Europa (2009). *Manual for relating language examinations to the Common European Framework of Reference for Languages*. Estrasburgo: Consejo de Europa.

Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Londres: Pearson

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: OUP.

Eguaras, M. (2014). *Qué tipografía usar en libros impresos y digitales*. Mariana Eguaras Consultoría Editorial (Edición Kindle).

Escandell Vidal, M. V. (1996). *Introducción a la pragmática*. Barcelona: Ariel Lingüística.

Field, A. P. (2009). *Discovering statistics using SPSS* (3ª ed.). Londres: Sage publications.

Field, J. (2008) *Listening in the language classroom*. Cambridge: CUP.

Figueras, N. (2014). "Procedimientos y procesos de vinculación de exámenes a sistemas de referencia externos". Seminario impartido en el curso *Desarrollo y gestión de exámenes vinculados al Marco común europeo de referencia* organizado por la Universidad Internacional Menéndez Pelayo y el Instituto Cervantes en Santander.

Flowerdew, J. y Miller, L. (2005). *Second language listening*. Nueva York: CUP.

Gozalo Ausín, A. (2013). "Retos de las EEOOII ante la LOMCE: niveles, currículo y certificaciones", en *Actas del VIII Congreso de Escuelas Oficiales de Idiomas*, Pamplona, 9 págs. :  
[https://dl.dropboxusercontent.com/u/60897855/Actas\\_VIII\\_Congreso\\_EEOOII/04.-Mesas\\_redondas-15/Retos\\_de\\_la\\_LOMCE/Gozalo-Retos\\_de las\\_EEOOII\\_ante\\_la\\_LOMCE\\_NIVELES\\_CURR%C3%8DCULO\\_Y\\_CERTIFICACIONES.pdf](https://dl.dropboxusercontent.com/u/60897855/Actas_VIII_Congreso_EEOOII/04.-Mesas_redondas-15/Retos_de_la_LOMCE/Gozalo-Retos_de las_EEOOII_ante_la_LOMCE_NIVELES_CURR%C3%8DCULO_Y_CERTIFICACIONES.pdf)  
(consulta: 01/03/2016)

Green, R. (2013). *Statistical Analyses for Language Testers*. Nueva York: Palgrave

Gruba, P. (1993). "A Comparison Study of Audio and Video in Language Testing", en *JALT Journal*/vol. 15 núm. 1, págs. 85-88.

Gruba, P. (1997) "The role of media in listening assessment", en *System* vol. 25 núm. 3, págs. 335-345.

Harris, T. (2003). "Listening with Your Eyes: The Importance of Speech-Related Gestures in the Language Classroom", en *Foreign Language Annals* vol. 36 núm. 2, págs. 180-187.

Hidalgo de la Torre, R. (coord.) (2013). *Prepara y practica el DELE B1*. Barcelona: Octaedro.

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

Lynch T. (2006). "Academic listening: Marrying top and bottom", en Martínez-Flor, A. y Usó-Juan, E. (eds.) *Current Trends in the Development and Teaching of the Four Language Skills*. Berlin: Mouton de Gruyter, págs. 91-110.

Lynch, T. (2009). *Teaching Second Language Listening*. Oxford: OUP.

Lynch, T. (2010). "Listening: sources, skills, and strategies", en Kaplan, R.B. (ed.), *The Oxford Handbook of Applied Linguistics* (2ª ed.). Oxford: OUP, págs. 74-87.

Lynch, T. (2012). "Traditional and Modern Skills. Introduction", en Eisenmann, M. y Summer, T. (eds.), *Basic issues in EFL teaching and learning*, Heidelberg: Winter, págs. 69-81.

Martín Leralta, S. (2011). "Certificación lingüística de nivel inicial para inmigrantes en contexto laboral: ejemplo de una prueba del examen DILE", en *Lengua y migración*, vol. 3 núm. 1, págs. 89-10:

<http://lym.linguas.net/Download.axd?type=ArticleItem&id=90>

(consulta: 01/03/2016)

Martín Peris, E. (1991). "La didáctica de la comprensión auditiva" en *Cable*, núm. 8. Madrid: Difusión. Rpt. en (2007) *marcoELE* núm. 5:

<http://marcoele.com/la-didactica-de-la-comprension-auditiva/>

(consulta: 03/02/2016)

Martínez Castillo, G. (2009). *Presentaciones Digitales y Comunicación Interactiva*. UNISON: México:

<http://clasificaciontipografica.blogspot.pt/>

(consulta: 03/02/2016)

Messick, S. (1996). "Validity and washback in language testing", en *ETS Research Report Series* vol. 1996 núm. 1, págs. 1-18.

Nunan, D. (1999). *Second language teaching and learning*. Boston: Heinle and Heinle.

Ockey, G.J. (2007). "Construct implications of including still image or video in computer-based listening tests", en *Language Testing* vol. 24 núm. 4, págs. 517-537.

