

PROCESO DE COMPARABILIDAD DE TITULACIONES DE EUSKARA Y ADECUACIÓN AL MARCO COMÚN EUROPEO DE REFERENCIA (MCER)

Íñaki Pikabea Torrano, José Francisco Lukas Mugika, Neus Figueras Casanovas

RESUMEN

En el País Vasco existen más de 140.000 personas con títulos acreditativos de conocimientos de euskara, títulos expedidos por diversos organismos oficiales del Gobierno Vasco pero no convalidados entre ellos. Al objeto de ser convalidables y adecuados al Marco Común de Referencia para las Lenguas, el Gobierno Vasco encargó un estudio a un grupo de expertos para dicho fin, cuyo procedimiento y sugerencias se detallan en este artículo.

***Palabras clave:** Evaluación de conocimientos lingüísticos, Títulos, Euskara, Marco europeo de referencia, Convalidaciones.*

PROCESS OF BASQUE LANGUAGE CERTIFICATE VALIDATION AND ADAPTATION TO THE COMMON EUROPEAN FRAMEWORK OF REFERENCE (CEFR)

ABSTRACT

There are over 140,000 individuals in the Basque Country with certificates of Basque language proficiency. These certificates are issued by different official bodies of the Basque Government but lack mutual recognition between them. In order for the certificates to be recognised and adapted to the Common European Framework of Reference for Languages, the Basque Government commissioned a group of experts to conduct a study for this purpose. The procedure for the study and suggestions are described in this article.

Correspondencia:

Íñaki Pikabea Torrano. Facultad de Filosofía y CCEE. Universidad del País Vasco-Euskal Herriko Unibertsitatea. Avda. Tolosa, 20. San Sebastián 20018. E-mail: i.pikabea@ehu.es
Jose Francisco Lukas Mugika. Facultad de Filosofía y Ciencias de la Educación. UPV-EHU.
Neus Figueras Casanovas. Departament d'Educació. Generalitat de Catalunya.

Key words: Assessment of language proficiency, Certificates, Basque language, European Framework of Reference, Validations.

I. INTRODUCCIÓN

En la Comunidad Autónoma Vasca existen actualmente diversos títulos (más de 20) que acreditan el conocimiento de la lengua vasca o euskara en niveles diferentes. Desde la instauración de dichos títulos, la totalidad de certificaciones expedidas hasta abril de 2008 es de 140.448. Dichos títulos son requeridos para desempeñar algunas de las funciones del personal de la administración (profesorado, profesionales de la salud, fuerzas de seguridad, etc.) así como de entes privados. Los títulos son expedidos a través de diversos departamentos o instituciones del Gobierno Vasco (Dpto. de Educación, Universidades e Investigación; Instituto Vasco de Administración Pública; Dpto. de Cultura, etc.).

Hasta la fecha, la persona que deseaba presentarse a un puesto de trabajo a alguna de las administraciones pública con requisito de un determinado nivel de conocimiento de euskara, debía superar la correspondiente prueba que, habitualmente, únicamente tenía validez para el departamento al que se presentaba. Si optaba por un puesto de otro departamento, debía superar la prueba de euskara exigida por ese otro ámbito (salvo equivalencias muy limitadas).

Sobre el estado de la investigación en este tema, hay que señalar que en algunas comunidades autónomas del estado (por ejemplo Cataluña, Comunidad Valenciana o Islas Baleares) existen marcos de convalidaciones de los distintos títulos y certificaciones acreditativos de conocimientos de la lengua propia, tanto los expedidos por órganos de sus respectivas administraciones como por otras instancias, donde se atribuye valor acreditativo de conocimientos lingüísticos a las enseñanzas primaria y secundaria.

En este marco, el 14 de julio de 2006, el Gobierno Vasco, a propuesta de las Consejeras de Hacienda y Administración Pública y de Cultura, adoptó el acuerdo de “elaborar un proyecto, soportado normativamente que tenga por objeto el establecimiento de un Sistema Marco de Acreditación de Conocimientos de Euskara que acoja distintas rutas para la certificación de conocimiento de la lengua desde la fijación de elementos comunes a todas ellas. El referido sistema deberá establecer las especificaciones comunes de capacidad lingüística para cada uno de los niveles previamente definidos, que posibilitarán, a su vez, la descripción de exámenes, así como el diseño de un cuadro de equivalencias de las distintas certificaciones de conocimientos de euskara”. A tal fin, se constituyó un grupo de trabajo bajo la dirección y coordinación de la Viceconsejería de Política Lingüística del Dpto. de Cultura, y con la participación de los organismos autónomos HABE e IVAP. Asimismo, contó con una representación del Dpto. de Educación, Universidades e Investigación en la fase de análisis del valor acreditativo de conocimientos lingüísticos que otorgase en la enseñanza primaria, secundaria y superior.

Para el establecimiento de un sistema de acreditación de conocimientos de euskara, la comisión delegada entendía que el Marco Común Europeo de Referencia para las lenguas del Consejo de Europa (en lo sucesivo MCER) podría resultar de gran utilidad, ya que ofrece una escala global de niveles comunes de referencia, de amplio espectro, a la que podrían reconducirse los distintos sistemas de acreditación de los conocimientos lingüísticos de euskara.

