

## LA CALIBRACIÓN DE LA COMPRENSIÓN LECTORA DENTRO DEL EXAMEN NACIONAL DE ELE SEGÚN EL MCER: ALGUNAS CUESTIONES DE EVALUACIÓN

### 1 INTRODUCCIÓN

El objetivo del presente trabajo es doble: por un lado, pretende presentar los resultados del proyecto concreto de calibración de la prueba de la comprensión lectora del examen de *matura*<sup>1</sup> para ganar una mayor equiparación con los niveles de referencia del *Marco Común Europeo de Referencia*. Por otro lado, y para garantizar una mayor transparencia de los resultados de calibración del examen según los niveles del MCER, introduce un análisis contrastivo de los resultados de calibración obtenidos mediante dos metodologías: la estadística – con los procedimientos de estandarización acreditados, y la descriptiva – con las escalas descriptivas del MCER especificadas en el manual *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. A manual* (Consejo de Europa 2009)<sup>2</sup>.

### 2 LA EVALUACIÓN EN EL MARCO COMÚN EUROPEO DE REFERENCIA PARA LAS LENGUAS

La rápida adopción del MCER tanto en el campo de la docencia como en el de la evaluación de lenguas ha mostrado la necesidad de tal documento en la actual Europa plurilingüe. Pero el documento no es un instrumento de evaluación prescriptivo, ya que no propone los contenidos, métodos o modalidades que deberían adoptar las evaluaciones de las lenguas extranjeras, sino funciona como «punto de referencia, [...] un esquema operativo más simple que sintetiza las categorías separadas del *Marco*» (Consejo de Europa 2002: 178). De ahí que los criterios para evaluar la competencia comunicativa en español vengán definidos en el MCER por las escalas de los descriptores correspondientes a cada una de las competencias (lingüísticas y no lingüísticas) que se cruzan en un acto comunicativo. A pesar de su carácter consultivo y no regulativo, las instituciones

---

\* *Dirección de las autoras*: Filozofska fakulteta, Univerza v Ljubljani, Aškerčeva 2, 1000 Ljubljana, Eslovenia. E-mail:marjana.sifrar@ff.uni-lj.si, andreja.trenc@ff.uni-lj.si.

- 1 El examen de *matura* es una prueba escrita y oral que se realiza en Eslovenia a los estudiantes que deseen acceder a los estudios universitarios. Los objetivos y el formato del examen se detallan en el capítulo 4 del presente trabajo.
- 2 De aquí en adelante: el *Manual*. También nos referimos a su versión abreviada *Highlights from the Manual* (Consejo de Europa 2011).

de toda Europa manifestaron su vinculación a los distintos niveles del MCER (A1-C2). Dado que muchos países ya disponían de pruebas propias para la evaluación de idiomas, el Consejo de Europa ha realizado esfuerzos importantes para establecer correctamente las correspondencias entre las pruebas y los estándares del MCER. Con estos fines se ha elaborado el *Manual* (directrices del Consejo de Europa, la Asociación Europea para la examinación y evaluación de lenguas – EALTA *Guidelines for Good Practice in Language Testing and Assessment* 2010), que incluye procedimientos de *standard setting*<sup>3</sup> aplicados al contexto del MCER. Los autores del *Manual* definen claramente la necesidad de contar con un sistema de certificación de calidad antes de empezar con el proceso de comparabilidad con los niveles del MCER. En la versión final del *Manual* (Consejo de Europa 2009) esta necesidad se concreta en tres fases: 1) la fase de especificaciones (concretar lo que se evalúa), 2) la fase de validez y fiabilidad como instrumentos imprescindibles, y 3) la fase de validez de las notas de corte o fijación de estándares.

Este manual se concibe como una ayuda para todos aquellos que quieran realizar este proceso con cierto rigor y garantías. Es un texto exhaustivo que aborda todas las fases y dimensiones de un correcto proceso de estandarización, pero que no conlleva una prescripción por parte del Consejo de Europa, sino una ayuda útil (García et al. 2012: 7). Ilc y Stopar (2014: 16) en su artículo sobre la validación del examen de comprensión lectora del examen nacional esloveno de inglés indican la suma importancia de los manuales, la formación y otras herramientas implementadas en relación con el MCER para así minimizar la influencia cultural, nacional y formativa de los jueces.

El fin último de este proceso de valoración o comparabilidad entonces no es el certificado oficial o un sello del Consejo de Europa que discrimine el nivel de una prueba, sino una aplicación ética, responsable y coherente del MCER y del *Manual* acorde con las necesidades de una sociedad.

### 3 LA COMPRESIÓN LECTORA

Desde los años 60 o 70 del siglo pasado, la comprensión de textos (escritos y orales) ha sido objeto de estudio principalmente en las áreas de psicología. Con el desarrollo del análisis de comunicación, la comprensión como actividad verbal comunicativa pasó a cobrar una mayor importancia en diferentes áreas: psicolingüística, neurolingüística, lingüística del texto, teoría del discurso, inteligencia artificial, y sobre todo, pragmática (Biere 1991: 1). En el centro de interés de todas estas disciplinas está el significado que el lector o intérprete<sup>4</sup> interpreta durante la actividad comunicativa de la comprensión (Ferbežar/Stabej 2008: 17). Los intérpretes pueden comprender un mismo texto de varias maneras, ya que el texto -según Halliday y Widdowson (en Alderson 2000: 6)- no contiene un significado determinado para ser despojado por el intérprete, sino sólo un potencial significativo que genera el significado en cada una de las interacciones entre

---

3 Se denomina como *standard setting* (establecimiento de estándares) al procedimiento de establecimiento de puntos de corte en las puntuaciones de los tests (Cizek/Bunch 2007).