O'Rourke, N. y Hatcher, L.(2013). *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling* (2ª ed.). Cary, Carolina del Norte: SAS Institute, Inc.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4ª ed.). Maidenhead: Open University Press/McGraw-Hill.

Pestana, M. H. y Gageiro, J. N. (2008). *Análise de Dados para Ciências Sociais: A Complementaridade do SPSS* (5ª ed.). Lisboa: Edições Sílabo.

Pastor Villalba, C. (2009). "La evaluación de la comprensión oral en el aula de ELE", en *marcoELE* núm. 9:

<http://marcoele.com/la-evaluacion-de-la-comprension-oral-en-el-aula-de-ele/>

(consulta: 01/03/2016)

Rabasco Macías, E. (2007). "Los peores años de nuestra vida", en *marcoELE* núm. 4, 38 págs.:

[http://marcoele.com/descargas/4/rabasco\\_peores.pdf](http://marcoele.com/descargas/4/rabasco_peores.pdf)

(consulta: 03/03/2016)

Rial, A. y Varela, J. (2008). *Estadística Práctica para la Investigación en Ciencias de la Salud*. Coruña: Netbiblo.

Richards, J.C. (1983). "Listening comprehension: approach, design, and procedure", en *TESOL Quarterly* vol. 17 núm. 2, págs. 219-240.

Richards, J.C. y Burns, A. (2012). *Tips for Teaching Listening*. Nueva York: Pearson Education.

Riley, P. (1979). «Viewing comprehension: "L'oeil écoute"», en *Mélanges CRAPEL* núm. 10, págs. 80-95.

Rost, M. (1990). *Listening in language learning*. Nueva York: Longman.

Rost, M. (2002). *Teaching and researching listening*. Harlow: Pearson Education.

Rubin, J. (1995). "An Overview to *A Guide for the Teaching of Second Language Listening*", en Mendelsohn, D. J. y Rubin, J. (eds.) *A Guide for the Teaching of Second Language Listening*. San Diego: Dominie Press, págs. 7-11.

Sans Martín, A. (2011). *La investigación de enfoque experimental*. Barcelona: Universidad Oberta de Cataluña.

Santos Gargallo, I. y Santos Gargallo, A. (2001). *De cine. Fragmentos cinematográficos para el aula de E/LE*. Madrid: SGEL.

Sen, A. y Srivastava, M. (1990). *Regression Analysis: Theory, Methods and Applications*. Nueva York: Springer-Verlag

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

Sperber, D. y Wilson, D. (1986). *Relevance. Communication and Cognition*. Oxford: Blackwell.

Sperber, D. y Wilson, D. (2004). "Relevance theory", en Horn, L.R. y Ward, G. (eds.). *The Handbook of Pragmatics*. Oxford: Blackwell, págs. 607-632.

Stempleski, S. y Tomalin, B. (2001). *Film*. Oxford: OUP.

Sueyoshi, A. y Hardison, D. (2005). "The Role of Gestures and Facial Cues in Second Language Listening Comprehension", en *Language Learning* vol. 55 núm. 4, págs. 661-669.

Surh, D. y Shay, M. (2009). "Guidelines for reliability, confirmatory and exploratory factor analysis", en *Actas de la Conferencia Anual de Western Users of SAS Software*. San José, California, 15 págs.

Suvorov, R. (2008). *Context visuals in L2 listening tests: The effectiveness of photographs and video vs. audio-only format*. Ames, Iowa: Iowa State University (tesina de máster).

Sweller, J., Ayres, P. y Kalyuga, S. (2011). *Cognitive Load Theory*. Londres: Springer

Tabachnick, B. G. y Fidell, L. S. (2007). *Using multivariate statistics* (5ª ed.). Boston: Pearson.

Trindade Natel, T. B. (2002). "La proximidad entre el portugués y el español, ¿facilita o dificulta el aprendizaje", en *Actas del XIII Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera*, págs. 825-832:  
[http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/asele/pdf/13/13\\_0825.pdf](http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/13/13_0825.pdf)  
(consulta: 01/03/2016)

Underwood, M. (1989). *Teaching Listening*. Londres: Longman.

Ur, P. (1999). *A Course in Language Teaching: practice and theory*. Cambridge: CUP.

Vandergrift, L. y Goh, C. C. M. (2012). *Teaching and Learning Second Language Listening*. Nueva York: Routledge.

Vizcaíno Rogado, I. (2008). "El cine para la clase de ELE: aprendizaje de español a partir de fragmentos cinematográficos", en *redELE* núm. 8, 140 págs. (tesina de máster):  
<http://www.mecd.gob.es/redele/Biblioteca-Virtual/2007/memoriaMaster/2-Trimestre/VIZCAINO-R.html>  
(consulta: 03/03/2016)

Wagner, E. (2002). "Video listening tests: A pilot study", en *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics* vol. 2 núm. 1:  
<http://tesol.columbia.edu/article/video-listening-tests/>  
(consulta: 01/03/2016)

Wagner, E. (2007). "Are they watching? Test-taker viewing behavior during an L2 video listening test", en *Language Learning and Technology* vol. 11 núm. 1, págs. 67-86:  
<http://llt.msu.edu/vol11num1/pdf/wagner.pdf>  
(consulta: 01/03/2016)

Wagner, E. (2008). "Video listening tests: What are they measuring?", en *Language Assessment Quarterly* vol. 5 núm.3, págs. 218–243.