El Gobierno Vasco también entendía que debía incorporarse en el estudio el trabajo realizado desde instituciones europeas como ALTE (Asociación Europea de Evaluadores de Idiomas) o EALTA (Asociación Europea para la Evaluación de Lenguas). Estas asociaciones se centran en la formación de profesorado en evaluación y pruebas de idiomas, tanto para el ámbito del aula, como para el ámbito de acreditaciones de instituciones y centros públicos de diversos países europeos.

En total, las instituciones implicadas en cuanto a las titulaciones expedidas en euskara en el País Vasco han sido las siguientes:

- HABE (Instituto para la Euskaldunización y Alfabetización de Adultos. Consejería de Cultura)
- IVAP (Instituto Vasco de Administración Pública)
- Dpto. de EDUCACIÓN, UNIVERSIDADES E INVESTIGACIÓN
- ERTZAINZA (Dpto. de Policía. Consejería de Interior)
- OSAKIDETZA (Consejería de Salud)

Es en este contexto donde la Viceconsejería de Política Lingüística del Gobierno Vasco contactó en diciembre de 2006 con los autores del artículo para llevar a cabo un estudio para, posteriormente, proponer al Gobierno un cuadro de convalidaciones, así como sugerencias para su adecuación al MCER.

En este artículo se expone brevemente el procedimiento seguido en este estudio, así como un conjunto de sugerencias sobre la base de observaciones realizadas. No se pretende mostrar los resultados detallados con relación a cada título puesto que dichos datos requieren de la confidencialidad correspondiente.

2. OBJETIVOS Y LIMITACIONES

Los *objetivos* propuestos fueron los siguientes.

1. Establecimiento de un conjunto común de especificaciones de dominio de la lengua, que permita comparar de modo objetivo niveles, pruebas y grados de aprovechamiento en los distintos sistemas de certificación en adultos.
2. Recogida y análisis de los actuales sistemas de acreditación de conocimientos del euskara en adultos.
3. Propuesta de adaptaciones que se consideren necesarias teniendo como base los criterios y los niveles del MCER.
4. Diseño de un cuadro de equivalencias de las distintas certificaciones de conocimientos de euskara en adultos, en el caso de que ello sea posible.

Entre las dificultades para llevar a cabo el estudio merecen especial atención el poco tiempo disponible para el proyecto y la constatación de que no existían datos empíricos sobre las pruebas. Vincular pruebas sobre las que no se dispone de datos empíricos con los niveles del Marco es necesariamente cuestionable, y cualquier resultado debe considerarse orientativo.

Tanto los objetivos como el contenido de este informe se basan en un estudio pormenorizado y fiabilizado a partir de la valoración de jueces expertos. Por otro lado, se quiere resaltar que este estudio tiene un carácter preliminar, puesto que deberá repetirse en cuanto se hayan llevado a cabo las recomendaciones que se listan en el último apartado del artículo.

3. TÍTULOS EVALUADOS

Los títulos evaluados han sido en total 19, procedentes de las siguientes instituciones:

- **HABE (Instituto de Euskaldunización y Alfabetización de Adultos).** HABE evalúa y otorga cuatro títulos: el 1º, el 2º, el 3º o EGA¹, y el 4º.
- **IVAP (Instituto Vasco de Administración Pública):** El IVAP evalúa a empleados o candidatos adultos al servicio de las administraciones públicas vascas. Otorga cuatro títulos: el 1º, el 2º, el 3º y el 4º.
- **Dpto. de EDUCACIÓN-Perfiles lingüísticos:** El Dpto. de Educación evalúa y otorga cuatro títulos destinados al personal docente y laboral de Educación: el *Perfil Lingüístico 1 (PL1)*, el *Perfil Lingüístico 2 (PL2)*, el *perfil Certificado de Euskara para los Profesionales no docentes de Educación Especial (HLEA)* y el *Certificado de Aptitud de Euskara (EGA)*.
- **Dpto. de EDUCACIÓN: Escuelas Oficiales de Idiomas.** Actualmente, el Dpto. de Educación, a través de las Escuelas Oficiales de Idiomas evalúa y otorga dos títulos del primer nivel: el *Ciclo Elemental (1º, 2º y 3º)* y el *Ciclo Superior (4º y 5º)*. En este estudio se analizaron los exámenes de 3º, 4º y 5º.
- **ERTZAINZA.** La Ertzaintza o Policía Autónoma tiene en vigor un único examen el PL1, aunque está previsto un PL2.
- **OSAKIDETZA.** El Servicio Vasco de Salud u Osakidetza evalúa y otorga tres perfiles (1º, 2º y 3º).

4. METODOLOGÍA

El grupo de trabajo acordó desde seguir las directrices del Consejo de Europa para vincular los exámenes con los niveles del MCER. En el *Manual "Preliminary Pilot version. Manual for relating examinations to the CEFR"* del Council of Europe (2003), para relacionar exámenes con los niveles del MCER se proponen una serie de procedimientos que definen el proceso a seguir para la vinculación de pruebas con los niveles del Marco en tres fases y con cuatro actividades diferenciadas: *familiarización, especificación, estandarización de juicios y validación empírica*.