4 Ferbežar y Stabej (2008) denominan al lector como intérprete.

el texto y su intérprete. La comprensión de textos en una lengua extranjera estudia esta misma interacción desde la perspectiva del acercamiento a un código perteneciente a otra lengua y otras culturas. En este sentido, la didáctica en su vertiente contrastiva entre L1/lengua extranjera ha intentado proponer una nueva reflexión sobre la creencia comúnmente aceptada en la lingüística tradicional de que «el conocimiento de la lengua ya forma una condición suficiente para la comprensión» (Ferbežar/Stabej 2008: 15).

El proceso comunicativo de lectura en una lengua extranjera difiere fundamentalmente del mismo proceso en la lengua materna en cuanto al tipo y al nivel de conocimientos (lingüísticos y psicolingüísticos) que el individuo emplea durante la lectura; el primer contacto con la lectura en la lengua materna presupone del niño un dominio considerablemente alto de la lengua y la falta de la destreza de decodificación; mientras al niño le quedan por aprender estas destrezas, un aprendiz de lenguas extranjeras ya entra equipado con el conocimiento sobre el proceso de decodificación y las estrategias, ya que lo obtuvo durante el proceso de la adquisición de la lengua materna, mientras que está todavía en el proceso de la adquisición del sistema de la nueva lengua. La discrepancia entre las estrategias de lectura y el dominio de la lengua objeto determina los factores que influyen sobre el proceso de comprensión en una lengua extranjera a diferencia de la materna.

Mikeln (2008: 387-398) menciona tres factores: los factores lingüísticos, esto es, el procesamiento de los textos, los factores individuales y de experiencia, y los factores socioculturales. La capacidad de lectura en una lengua extranjera no puede por lo tanto ser entendida por completo sin la implicación íntegra de procesos heurísticos, cognitivos, psicolingüísticos por encima de la básica capacidad de decodificación del sistema lingüístico del idioma. Un número considerable de trabajos sobre los procesos de comprensión de una lengua extranjera (Acquaroni Muñoz 2003; Gospodarič/Šifrar Kalan 2008; Lah 2012; Kintch, Van Dijk 1983) han coincidido en detectar estos procesos como lo que Alderson (2000: 33-40) define como «una serie de conocimientos lingüísticos, el reconocimiento de clases morfológicas, asimismo como el conocimiento metalingüístico, metacognición, el conocimiento de la temática, el conocimiento general sobre el mundo o la experiencia y el conocimiento sociocultural» y han consolidado una clara correlación con las competencias comunicativas no lingüísticas o generales del MCER. No deberíamos dejar al lado asimismo la competencia ortoépica, una competencia subestimada en las clases de LE según subraya Lah (2013).

#### **4 EL EXAMEN DE *MATURA* DE ESPAÑOL COMO LENGUA EXTRANJERA EN ESLOVENIA**

En Eslovenia, desde el año 1995 y con la fundación del Centro Nacional de Exámenes de Eslovenia (Državni izpitni center), los estudiantes de bachillerato concluyen sus estudios de la enseñanza secundaria con un examen nacional objetivo llamado *matura* que consta de una parte obligatoria y otra optativa. La modalidad optativa le posibilita al candidato examinarse, entre otras asignaturas, también de ELE. Este examen nacional tiene un doble objetivo: es, por una parte, el examen de aprovechamiento para medir los conocimientos conseguidos, y, por otra, el examen de dominio de nivel o de selección que le sirve para seguir con los estudios universitarios en universidades

públicas, y forma parte de las Pruebas de acceso a estudios universitarios. El candidato se somete a una parte objetiva (comprensión lectora y auditiva, y uso de la lengua) y a una parte subjetiva (expresión oral y escrita) en el examen.

El formato de las pruebas tiene como objetivo general evaluar la capacidad comunicativa del hablante sea al nivel de percepción de textos escritos y orales, sea a nivel de emisión de mensajes. Este enfoque eminentemente comunicativo toma como modelo las directrices sobre la evaluación comunicativa recogida en el diseño de pruebas de la Universidad de Cambridge (Bachmann 1990; Alderson 1995; Swain 1985). A pesar de la dificultad de la adopción completa de un enfoque comunicativo en las pruebas de la comprensión lectora objetivas a través de unos ítems discretos, pues cada examen evalúa solo una simulación y no una actuación real comunicativa, se intenta garantizar la autenticidad de la prueba a partir de las generalizaciones del dominio comunicativo en una serie de ámbitos, taxonomías<sup>5</sup> y tipologías textuales que se consideran factibles y significativas en el contexto educativo del alumno. La autenticidad de la prueba está garantizada también con la reproducción de una situación lectora en su dimensión compleja, donde se emplean las mismas estrategias metacognitivas que el candidato utilizaría durante la decodificación real de los textos: conocimientos generales y estrategias de comprensión, como la anticipación, lectura global (*scanning*) y detallada (*skimming*), lectura intensiva e inferencia o deducción (Gospodarič/Šifrar Kalan 2008).

## 5 EL PROYECTO DE CALIBRACIÓN DEL EXAMEN DE *MATURA* DE ESPAÑOL SEGÚN LAS ESCALAS DEL MCER

Respetando el modelo de la política lingüística europea sobre la evaluación de las lenguas extranjeras recogidas en el *Manual* (Consejo de Europa 2009) se ha creado también en Eslovenia la necesidad de calibrar el nivel de los candidatos en los exámenes nacionales de lenguas extranjeras elaborados por el Centro Nacional. A base de esta iniciativa y con el objetivo de crear una transparencia equiparable internacionalmente de los niveles de los exámenes de *matura* de lenguas extranjeras, se emprendió en 2008 el proyecto de calibración de los exámenes de *matura* para todas las lenguas extranjeras según el MCER. El proyecto está siendo llevado a cabo de acuerdo con métodos aceptados internacionalmente y procedimientos y estándares para las lenguas extranjeras recopilados en el *Manual*.