Wagner, E. (2010). "Test-takers' interaction with an L2 video listening test", en *System* vol. 38 núm. 2, págs. 280-291.

Wagner, E. (2013). "Assessing listening", en Kunnan, A. J. (ed.). *The companion to language assessment*. Oxford: Wiley-Blackwell, págs. 47-63.

Weir, C. (1993). *Understanding and developing language tests*. London: Prentice Hall.

Whittaker, S. (1995). "Rethinking video as a technology for interpersonal communications: theory and design implications", en *International Journal of Human-Computer Studies* vol. 42 núm. 5, págs. 501-529.

NOMBRE:

Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

1. ¿De qué temas habla el chico pelirrojo?

- a) De la cena de Navidad y de una antigua novia.
- b) De la cena de Navidad y de pintura.
- c) De pintura y de su aspecto físico.
- d) De su aspecto físico y de una antigua novia.

2. ¿Cómo se llama la chica?

- a) Marina
- b) María
- c) Margarita
- d) Maite

3. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico pelirrojo por la chica? Justifica brevemente la respuesta. Puedes responder en español, portugués o inglés.

## ANEXO 2

[NOMBRE] *[Se pregunta cuando es necesario para trazar relaciones con otras informaciones]*

TAREA 2. Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

5. ¿De qué temas habla el chico pelirrojo?

- a) Del tiempo atmosférico y de una antigua novia.
- b) Del tiempo atmosférico y de pintura.
- c) De pintura y de su aspecto físico.
- d) De su aspecto físico y de una antigua novia.

6. ¿Cómo se llama la chica?

- a) Marina
- b) María
- c) Margarita
- d) Maite

7. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico pelirrojo por la chica? Justifica brevemente la respuesta. Puedes responder en español, [en...] *[también se mencionan las lenguas de los informantes conocidas por el investigador o por personas que puedan ayudarlo, v. § 2.1.9.]*

NOMBRE:

Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

1. ¿Cómo se llama la chica?

- a) Vanesa
- b) Violeta
- c) Verónica
- d) Virginia

2. ¿De qué hablan?

- a) la familia y los estudios de la chica
- b) la familia y el trabajo de la chica
- c) la casa y el trabajo de la chica
- d) la casa y los estudios de la chica

3. ¿Cuál es el teléfono de la chica?

4. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico por la chica? Justifica brevemente la respuesta. Puedes responder en español, portugués o inglés.

## ANEXO 4

[NOMBRE] *[Se pregunta cuando es necesario para trazar relaciones con otras informaciones]*

TAREA 1. Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

1. ¿Cómo se llama la chica?

- a) Vanesa
- b) Violeta
- c) Verónica
- d) Virginia

2. ¿De qué hablan?

- a) la familia y los estudios de la chica
- b) la familia y el trabajo de la chica
- c) la casa y el trabajo de la chica
- d) la casa y los estudios de la chica

3. ¿Cuál es el teléfono de la chica?

4. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico por la chica? Justifica brevemente la respuesta. Puedes responder en español, [en...] *[también se mencionan las lenguas de los informantes conocidas por el investigador o por personas que puedan ayudarlo, v. § 2. 1. 9.]*

NOMBRE:

TAREA 1. Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

1. ¿Cómo se llama la chica?

- a) Vanesa
- b) Violeta
- c) Verónica
- d) Virginia

2. ¿De qué hablan?

- a) la familia y los estudios de la chica
- b) la familia y el trabajo de la chica
- c) la casa y el trabajo de la chica
- d) la casa y los estudios de la chica

3. ¿Cuál es el teléfono de la chica?

4. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico por la chica? Justifica brevemente la respuesta. Puedes responder en español y en portugués.

NOMBRE:

TAREA 2. Escucha y observa el vídeo. En las preguntas de elección múltiple solo una opción es correcta.

5. ¿De qué temas habla el chico pelirrojo?

- a) Del tiempo atmosférico y de una antigua novia.
- b) Del tiempo atmosférico y de pintura.
- c) De pintura y de su aspecto físico.
- d) De su aspecto físico y de una antigua novia.

6. ¿Cómo se llama la chica?

- a) Marina
- b) María
- c) Margarita
- d) Maite

7. Considerando todo el fragmento, ¿qué sentimientos puede tener el chico pelirrojo por la chica? Justifica brevemente la respuesta. Puedes responder en español y en portugués.

FECHA DE ENVÍO: 7 DE MARZO DE 2016