Se concretaron los pasos a seguir y el calendario, y se elaboró un protocolo de actuaciones que se hizo llegar a todas las instituciones participantes en el proyecto, solicitando:

- el mismo número de pruebas (correspondientes a los mismos años)
- completar las parrillas que se diseñaron a este propósito
- facilitar un número de actuaciones escritas y orales de alumnos, con unas características fijadas

1 Este título procede del Dpto. de Educación.

Para llevar adelante los objetivos propuestos se llevaron a cabo dos procedimientos. Por un lado, se prepararon unos protocolos o parrillas para el análisis de los exámenes y, por otro, mediante un sistema de jueces, se intentaron buscar las equivalencias entre los distintos niveles. Pasamos a detallar cada uno de los procedimientos.

4.1. Parrillas de análisis: especificación

El *Manual* para relacionar exámenes con los niveles del Marco (2003) publicado por el Consejo de Europa propone un protocolo de análisis de prueba muy sistemático, para el que se utilizan formularios o parrillas que ayudan tanto a la descripción general como a la descripción pormenorizada de los exámenes o pruebas que se pretende vincular a los niveles del MCER. El *Manual* en su Capítulo 4 presenta 23 parrillas que pueden agruparse de la forma siguiente:

- Parrillas para la descripción general del examen
- Parrillas para la descripción del proceso de elaboración, corrección, puntuación y análisis del examen
- Parrillas para la descripción de cada una de las partes del examen

En estas parrillas se requiere a la institución que se ocupa que describa su examen, primero con una descripción general, y a continuación con una descripción en relación con los contenidos y niveles del MCER.

Para facilitar la tarea de las instituciones que participaban en el estudio, el equipo de trabajo —a partir del estudio de la documentación de que se disponía al inicio de los trabajos— realizó una síntesis de las parrillas del *Manual*. Esta síntesis fue la que permitió elaborar una parrilla única con cinco grandes apartados (los que el equipo consideró más relevantes para poder llevar a cabo el trabajo encomendado) que se envió a las instituciones como parte del protocolo de trabajo.

4.2. Sistema de jueces: estandarización

Se optó por fiabilizar la información facilitada por cada Institución en la parrilla descrita en el apartado anterior mediante un análisis sistemático de las tareas de examen propiamente dichas, por parte de un panel de jueces expertos nombrado para la tarea.

Este análisis se basó en el estudio de diversas tareas/actividades utilizadas para evaluar las cuatro destrezas lingüísticas: la comprensión escrita y la oral, la expresión escrita y la oral.

El proceso estribaba en utilizar jueces expertos en evaluación y docencia del euskara para evaluar las distintas tareas en relación con los niveles del MCER. La selección de jueces expertos se realizó desde las distintas instituciones a partir de las directrices de los expertos. Los jueces expertos (un total de 30), de procedencia institucional diversa, tenían la particularidad de que eran experimentados en actividades docentes y de evaluación con respecto al euskara. Previamente se les envió y requirió la lectura de algunos textos del MCER.

Todas las tareas se formatearon de forma que resultara difícil saber a qué institución pertenecían y se recogieron en cuatro cuadernos, uno para cada destreza. Una vez compilados los cuadernos se seleccionaron dos tareas para cada destreza que pudieran considerarse “modelos” de nivel. Estas tareas funcionaron como ejercicio de entrenamiento y como ejemplo ilustrativo de la tarea que debían llevar a cabo los jueces. Fueron necesarias dos sesiones de trabajo de día completo con los jueces. El desarrollo de las sesiones de la primera fase se presenta a continuación:

SESIÓN I

Familiarización

En la primera parte de la Sesión se llevaron a cabo las actividades de familiarización recomendadas en el *Manual* (Capítulo 3), actividades que se retomaron en la segunda parte de la sesión, antes de proceder a la evaluación de las tareas de comprensión oral.

1. Presentación del MCER.
2. Explicación de los 6 niveles del Marco.
3. Formación y adiestramiento en los descriptores del Marco en lo referente a Comprensión Escrita. Para ello, se realizó una actividad en la que los jueces tuvieron que ordenar y clasificar una serie de descriptores en función de los niveles del Marco. Se llevó a cabo una tarea de puesta en común para discutir las características diferenciales de los distintos descriptores para cada nivel, y se recogieron las clasificaciones de los jueces para su posterior análisis obteniéndose un coeficiente de fiabilidad aceptable.
4. A continuación se procedió a introducir la tarea de análisis y valoración de las tareas e ítems de examen que debía llevarse a cabo. Se pidió a los jueces estimar la dificultad de ítems de comprensión escrita. A tal efecto se informó a los jueces sobre el trabajo realizado por el proyecto *CEFR Dutch Construct Project* (Alderson, 2006) en el que se incide especialmente en la importancia de considerar la dificultad de una tarea de comprensión como el fruto de la interacción entre la dificultad del texto sobre el que se basa la pregunta o preguntas y la dificultad de la pregunta propiamente dicha. Se presentó al grupo una tarea-ejemplo elegida por el equipo de trabajo como representativa de un nivel concreto (de entre las tareas recibidas desde las distintas instituciones) y se discutió su dificultad y nivel en relación con los niveles del Marco. La corrección entre todo el grupo de jueces y los instructores mostró que la unificación de criterios era más que evidente. A partir de este momento se procedió al análisis de tareas correspondientes a pruebas reales.