El proyecto de calibración del examen de español se está implementando en dos fases sucesivas en las que se realizarán cuatro procedimientos de la calibración. La primera fase, realizada entre finales del año 2013 y mediados del año 2014, consistió en procedimientos introductorios de familiarización con los métodos e instrumentos de calibración, especificación y estandarización de los dos niveles de pruebas de comprensión lectora. La segunda fase, planificada para finales del 2014, comprende la especificación y estandarización de un ejemplar de prueba de *matura*, esa vez de todas las partes del examen (incluso la expresión oral), así como los procedimientos finalizadores de validación. La

---

5 El diseño de tareas baraja diferentes modalidades definidas por el Catálogo curricular de los contenidos del examen (*Predmetni izpitni katalog*).

finalización del proyecto se prevé para el año 2015 cuando se haya calibrado todo el examen de bachillerato de ELE según los estándares europeos y se haya llegado a una estimación final del nivel de la examinación según las escalas del MCER.

## 6 METODOLOGÍA DE LA CALIBRACIÓN

Una de las finalidades del *MCER* es ayudar a los usuarios a describir los niveles de dominio lingüístico exigidos por los exámenes y programas de evaluación existentes, con el fin de facilitar las comparaciones entre distintos sistemas de certificados. Con este propósito se han desarrollado el esquema descriptivo y los niveles comunes de referencia. Entre los dos, se proporciona un cuadro conceptual que los usuarios pueden utilizar para describir su sistema (Consejo de Europa 2002: 23).

El proceso de relacionar un examen con el MCER contiene cinco conjuntos de procedimientos que no deben seguirse lineal sino cíclicamente, siempre y cuando el examen respete los criterios de validez y confiabilidad: 1) familiarización de los involucrados en el proyecto con el MCER, con sus niveles y descriptores; 2) especificación de los contenidos y estructura del examen; 3) prácticas de estandarización (comprensión de niveles) y *benchmarking*; 4) estandarización; 5) validación (Consejo de Europa 2011: 16–18).

Para calibrar cada examen de LE se formó un panel de jueces compuesto por profesores de escuelas secundarias, profesores universitarios, diseñadores y evaluadores externos de exámenes, junto a una coordinadora profesional del proyecto. El proyecto empezó con unos seminarios de familiarización con los procedimientos y herramientas de calibración de exámenes de lenguas para todos los equipos usando las muestras del examen de matura de inglés. A continuación el procedimiento se realizó en parte en equipos de diferentes LE, con debates sobre el tema, y en parte individualmente. En la fase de familiarización con los descriptores específicos de comprensión lectora (MCER), los jueces del equipo español tuvieron que aceptar dichos descriptores a pesar de algunos desacuerdos con ellos, ya que solo el mutuo consentimiento posibilitó el progreso de la calibración. Por lo tanto, es imprescindible una discusión en grupo después de haber clasificado individualmente cada ítem y haberlo registrado. Cabe destacar que cada reunión empezó con unos ejercicios de conocimiento de descriptores, ya que solo así se puede garantizar una estandarización fiable y válida. En la fase de estandarización se pudo elegir entre varios métodos (*The body of work method*, *The Tucker-Angoff method*, *The Basket method*, *The Bookmark method*). Los métodos escogidos en este proyecto, *the Tucker-Angoff method* y *the Basket method*, se enfocan en los ítem-tests. De acuerdo con los métodos de estandarización, «no hay un único procedimiento adecuado para establecer puntos de corte, ya que incluso un mismo método puede ser aplicado de muy diversas maneras, dependiendo de las condiciones concretas en que se ejecuta. No es tanto el método concreto que se emplea, de hecho, sino el conjunto de decisiones adoptadas, lo que hacen de un proceso de *standard setting* un buen o mal procedimiento» (Cizek/Bunch 2007 en García et al. 2012: 9).

El más utilizado entre los métodos centrados en las pruebas (ítems o tareas) es el método *Tucker-Angoff* (Consejo de Europa 2009: 61). Los jueces deben analizar cada

ítem o tarea y estimar la probabilidad (expresada en %) que tendría un candidato mínimamente competente de resolver un ítem. La nota de corte se fija una vez hecha la media de los juicios agregados de cada juez. En nuestra versión del método el panel de diez jueces tenía que imaginar cuántos de los 100 candidatos límite (A2/B1, B1/B2) podrían dar la respuesta correcta. El método *Basket* está enfocado en los descriptores del MCER. El panel de los jueces tiene que responder a la pregunta: ¿a qué nivel del MCER un candidato hipotético ya puede dar con la respuesta correcta para cada ítem? Es decir, cuál es el nivel mínimo suficiente para responder correctamente. Eso no quiere decir que los candidatos de niveles inferiores no puedan dar soluciones correctas, sino que se sobre entiende que los candidatos de niveles avanzados van a dar la respuesta correcta.

El panel de los jueces evaluó los ítems de comprensión lectora en dos rondas: la primera sin ver los resultados de los test y la segunda dentro de una semana con los datos revelados de la primera ronda y los índices de dificultad, calculados en el Centro Nacional de Exámenes. Se calculó la media para cada ítem, así que en el método *Angoff* la media de 80 significaría que 80 de 100 candidatos límite resolverían el ítem. En el método *Basket* se usaron los números de 1 a 6 para marcar los 6 niveles de A1 a C2 (A1 siendo 1). Con este método una media de, por ejemplo, 2,2 significaría que se trata de un ítem A2 o un 3 significaría que el nivel suficiente en este ítem para dar con la respuesta correcta sería B1. Si la mayoría de los ítems exige un nivel de B1, entonces se puede concluir que el examen certifica este nivel. En los dos métodos los jueces se apoyaron en los descriptores de comprensión lectora del MCER.

Las hipótesis en esta correlación o calibración de comprensión lectora en los exámenes de madura de español LE con los niveles del MCER son: 1) El examen de nivel básico certifica el nivel B1 y determina asimismo el grado alcanzado por la mayoría de los candidatos de este nivel; 2) El examen de nivel avanzado certifica el nivel B2 y determina asimismo el grado alcanzado por la mayoría de los candidatos de este nivel.