Análisis y valoración

5. Se presentaron a los jueces todas las pruebas que utilizaron las entidades que se estaban analizando durante el año 2006 para medir la *Comprensión Escrita* de los aspirantes. Todas las pruebas fueron presentadas sin ningún tipo de ordenación y sin ningún tipo de identificación. Es decir, los jueces desconocían a qué entidad en concreto correspondía cada una de las pruebas.

6. Junto con el cuadernillo de las pruebas se les entregó una hoja de respuestas en la que debían señalar el nivel que, según ellos, correspondía a la tarea en sí y explicitar aquellas sugerencias que consideraran oportunas. En concreto la pregunta que se les planteaba era la siguiente: *“Si una persona debe enfrentarse a la siguiente tarea, ¿qué nivel del MCER debe tener para acertar, al menos, el 50 o 60 % de las cuestiones planteadas?”*. Aunque habitualmente el trabajo de valoración de pruebas para fijar estándares se lleva a cabo para cada ítem o pregunta, en este caso se procedió a la valoración de las tareas como tal con todos los ítems o preguntas vinculadas a las mismas.
7. En la segunda parte de la reunión se continuó con el trabajo de la primera sesión. Primeramente se les formó y adiestró en los descriptores del Marco en lo referente a Comprensión Oral. Para ello, se realizó una actividad en la que tuvieron que ordenar y clasificar una serie de descriptores en función de los niveles del Marco, realizándose el mismo proceso que para la *Comprensión Escrita*, ya descrito en los apartados 3 y 4.
8. Se presentaron a los jueces todas las pruebas que utilizaron las instituciones participantes en el estudio durante el año 2006 para medir la *comprensión oral* de los aspirantes. Todas las pruebas fueron presentadas sin ningún tipo de ordenación y sin ningún tipo de identificación. Es decir, los jueces desconocían a qué entidad en concreto correspondía cada una de las pruebas.
9. Junto con el cuadernillo de las pruebas se les entregó una hoja de respuestas en la que debían señalar el nivel que, según ellos, correspondía a la tarea en sí y explicitar aquellas sugerencias que consideraran oportunas. En concreto la pregunta que se les planteaba era la siguiente: *“Si una persona debe enfrentarse a la siguiente tarea, ¿qué nivel del MCER debe tener para contestar, al menos, el 50 o 60 % de las cuestiones planteadas?”*

SESIÓN 2

En esta sesión se siguió la misma estructura seguida en la Sesión 1, aunque se añadieron dos actividades previas:

- Se hizo un recordatorio de las características principales del MCER.
- A continuación, y a modo de “repaso” se pidió a los jueces que volvieran a resolver dos de los ejercicios resueltos en la primera sesión (uno de comprensión escrita y otro de comprensión oral). En realidad esto se utilizó para comprobar la *fiabilidad* de los jueces, es decir, si la clasificación realizada en la sesión primera coincidía con la realizada este segundo día. Hay que señalar que entre las dos sesiones mediaron más de 15 días por lo que con toda probabilidad los jueces no recordaban qué es lo que habían contestado en esos ejercicios concretamente.

Familiarización

1. Formación y adiestramiento en los descriptores del Marco en lo referente a la *Expresión Oral*. Para ello, se realizó una actividad en la que los jueces tuvieron que ordenar y clasificar una serie de descriptores en función de los niveles del MCER.

Además de trabajar con los descriptores propios de la destreza de la expresión oral, se añadieron descriptores relativos a los aspectos lingüísticos en la expresión oral, puesto que se consideró que los elementos lingüísticos en la actuación de los alumnos/examinandos son muy informativos para los correctores a la hora de adjudicar un nivel. Se utilizó para esta actividad el cuadro 3 del MCER: *Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada*. Se facilitó además a los asistentes una lista completa de las escalas de descriptores que pueden encontrarse en el MCER para la expresión, con la intención de que los jueces tuvieran el máximo de información sobre la naturaleza de la expresión. El equipo recogió el trabajo de los jueces con los descriptores para llevar a cabo análisis con resultados que se describen más adelante.

2. A continuación, y siguiendo el ejemplo de la Sesión I, se evaluó entre todo el grupo de jueces y los instructores un primer ejemplo de actuación oral de un alumno, comprobándose que la unificación de criterios era más que evidente. A partir de aquí se inició el trabajo con todas las actuaciones reales que se pretendían analizar.

Análisis y valoración

3. Se presentaron a los jueces todas las actuaciones seleccionadas de entre las recogidas por las entidades en situación de examen real. Todas las pruebas fueron presentadas sin ningún tipo de ordenación y sin ningún tipo de identificación. Es decir, los jueces desconocían a qué entidad en concreto correspondía cada una de las pruebas.
4. Para poder llevar a cabo y recoger la valoración de las actuaciones, junto con el cuadernillo de las pruebas, se entregó a los jueces una hoja de respuestas en la que debían señalar el nivel al que, según ellos, correspondía la actuación en sí y explicitar aquellas sugerencias que consideraran oportunas. En concreto la pregunta que se les planteaba era la siguiente: “¿En qué nivel del Marco encuadraría a la persona cuya producción escucha? Es decir ¿cuál es el nivel que superaría no holgadamente?”
5. Tras el almuerzo se continuó con el trabajo de la primera sesión. Primeramente se les formó y adiestró en los descriptores del MCER en lo referente a *Expresión Escrita*. Para ello, se realizó una actividad en la que tuvieron que ordenar y clasificar una serie de descriptores en función de los niveles del Marco. Al igual que con el trabajo para la *Comprensión* y para la *Expresión Oral*, se recogieron las respuestas que dieron los jueces y después fueron analizadas obteniéndose un coeficiente de fiabilidad aceptable. Asimismo se procedió también a un primer análisis del nivel de un escrito como entrenamiento para el trabajo que se iba a realizar.
6. Se presentó a los jueces un cuaderno con todas las actuaciones escritas que debían analizar (36). Todas las actuaciones fueron presentadas sin ningún tipo de ordenación y sin ningún tipo de identificación. Es decir, los jueces desconocían a qué entidad en concreto correspondía cada una de las pruebas.