## 7 RESULTADOS

La prueba del nivel básico<sup>6</sup> se compone de tres textos auténticos (3 tipos de tareas) con 19 ítems y 19 puntos en total. Se presentaron 189 candidatos y consiguieron un 89% de promedio de soluciones correctas (16,87 puntos de los 19). Con el método *Angoff* se ha tratado de marcar el límite entre A2 y B1. Según las evaluaciones de la segunda ronda el corte es del 79% (15 de 19 puntos) para B1. Con el método *Basket* se ha evaluado el nivel 2,7 (2=A2; 3=B1) y 17,56 puntos para el mismo nivel. Tomados en cuenta los dos métodos, el corte para el nivel B1 es 16 puntos o 84% de todos los puntos del examen de comprensión lectora. Por lo tanto, en nuestra muestra de candidatos examinados el 74% de ellos han conseguido el nivel B1 lo que viene a confirmar la primera hipótesis.

La prueba del nivel avanzado se compone de dos textos auténticos (2 tipos de tareas) con 15 ítems puntuados con 21 puntos en total (una de las tareas puntuó el doble

---

6 Las pruebas se pueden consultar directamente en la página [http://www.ric.si/splosna\\_matura/predmeti/spanscina/](http://www.ric.si/splosna_matura/predmeti/spanscina/), convocatoria del 12 de junio del 2013.

cada uno de los ítems correctos). Se presentaron 196 candidatos y consiguieron un 86% de promedio de soluciones correctas (18 puntos de los 21). Con el método *Angoff* se ha tratado de marcar el límite entre B1 y B2. Según las evaluaciones de la segunda ronda el corte es de 81% (17 de 21 puntos) para B2. Con el método *Basket* se ha evaluado el nivel 3,35 (3=B1; 4=B2) y 20,8 puntos para el mismo nivel. Tomados en cuenta los dos métodos, el corte para el nivel B2 es 19 puntos o 90% de todos los puntos del examen de comprensión lectora del nivel avanzado. Por lo tanto, en nuestra muestra de candidatos examinados 58% de ellos han conseguido el nivel B2 y se puede confirmar la segunda hipótesis, pero solo en parte, ya que el número de ítems superiores al nivel 3,5, ronda en torno a los 7 de los 15 ítems en total, ubicando la gran mayoría de los ítems un poco por debajo del B2.

En los dos niveles del examen, la primera ronda del método *Angoff* ha proporcionado un corte más bajo que la segunda ronda, mientras que en el método *Basket* no ha habido discrepancia sustancial entre la primera y la segunda ronda. Se ha demostrado, igual que en la estandarización de otros exámenes (Bitenc Peharc/Tratnik 2014), que el método *Basket* proporciona resultados menos rígidos, esto es, más indulgentes. Es importante subrayar que el uso de dos métodos independientes aporta mayor validez externa. El método *Angoff*, a pesar de basarse en la valoración subjetiva de cada juez de quién es el candidato límite, es considerado como el método más frecuentemente usado y científicamente investigado entre los métodos de estandarización (Bitenc Peharc/Tratnik 2014: 10).

## 8 DISCUSIÓN

El panel de jueces ha llegado a la conclusión de que la metodología utilizada ha dado unos resultados por debajo del nivel esperado, pero los descriptores generales (Consejo de Europa 2002: 72) han resultado imprecisos para calibrar la comprensión lectora, puesto que no abarcan todas las competencias comunicativas a las que se somete dicha destreza. Partiendo de la concepción de la comprensión lectora como una destreza interdisciplinaria e integradora de varias subcompetencias o subdestrezas, hemos usado los criterios de selección de las (sub)competencias para evaluar varios aspectos de la comprensión lectora adoptados por el *Manual* (Consejo de Europa 2011: 42–43) que no están incluidos en los descriptores generales usados en la estandarización. Esta selección, a diferencia de las escalas generales del nivel de la comprensión lectora como una actividad global, presenta la comprensión lectora desde una perspectiva multidimensional y tiene en cuenta unos criterios más precisos y detallados que se ponen a prueba diferentes competencias generales y lingüísticas:



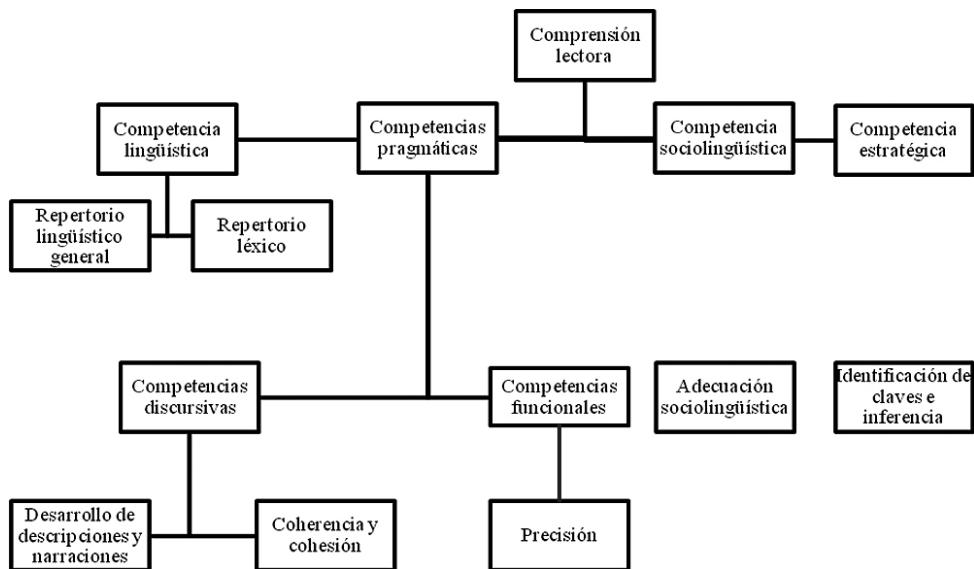


Figura 1: Subcompetencias integradas en la comprensión lectora según el *Manual* (Consejo de Europa 2011: 42–43)

La estimación del nivel de la actuación comunicativa de un aprendiz (durante la actividad lectora, en este caso) depende tanto de los componentes lingüísticos que el candidato emplea durante la comprensión como también de las competencias generales que «contribuyen de una forma u otra a la capacidad comunicativa del alumno» y se pueden considerar por lo tanto «solo aspectos de la misma» (Consejo de Europa 2002: 99). De ahí que, en una nueva fase experimental de especificación, elaborada con fines de una comparación de los resultados obtenidos por el método adoptado de *Angoff* y *Basket*, y la estimación del nivel de cada ítem según los descriptores de las siete competencias propuestas en el *Manual*, se llegue a una evaluación más rigurosa y exhaustiva. Partimos de la hipótesis de que el nivel especificado según los descriptores generales de la comprensión lectora en el MCER no se distingue sustancialmente de la evaluación del nivel usando los descriptores más detallados propuestos por el *Manual* sino que el uso de éstos contribuye a una mayor transparencia del proceso de la comprensión lectora. Asimismo, consideramos que una revaloración de criterios de calibración y una descripción más detallada de la destreza lectora podrían contribuir a una mayor precisión y fiabilidad de la evaluación según el MCER.

A continuación se presentan los ejemplos seleccionados de los ítems que han sido sometidos a una doble especificación, la estadística basada en los descriptores generales y los métodos *Angoff* y *Basket*, y en segundo lugar a la especificación según los descriptores detallados propuestos por el *Manual*.



**Examen de nivel básico, ítem 1.3.** La comprensión lectora y la resolución correcta del ítem se centran en la interpretación semántica correcta de la colocación *darse cita*. Para poder encontrar el vínculo semántico entre esta expresión y su interpretación sinónima en la respuesta correcta (*comer y hablar con los artistas*), el candidato tiene que recurrir a un repertorio de vocabulario frecuente y propio a las actividades cotidianas (relacionadas en este caso, con el ocio, la concertación de citas y el tiempo libre) lo que exige el nivel de competencia léxica de A2 según el descriptor en cuestión: «Tiene suficiente vocabulario para desenvolverse en actividades habituales y transacciones cotidianas que comprenden situaciones y temas conocidos» (Consejo de Europa 2002: 109). Asimismo se pone en marcha la competencia estratégica de inferencia al entender otra clave textual léxica (*los artistas*) como el sinónimo o el hiperónimo de *músicos, fotógrafos, diseñadores y demás gente de mal vivir*, puesto que «sabrán cómo utilizar una idea del significado general de textos y enunciados cortos que tratan temas cotidianos concretos para inferir del contexto el significado probable de las palabras que desconoce» (Consejo de Europa 2002: 74). Esta estimación descriptiva del nivel de competencias léxicas y estratégicas (A2) se vino a comprobar también con los métodos estadísticos ya que según la metodología *Angoff* el 83,75 % de los candidatos límites serían capaces de resolver el ítem a un nivel A2 (estimado por el método *Basket* como 2,3).

**Examen de nivel básico, ítems 2.2. y 2.3.** Para el análisis de los dos ítems en cuestión se ha tenido en cuenta sobre todo la competencia discursiva, puesto que se favorece por la propia tipología textual de la prueba (una entrevista con intervenciones desordenadas). Los dos ítems corresponden a la competencia discursiva en el nivel B1 porque el candidato tiene que relacionar secuencias discursivas sencillas y «unos elementos breves y concretos [...] para crear una secuencia cohesionada y lineal» (Consejo de Europa 2002: 122), como por ejemplo, la refutación o la comprobación de una afirmación incluida en el ítem 2.2 o la propuesta de un juicio de valor a una pregunta abierta que se plantea en el ítem 2.3.

El léxico que se debe activar para la resolución del ítem abarca el repertorio de las nociones relacionadas con el tiempo y los fenómenos meteorológicos (la nieve, las condiciones climáticas) y el tiempo libre (las actividades y espacios deportivos) que forman parte de un área inmediato, cercano y de interés para el candidato. De ese modo el dominio estimado de la competencia léxica corresponde a un nivel B1 donde el candidato tiene suficiente vocabulario para comprender la mayoría de información acerca de «los temas pertinentes para su vida diaria como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad» (Consejo de Europa 2002: 109). La calibración de los ítems con el método *Basket* ha coincidido en el nivel de la dificultad dando un B1 o A2+ (en el ítem 2.3. por debajo de la dificultad del segundo cifrándose en un 2,5 por plantear una respuesta más directa a la pregunta). La comparación de los resultados obtenidos con los dos métodos muestra una correlación con el método *Angoff* (78,75% y 85,62%), sin embargo, se ha descartado de los resultados el factor de la competencia discursiva que también influye en el proceso de la resolución de los ítems y, como consecuencia, también en el nivel de dificultad estimado.

**Examen de nivel básico, ítem 3.2.** En muchos ítems se ha puesto de relieve la importancia de la competencia estratégica de inferencia, uno de los aspectos clave de la competencia comunicativa y la actividad de lectura que, por otro lado, no se incluye en los descriptores generales de la comprensión lectora. La competencia estratégica estimada del ítem comprende una variedad de estrategias muy complejas y variadas (la inferencia del significado por el contexto o por el andamiaje con el léxico de otras lenguas extranjeras que se tienen dominadas) que se articulan para descifrar la semántica de las palabras desconocidas<sup>7</sup> en este ítem por lo que el nivel de la competencia estratégica exigido sea alto y correspondiente al nivel B2 (Consejo de Europa 2002: 74).

Del mismo modo, la competencia sociolingüística corresponde a un nivel B2, ya que para resolver el ítem correctamente es necesario entender y desenvolverse, aparte de en el registro informal, también en «el registro formal, que sea adecuado a la situación» como propone el descriptor del nivel B2 para esta competencia (Consejo de Europa 2002: 119).