7. Junto con el cuadernillo de las actuaciones escritas, se entregó a los jueces una hoja de respuestas en la que debían señalar el nivel que, según ellos, correspondía la tarea en sí y explicitar aquellas sugerencias que consideraran oportunas. En concreto la pregunta que se les planteaba era la siguiente: “¿En qué nivel del Marco encuadraría a la persona cuya producción escrita acaba de leer? Es decir ¿cuál es el nivel que superaría no holgadamente?”

5. RESULTADOS

En el apartado de los resultados vamos a exponer tres tipos de análisis. Por un lado los recogidos mediante la parrilla que completaron los responsables de las instituciones implicadas en la evaluación. En segundo lugar, aparecen los análisis técnicos-cualitativos realizados sobre las propias pruebas. Por último, en el tercero de los apartados, además de mostrar los datos relativos de la fiabilidad de los jueces se muestran las equivalencias de las distintas pruebas analizadas con respecto al MCER.

5.1. Resultados del análisis de las Parrillas

En este análisis se han tenido en cuenta los aspectos relativos a la elaboración y administración de las pruebas, la corrección y puntuación y el análisis de datos y revisión que hacen a posteriori de las pruebas.

5.1.1. Elaboración y administración de pruebas

Los aspectos que se van a señalar a continuación hacen referencia a las situaciones más comunes en los exámenes de las instituciones evaluadas, aunque no quiere decir que en todas las situaciones se dé. Se observó que existe una amplia *variedad de tareas* en los exámenes ya que se utilizan las siguientes: ítems de elección múltiple, verdadero/falso, emparejamiento, respuestas extensas, interacción con el examinador y en algún caso completar frases. En general, *la información que se publica* y que está a disposición de los candidatos y los profesores es adecuada dado que se publican el objetivo general, los ámbitos principales, las subpruebas, las tareas más frecuentes con las que se van a tener que enfrentar y, en algunos casos, muestras de exámenes utilizados en convocatorias anteriores. Después del examen, sólo se publica la nota global del alumno, obviando las notas parciales por subtests. Los exámenes son elaborados por *personal cualificado* de la entidad perteneciente en la mayoría de los casos al departamento de evaluación de la misma. Para garantizar la *calidad de las pruebas*, los redactores de las mismas han recibido algún tipo de formación, disponen de ciertas directrices y ejemplos de tareas y además, tienen en cuenta las descripciones calibradas de los niveles del MCER, aunque esto no es generalizable a todas las instituciones analizadas. De la misma manera, antes de utilizar las pruebas, son discutidas y analizadas con un comité de expertos. También señalan que realizan pruebas piloto de los exámenes, aunque la forma concreta de hacerlo no sea quizás la más adecuada puesto que se hacen de una manera informal y no sistemática. Además, da la impresión de que la prueba piloto se centra en los aspectos cualitativos de los exámenes y que no se realiza ningún tipo de análisis estadístico para comprobar

la eficacia de los ítems salvo en algún caso aislado. En cuanto al análisis de la *fiabilidad y la validez* de los exámenes, prácticamente en todos los casos señalan que son calculados, pero no se han facilitado.

5.1.2. Corrección y puntuación

Las preguntas objetivas son corregidas de manera automática en la mayoría de los casos. En cuanto a la *corrección* de las preguntas abiertas los examinadores son personas cualificadas y adiestradas. Dan mucha importancia al hecho de que aquellos exámenes que recogen la producción oral o escrita del alumnado, dado que su corrección es más subjetiva, deben ser analizados por dos correctores y comprobar después su concordancia. La misma puede conseguirse por consenso entre los dos correctores, calculando la media de ambos, o bien enviando el examen a un tercer corrector. Los *controles sistemáticos* por parte del coordinador de los exámenes para comprobar la adecuación de las correcciones y las sesiones de unificación para estandarizar juicios son acciones que se llevan a cabo con periodicidad. Además disponen de una serie de documentos que se utilizan en esta tarea. Para garantizar la estabilidad y consistencia de las *notas de corte* en las distintas pruebas y convocatorias se utilizan muestras de convocatorias anteriores en las mencionadas reuniones de identificación y además se observan los resultados estadísticos de las distintas convocatorias. En todos los casos los examinados tienen derecho a una *segunda corrección* del examen y en ese caso tienen también la posibilidad de revisar el mismo.

5.1.3. Análisis de datos y revisión a posteriori

Cuando se realiza un examen en la mayoría de las instituciones se recoge de manera sistemática información acerca de los resultados, incidencia, opiniones del alumnado y profesorado, etc. Esta información es incorporada en futuros exámenes. En todos los casos se intenta proteger la *confidencialidad* de los resultados y datos recogidos.

5.2. Análisis técnico cualitativo de las pruebas

Para el análisis técnico de las pruebas se han tomado en consideración, además de las directrices de EALTA, las recomendaciones más usuales en la redacción de ítems. El resumen de las observaciones realizadas se expondrá por destrezas.

5.2.1. Comprensión escrita

En la práctica totalidad de los exámenes analizado se utilizan artículos o textos extraídos de una revista o libro con el riesgo de que haya candidatos que lo conozcan (bien porque lo hayan leído a título individual o bien porque lo hayan trabajado en clase). En algunas de las instituciones tienden a utilizar preguntas de verdadero/falso con el problema que se genera con el del acierto al azar. Aunque para niveles iniciales, sería más conveniente pensar en ítems de elección múltiple. En ítems de elección múltiple, para minimizar los efectos del azar convendría usar cuatro opciones en vez de tres. Es

común que para medir la comprensión escrita del texto propuesto el número de preguntas sea muy reducido. En ocasiones menos de 5 ítems. Obviamente tan pocos ítems son insuficientes para medir la comprensión de un texto. El formato donde los ítems se presentan como enunciados con una frase sin completar, no es muy adecuado cuando se realiza en idiomas como el euskara donde se tiende a poner el verbo al final de la frase. En las instrucciones, debería informarse del número de cuestiones a contestar correctamente y la posible corrección del azar para que el sujeto sepa si debe contestar a todas las cuestiones o sólo aquellas cuya respuesta conoce.

5.2.2. *Comprensión oral*

Al igual que en el caso de la comprensión escrita, el enunciado como frase sin completar no es muy adecuado cuando se realiza en idiomas como el euskara donde se tiende a poner el verbo al final de la frase. Convendría utilizar ítems en donde se haga la pregunta correspondiente. En un ítem de elección múltiple, el enunciado debe dejar claro y sin ambigüedades, concretando qué es lo que se le pregunta al sujeto, aspecto que en muchas pruebas no se recoge. Al igual que en la comprensión escrita hay ocasiones en las que tras la escucha de algún fragmento, el alumnado debe contestar a una serie de preguntas acerca de lo que ha comprendido. El problema reside en que en la mayoría de los exámenes el número de ítems es insuficiente para garantizar la fiabilidad y validez de las pruebas. Hay ocasiones en los que los ítems no son independientes entre sí, aspecto que habría que garantizar.

El uso de preguntas de respuesta de verdadero/falso conlleva el problema del acierto al azar. Aunque para niveles iniciales podría utilizarse, sería más conveniente pensar en ítems de elección múltiple de tres o cuatro opciones. En caso contrario podría pensarse en realizar la corrección del azar. En las instrucciones, debería informarse del número de cuestiones a contestar correctamente y la posible corrección del azar para que el sujeto sepa si debe contestar a todas las cuestiones o sólo aquellas cuyas respuestas conoce. Para minimizar los efectos del azar convendría añadir un distractor más en los ítems de tal manera que quedaran con cuatro opciones en vez de con tres.

5.2.3. *Expresión escrita*

En general se presentan las instrucciones para el desarrollo de la prueba de manera adecuada. Las cuestiones a contestar por el examinado, así como el tema a tratar están expuestos claramente. No obstante, convendría especificar al examinado en todos los casos qué aspectos de la redacción serán los que se tendrán en cuenta.

5.2.4. *Expresión oral*

En general, se presentan las instrucciones para el desarrollo de la prueba de manera adecuada. Las preguntas a contestar por el examinado, así como el tema a tratar están expuestos claramente, al igual que la finalidad que se persigue. Aunque en la mayoría de los casos se señalan los criterios de evaluación, el nivel de especificación quizás no es suficiente para que distintos evaluadores lleguen a la misma calificación.

5.3. Análisis de las valoraciones de los jueces sobre las pruebas

5.3.1. Descripción y fiabilidad de los jueces

Los 30 jueces que participaron como expertos en la clasificación de los exámenes analizados en cuestión fueron seleccionados atendiendo sobre todo a su experiencia docente con personas adultas en el ámbito del euskara y a su vinculación con temas relativos a la medición y evaluación de las competencias lingüísticas en euskara. Entre las características de los jueces seleccionados destacamos las siguientes: el 62% es hombre y el 38% mujer, la media aritmética de años dedicados a la docencia se sitúa en 20, mientras que la media en tareas evaluativas está en 13 años. La *fiabilidad* hace referencia a la precisión de los juicios aportados por los jueces, por un lado y, por otro, a la consistencia de esos mismos juicios. En cuanto a la precisión de los juicios, en la fase de adiestramiento se les pidió que, a modo de entrenamiento, ordenaran y clasificaran una serie de descriptores en base a los 6 niveles de referencia del MCER. Esto se realizó en las pruebas de *comprensión escrita*, *comprensión oral*, *expresión escrita* y *expresión oral*. En la siguiente Tabla 1 pueden apreciarse los resultados correspondientes a los jueces de la primera fase.