Para poder hacerse con este ítem complejo cuya considerable dificultad se comprobó en el examen con el índice de dificultad 0,57, además también durante el proceso de estandarización de *Angoff* (sólo el 55,6% de los candidatos límite podrían resolver con éxito este ítem) y *Basket* (3,5 correspondientes a un nivel B1+), el candidato debería disponer de un vocabulario bastante abstracto y complejo. Su repertorio lingüístico general debería superar el nivel B1, disponiendo de «suficientes elementos lingüísticos como entender situaciones impredecibles, para captar los puntos principales de una idea o un problema con razonable precisión y para comprender pensamientos sobre temas abstractos o culturales [...]» para ubicarse en el nivel medio-alto (B1+) (Consejo de Europa 2002: 107).

**Examen de nivel alto, ítem 2.5.** La competencia gramatical involucrada en la comprensión lectora se basa en la interpretación correcta de estructuras sintácticas complejas de perífrasis verbales que aparecen dos veces en el texto en la función terminativa (*dejar de servir, dejar de ser útil*). El manejo de estas estructuras gramaticales complejas manifiesta «un grado relativamente alto del control gramatical» (B2) que ya trasgrede un «repertorio de estructuras o «fórmulas» habituales relacionadas con las situaciones más predecibles» (Consejo de Europa 2002: 111) característico de unos niveles inferiores (B1 en este caso) donde el aspecto terminativo de una acción suele expresarse a través de fórmulas fijas lexicalizadas en la semántica léxica o el uso de los tiempos verbales perfectivos (*ya no sirvió más*, p. ej.).

Del mismo modo, el candidato debe poner en marcha la competencia sociolingüística y la competencia léxica a un nivel B2 para entender adecuadamente un contexto formal (Consejo de Europa 2002: 119), por un lado la noción antagónica que existe en el mundo real entre la palabra *invasores* y *el pueblo propio*, y, por otro relacionarla con el significado léxico de estas dos expresiones. El grado del vocabulario prueba un amplio repertorio de nociones «sobre asuntos relativos a su especialidad y temas más generales» (Consejo de Europa 2002: 109). La comparación con la evaluación del nivel de competencia en las

7 El ítem requiere relacionar la frase siguiente del texto auténtico *Un hombre es exculpado tras 35 años en la cárcel: La sala irrumpió en una ovación que certificaba la reparación de una injusticia* y la frase: *El público estaba de acuerdo con la libertad de James*, marcando si esto es verdadero o falso.

escalas generales de la comprensión lectora mostró una correspondencia entre el nivel estimado según los descriptores detallados de competencias lingüísticas (B2) y el resultado obtenido con los métodos estadísticos (un 4 ó B2 según el método *Basket* o el 72,14% de éxito entre los candidatos límites entre B1/B2 en la metodología de *Angoff*).

Tanto las herramientas estadísticas como los descriptores detallados de la calibración han coincidido en el nivel del examen estimado: B1 para el nivel de prueba básico y B1/B2 para el nivel de prueba alto. En la discusión se ha comprobado que el panel de jueces ha insistido durante la estandarización en la aplicación de los descriptores detallados según el *Manual* y en la mayoría de las veces durante el proceso sólo se ha visto satisfecho con un enfoque analítico abarcando todas las competencias. Desde esta perspectiva analítica el panel de jueces ha podido ofrecer una reflexión sobre unos criterios más precisos para tener en cuenta durante las futuras elaboraciones de exámenes con el fin de garantizar una mayor validez de pruebas de comprensión lectora y una mayor adecuación a los niveles del MCER.

## 9 CONCLUSIÓN

La aplicación de unos métodos de calibración más detallados y amplios, que abarquen la comprensión lectora como una actividad tridimensional integradora de varias (sub)competencias generales y lingüísticas, proponen unos estándares alternativos para tener en cuenta durante las futuras calibraciones de exámenes de lengua extranjera según el MCER, de cara a garantizar una mayor correspondencia y validez entre los aspectos que se ponen a prueba durante la prueba real y los resultados obtenidos estadísticamente. Los equipos de todos los ámbitos involucrados en el proceso de evaluación, elaboración o estandarización de exámenes de lenguas extranjeras (los jueces del panel, los comités organizativos y los elaboradores de exámenes) deberían «cruzar los descriptores seleccionados entre las tres escalas para obtener una visión multidimensional de lo que tienen que hacer los alumnos en esos ámbitos y a ese nivel» (Pernas Izquierdo 2012: 37–38)<sup>8</sup>.

---

8 Pernas Izquierdo (2012: 37–38) en su artículo sobre diversificación curricular y plurilingüismo, compara el trabajo del programador curricular al del Jefe de la Unidad de Pre-Crimen, John Anderton, protagonista de la película *Minority Report* de S. Spielberg (2002) «que busca evidencias que puedan servirle para identificar a los asesinos antes de que cometan el crimen. Esta unidad cuenta con un novedoso sistema de prevención de delitos: tres *pre-cogs*, seres que ven el futuro, dan información a la policía, en forma de archivos o imágenes, sobre los homicidios que van a cometerse. Cada uno de los *pre-cogs* genera unas visiones distintas, pero complementarias, del mismo delito y el agente Anderton prevé tridimensionalmente la escena de crimen [...]. La visión de cada *pre-cog* representa cada una de las escalas del *Marco* (actividades de la lengua, estrategias y competencias). Las visiones de estos seres aportan una perspectiva complementaria del mismo evento delictivo, al igual que cada una de las escalas nos da una dimensión distinta del mismo fenómeno, la lengua en uso. De la misma manera que el agente Anderton analiza las imágenes generadas, [...] así el programador de un curso de idiomas extrae de cada una de las escalas los datos que considera más relevantes y los cruza para obtener una visión multidimensional de cómo se tienen que comportar, comunicativamente hablando, los aprendices de lenguas en un determinado nivel del dominio.»