TABLA 1
RESULTADOS SOBRE JUECES

	Nº de ítem (descriptores)	Nº de jueces	Media	Des. Tip.	W de Kendall	Alfa de Cronbach
<i>Comprensión Escrita</i>	16	30	11,83	2,51	0,954	0,643
<i>Comprensión Oral</i>	19	30	16,17	2,33	0,946	0,613
<i>Expresión Escrita</i>	20	30	14,53	4,01	0,890	0,784
<i>Expresión Oral</i>	16	30	12,11	3,16	0,855	0,715

Como puede observarse en la tabla precedente, la media de acierto de los jueces con respecto al Marco en todos los casos se sitúa en puntuaciones altas superando en todos los casos el 75%. Además, la *desviación típica* resultante no es muy grande por lo que la homogeneidad de los jueces parece evidente, aunque en todo caso, en la fase de análisis de los exámenes propiamente dichos se tuvo que comprobar la adecuación de cada uno/a. Se calculó el *índice de concordancia* entre los jueces con respecto a cada uno de los descriptores analizados mediante la W de Kendall. La prueba W de Kendall se puede interpretar como el *coeficiente de concordancia*, que es una medida de acuerdo entre evaluadores. La W de Kendall varía de 0 (sin acuerdo) a 1 (acuerdo completo). Como puede observarse en la tabla precedente, los valores alcanzados son netamente más que aceptables.

La *fiabilidad* de las respuestas dadas por los jueces en cuanto a precisión de los juicios fue calculada mediante el Alfa de Cronbach. Los valores obtenidos se sitúan por encima del 0,60 que, aunque no son excesivamente altos, deben ser considerados como aceptables debido a dos razones principalmente: el bajo número de ítems de cada una de las pruebas y, sobre todo, a la muestra reducida de jueces, factores ambos que condicionan y limitan el coeficiente de fiabilidad obtenido (ver procedimiento sesión 2). Posteriormente se calculó el coeficiente de correlación como coeficiente de fiabilidad de los juicios de los expertos (test-retest). Los resultados obtenidos pueden comprobarse en la Tabla 2.

TABLA 2
COEFICIENTE DE FIABILIDAD DE LOS JUECES

	Coeficiente de Fiabilidad
Comprensión Escrita	0,70
Comprensión Oral	0,65

En ambos casos la correlación fue significativa.

5.3.2. Análisis de los niveles y exámenes

En total fueron 39 ítems de comprensión escrita, 16 ítems de comprensión oral, 36 producciones escritas de los alumnos y otros 34 producciones orales de los alumnos las lo que los jueces evaluaron. A continuación vamos a analizar qué es lo que sucede en cada una de las cuatro destrezas analizadas.

Primeramente, para analizar la competencia de los jueces se realizó un estudio para comprobar si alguno de ellos no resultaba apto. Para ello, se analizó en qué medida cada uno de los jueces discrepaba de la opción mayoritaria en cada examen. De tal forma que, si alguno de los jueces presentaba un *índice de discrepancia* elevado sería eliminado de la muestra. No obstante, el análisis realizado reveló que todos los jueces seleccionados podían ser considerados como aptos para la tarea en cuestión.

Los resultados de los jueces fueron recogidos por destrezas (*comprensión escrita, comprensión oral, expresión escrita, expresión oral*). A modo de ejemplo, se muestra la Tabla 3 donde se recogieron los datos de la *comprensión escrita* (cuando existía). En la 1ª columna se indica la institución o dpto. del cual procede el ejercicio²; en la 2ª los títulos³; en la 3ª los ejercicios analizados —codificados— según sus niveles, a continuación,

2 Por respetar la privacidad, se han sustituido los nombres de la institución por números.

3 Por respetar la privacidad, se han sustituido los nombres reales por letras.

TABLA 3
RESULTADOS DE JUECES SOBRE COMPRESIÓN ESCRITA

DPTO. ²	TÍTU. ³	EJERCICIO	NIVELES DEL MCER						RESUMEN DEL NIVEL DE CADA TÍTULO
			A1	A2	B1	B2	C1	C2	
I	A	1.		43 ⁴	48	9			B1
		2.	4	61	35				
		3.		9	73	18			
	B	4.		22	57	17	4		B2
		5.			22	78			
		6.			17	83			
		7.		4	13	66	17		
		8.			17	61	18	4	
	C	9.				30	48	22	C1
		10.				5	95		
		11.				10	90		
II	A	12.			44	52	4	B1 - B2	
	B	13.			4	73	23	B2	
	C	14.			9	26	65	C1	
	D	15.					22	78	C2
III	A	16.		4	13	70	13	B2	
	B	17.		9	61	30		B1	
IV	A	18.			48	52		B1- B2	
	B	19.			4	78	18	B2	
	C	20.			4	39	53	4	C1
V	A	21.			16	52	32		B2
		22.				17	79	4	C1
		23.				48	52		B2-C1
	D	24.			52	40	8		B1-B2-C1
		25.			4	32	60	4	
		26.		4	55	36	5		
		27.			28	56	16		
		28.		4	17	31	48		
	E	29.			28	56	16		B2
		30.			12	64	24		
		31.			36	52	12		
		32.			20	68	12		
		33.			24	40	36		
		34.			12	40	48		
		35.			32	48	20		
	F	36.				64	36		B2
		37.			8	60	32		
38.				16	40	44			
39.				4	4	92			

4 Los números reflejan el porcentaje de jueces que ha clasificado dicho examen en la categoría en cuestión. Obviamente, la suma de las filas debe ser 100.

los niveles del MCER donde los jueces han catalogado cada ejercicio y, finalmente, el resumen general del grupo de ejercicios según niveles del MCER (en caso de duda, se indicaron dos o más). Las celdas oscuras indican la o las puntuaciones más altas, siempre y cuando fueran similares.