No cabe duda de que el MCER presenta una gran aportación e impulso en el desarrollo de la enseñanza de lenguas extranjeras, pero de acuerdo con Weir (2005) tenemos que admitir que no es lo suficientemente comprensivo, coherente y transparente para las pruebas de lenguas extranjeras, y, como dice Alderson (2010), tampoco es adecuado por sí solo para el desarrollo de las pruebas, o incluso de los manuales o currícula. A todos estos campos por desarrollar e investigar se le debería añadir también el proceso de estandarización y los problemas que pueden tener los jueces a la hora de tomar la decisión sobre los puntos de corte, tema que ya ha empezado a investigar Papageorgiou (2010), y en Eslovenia Ilc y Stopar (2014), y que hemos experimentado en nuestro proceso de calibración.

## Fuentes

Spomladanski izpitni rok 2013. Španščina. Izpitna pola 1 OR. Izpitna pola 1 VR. 12 de junio del 2013. [http://www.ric.si/splosna\\_matura/predmeti/spanscina/](http://www.ric.si/splosna_matura/predmeti/spanscina/).

## Bibliografía general

- ACQUARONI MUÑOZ, Rosana (2003) «La comprensión lectora.» In: J. Sánchez Lobato/I. Santos Gargallo (eds), *Vademécum para la formación de profesores*. Madrid: SGEL, 943–964.
- ALDERSON, Charles James/Caroline CLAPHAM/Diane WALL (1995) *Language Test Construction and Evaluation*. Cambridge: CUP.
- ALDERSON, Charles James/Caroline CLAPHAM/Diane WALL (1998) *Exámenes de idiomas. Elaboración y evaluación*. Cambridge: CUP.
- ALDERSON, Charles James (2000) *Assesing Reading*. Cambridge: CUP.
- ALDERSON, Charles James (2010) «Language testing-informed SLA? SLA-informed language testing?» In: I. Bartning/M. Martin/I. Vedder (eds), *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*. Roma: Eurosla Monographs Series, 239–248.
- BACHMANN, Lyle (1990) *Fundamental considerations in Language Testing*. Oxford: OUP.
- BIERE, Bernd Ulrich (1991) *Textverstehen und Textverständlichkeit*. Heidelberg: Groos.
- BITENC PEHARC, Suzana/Alenka TRATNIK (2014) *Umetstitev nacionalnih izpitov iz angleščine v skupni jezikovni evropski okvir. Zaključno poročilo o izvedbi projekta*. Ljubljana: Državni izpitni center.
- CIZEK, Gregory J./Michael B. BUNCH (2007) *Standard Setting, A Guide to Establishing and Evaluating Performance Standards on Tests*. Thous and Oaks/London/New Delhi: SAGE Publications.
- CONSEJO DE EUROPA (2009) *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): a Manual*. Strasbourg: Consejo de Europa [en línea]. 4 de julio de 2014. [http://www.coe.int/t/dg4/linguistic/manuel1\\_en.asp](http://www.coe.int/t/dg4/linguistic/manuel1_en.asp).
- CONSEJO DE EUROPA/CENTRO EUROPEO PARA LENGUAS MODERNAS (2011) *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): a Manual*. Highlights from the Manual. Strasbourg: Consejo de Europa/Graz: Centro europeo de lenguas modernas.

- FERBEŽAR, Ina/Marko STABEJ (2008) «Razumeti razumevanje.» *Jezik in slovstvo* 53/1, 15–31.
- FERBEŽAR, Ina (2012) *Razumevanje in razumljivost besedil*. Ljubljana: Znanstvena založba Filozofske fakultete.
- FIGUERAS CASANOVA, Neus (2012) «Hacia el mutuo reconocimiento de certificaciones: criterios y procedimientos de comparabilidad.» In: Ministerio de Educación, Cultura y Deporte (ed.), *Las Escuelas oficiales de idiomas en el desarrollo de las políticas lingüísticas del Consejo de Europa* [libro electrónico].
- GARCÍA, Pablo E./Francisco J. ABAD/Julio OLEA/David AGUADO (2012) *Un nuevo método de standard setting basado en la TRI: aplicación a eCat-Listening*. Madrid: Universidad Autónoma de Madrid. 4 de julio de 2014. [http://www.iic.uam.es/catedras/Publicaciones/Informe\\_Standard\\_Setting.pdf](http://www.iic.uam.es/catedras/Publicaciones/Informe_Standard_Setting.pdf).
- GOSPODARIČ, Katarina/Marjana ŠIFRAR KALAN (2008) «Matura: examen final de E/LE de bachillerato en Eslovenia.» In: S. Pastor Cesteros/S. Roca Marín (eds), *XVIII. Congreso ASELE. Actas*. Alicante: Universidad de Alicante, 323–335. 4 de julio de 2014. [http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/asele/pdf/18/18\\_0323.pdf](http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/18/18_0323.pdf).
- ILC Gašper/Andrej STOPAR (2014) «Validating the Slovenian national alignment to CEFR: The case of the B2 reading comprehension examination in English.» *Language Testing*, 1–20. 14 de enero de 2015. <http://ltj.sagepub.com/content/early/2014/12/17/0265532214562098.full.pdf>
- KINTSCH, Walter/Teun A. VAN DIJK (1983) *Strategies of discourse comprehension*. New York: Academic Press.
- LAH, Meta (2012) *Bralno razumevanje pri pouku tujega jezika: vloga tipa in avtentičnosti besedila*. Ljubljana: Znanstvena založba FF.
- LAH, Meta (2013) «La lecture à haute voix en classe de langue étrangère - une activité à réhabiliter?» *Linguistica* 53/1, 199–207.
- Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (2002). Madrid: Ministerio de Educación y Ciencia/Instituto Cervantes.
- MIKELN, Petra (2008) «Pomen besedišča pri branju v tujem jeziku.» In: J. Skela (ed.) *Učenje in poučevanje tujih jezikov na Slovenskem*. Ljubljana: Tangram, 392–402.
- PAPAGEORGIOU, Spiros (2010) «Investigating the decision-making process of standard setting participants.» *Language Testing* 27/2, 261–282.
- PERNAS IZQUIERDO, Almudena (2012) «Diversificación curricular y plurilingüismo: lenguas, ámbitos, actividades, competencias.» In: Ministerio de Educación, Cultura y Deporte (ed.), *Las Escuelas oficiales de idiomas en el desarrollo de las políticas lingüísticas del Consejo de Europa* [libro electrónico].
- SWAIN, Merrill (1985) «Large-scale Communicative Language Testing: A Case Study.» In: Y. P. Lee/A. C. Y. Y. Fok/R. Lord/G. Low (eds), *New Directions in Language Testing*. Oxford: Pergamon Press, 35–46.
- WEIR, Cyril J. (1990) *Communicative Language Testing*. Nueva York: Prentice Hall.
- WEIR, Cyril J. (2005) «Limitations of the Common European Framework for developing comparable examinations and test.» *Language Testing* 22/3, 281–300.

Resumen  
LA CALIBRACIÓN DE LA COMPRENSIÓN LECTORA DENTRO  
DEL EXAMEN NACIONAL DE ELE SEGÚN EL MCER:  
ALGUNAS CUESTIONES DE EVALUACIÓN

El objetivo del presente trabajo es presentar los resultados del proyecto de calibración de la prueba de la comprensión lectora dentro del examen nacional de matura en español como lengua extranjera (ELE) en relación con el *Marco Común Europeo de Referencia* (MCER). El documento del MCER y las directrices para la evaluación (Consejo de Europa 2009) se presentan como puntos de referencia y herramientas de estandarización para establecer correspondencias entre las pruebas de matura y los niveles del MCER. Asimismo se analiza el proceso comunicativo de lectura en la lengua extranjera en comparación con la lengua materna. Después de una breve introducción del proyecto de calibración de pruebas de lenguas extranjeras y las metodologías aplicadas se presenta un análisis contrastivo de los resultados de calibración obtenidos mediante dos metodologías: la estadística con los métodos de estandarización *Angoff* y *Basket*, y la descriptiva con las escalas descriptivas del MCER. Se presenta el punto de corte establecido por el panel de jueces para comprobar el nivel B1 en el examen básico y el nivel B2 en el examen avanzado de matura. Sin embargo, a partir del análisis descriptivo detallado de los ítems se ha estimado que los descriptores del MCER deberían aportar mayor transparencia, exhaustividad y precisión, ya que se concibe la comprensión lectora como una actividad tridimensional integradora de varias (sub)competencias generales y lingüísticas.

**Palabras clave:** los niveles de referencia del MCER, los descriptores del MCER, la evaluación de la comprensión lectora, la estandarización de exámenes, los exámenes de ELE.

Abstract  
RELATING READING COMPREHENSION IN THE SPANISH AS A FOREIGN  
LANGUAGE NATIONAL EXAM TO THE CEFR: SOME ASPECTS OF  
EVALUATION

The aim of the article is to present the results of a project which tried to relate the reading comprehension test in the Spanish as a foreign language national *matura* exam to the Common European Framework of Reference for Languages (CEFR). The CEFR and the evaluation guidelines (Council of Europe 2009) present reference points and standardization tools for establishing the correspondence between the knowledge required in the exam on the one hand and the levels of linguistic competence proposed by the CEFR on the other hand. Reading comprehension is analysed as a communicative process in foreign and native language. After introducing the project of calibrating the foreign language exam to the CEFR and presenting the methodology employed in the



study, a comparative analysis of the results is presented. The analysis combines two methods: *Angoff* and *Basket* standard setting methods and a descriptive method based on the CEFR descriptive scales. Cut scores for B1 level in the elementary examination and for B2 level in the advanced examination as established by the panel of judges are proposed. However, a detailed descriptive analysis of each exam item has shown that the CEFR descriptors should be more transparent, comprehensive and precise, given that reading comprehension is a threedimensional skill consisting of various general and linguistic (sub)competences.

**Keywords:** CEFR reference levels, CEFR descriptors, reading comprehension evaluation, exam standard setting, exams in Spanish as a foreign language.

#### Povzetek

### UMESTITEV BRALNEGA RAZUMEVANJA MATURITETNEGA IZPITA IZ ŠPANŠČINE V SKUPNI EVROPSKI JEZIKOVNI OKVIR: NEKATERI VIDIKI VREDNOTENJA

Namen pričujočega članka je predstaviti rezultate projekta umeščanja preverjanja bralnega razumevanja pri maturitetnem izpitu iz španščine kot tujega jezika v Skupni evropski jezikovni okvir (SEJO). Dokument SEJO in smernice za vrednotenje (Svet Evrope 2009) predstavljajo izhodišča ter standardizacijska sredstva za oceno ujemanja med zahtevnostjo maturitetnega izpita in referenčnimi ravni. Članek razčleni bralno razumevanje kot sporazumevalni proces v tujem in maternem jeziku. Kratki predstavitvi projekta umeščanja maturitetnih izpitov iz tujih jezikov in uporabljenih metod sledi primerjalna analiza rezultatov umeščanja, pridobljenih z dvema metodama: statistično – standardizacijski metodi *Angoff* in *Basket* – ter opisno z opisniki lestvic SEJO. Predstavljene so meje (minimalni prag točk) za ravni, ki so jih presojevalci prisodili za raven B1 na osnovni ter B2 na višji izpitni ravni. Na podlagi podrobnejše opisne analize izpitnih postavk se je izkazalo, da bi opisniki lestvic SEJO morali zagotavljati večjo transparentnost, izčrpnost in natančnost, saj se bralno razumevanje pojmuje kot tridimenzionalna dejavnost, v kateri se prepletajo raznolike splošne in jezikovne (pod)kompetence.

**Ključne besede:** referenčne ravni SEJO, opisniki SEJO, vrednotenje bralnega razumevanja, standardizacija izpitov, izpiti iz španščine kot tujega jezika.