El sistema para la recogida de opinión de los jueces sobre la *comprensión oral* fue similar. Para evaluar la *expresión escrita* y la *expresión oral*, se pidió a los jueces que clasificaran exámenes ya realizados sin que conocieran cuál era el nivel que había otorgado la institución en cuestión. Para ello se pidió a cada institución una muestra representativa (de producción escrita y oral) de 6 examinados que hubieran aprobado no holgadamente el nivel al que se presentaron. Los jueces debían tener en cuenta el nivel del MCER donde encuadrarían a la persona cuya producción escuchaban.

Los datos recopilados fueron analizados exhaustivamente y, a modo de conclusión se confeccionó la Tabla 4, donde se refleja la posible comparabilidad de títulos, así como adecuaciones a los niveles del MCER. Para ello se siguieron los siguientes criterios:

- Análisis de las pruebas según los constructores
- Análisis cualitativo-técnico de las pruebas
- Análisis cuantitativo de los jueces sobre las pruebas.

Al lado de cada título, si se precisa, se indica en tonos grises diferentes en qué ámbitos debería ajustarse dicho título para poder darse una equivalencia integral a los niveles del marco (este es un ejemplo orientativo, donde se han eliminado nombres de instituciones y títulos expedidos).

Abreviaturas:

ce: comprensión escrita

ee: expresión escrita

co: comprensión oral

eo: expresión oral

Interpretación de posiciones y colores:

Misma columna	Equivalencia, en términos generales
	Ausencia de prueba específica
	No llega al nivel exigido
	Sobrepasa el nivel exigido

La Tabla 4 resumiría simbólicamente las conclusiones aportadas de este estudio para la futura toma de decisiones por parte de la administración.

A modo de ejemplo, la interpretación relacionada con la **Institución nº 1** significaría lo siguiente:

- No dispone pruebas para acreditar el nivel A1 y A2 del *Marco*.

TABLA 4
PROPUESTAS DE EQUIVALENCIA DE TÍTULOS Y ADECUACIONES AL MCER

NIVELES MCER										
INSTITUCIÓN	A1	A2	B1		B2		C1		C2	
1	ce	ce	Título A	ce	Título B	ce	Título C	ce	Título D	ce
	ee	ee		ee		ee		ee		
	co	co		co		co		co		
	eo	eo		eo		eo		eo		
2	ce	ce	Título A	ce	Título B	ce	Título C	ce	Título D	ce
	ee	ee		ee		ee		ee		
	co	co		co		co		co		
	eo	eo		eo		eo		eo		
3	ce	ce	Título A	ce	Título B	ce	Título C	ce	Título D	ce
	ee	ee		ee		ee		ee		
	co	co		co		co		co		
	eo	eo		eo		eo		eo		
4	ce	ce	Título A	ce	Título B	ce	Título C	ce	Título D	ce
	ee	ee		ee		ee		ee		
	co	co		co		co		co		
	eo	eo		eo		eo		eo		
5	ce	ce	Título A	ce	Título B	ce	Título C	ce	Título D	ce
	ee	ee		ee		ee		ee		
	co	co		co		co		co		
	eo	eo		eo		eo		eo		

- La prueba para acreditar el nivel B1 (título A) no requiere de ajustes de nivel. El diseño de pruebas y nivel exigido de *comprensión escrita* (ce), *expresión escrita* (ee), *comprensión oral* (co) y *expresión oral* (eo) está bien ajustado al Marco.
- La prueba para acreditar el nivel B2 (título B) está bien ajustada al Marco salvo en la *expresión oral* (eo), ya que no acredita al nivel requerido.
- La prueba para acreditar el nivel C1 (título C) está bien ajustada al Marco salvo en la *comprensión oral* (co), ya que no requiere al nivel exigido.
- La prueba para acreditar el nivel C2 (título D) está bien ajustada al Marco en la exigencia de la *expresión oral* (eo). Pero en la *comprensión escrita* (ce) y *oral* (co) no existe prueba específica —y habría que incorporarla—, y en la *expresión oral* (eo) la prueba no requiere el nivel deseado.

6. CONCLUSIONES, SUGERENCIAS Y RECOMENDACIONES

Una vez realizada una propuesta de convalidaciones de titulaciones de euskara y propuestas para la adecuación al MCER, en este estudio también se recogieron algunas sugerencias para un mejor uso y diseño de las pruebas de medición, entre las que destacaríamos las siguientes:

1. Formar en medición y evaluación educativa a los constructores de las pruebas.
2. Redactar un documento que describa con precisión la prueba y sus contenidos. Así mismo deberían mostrarse modelos de corrección y criterios de puntuación.
3. Las pruebas deberían expresar su nivel en referencia a las escalas del MCER.
4. Redactar un documento para la elaboración de pruebas y redacción de ítems.
5. Idear un sistema de pilotaje de las pruebas antes de su aplicación definitiva.
6. Realizar análisis estadísticos tras la aplicación de las pruebas para su validación (análisis de ítems, sesgos, fiabilidad, validez, etc.).
7. Establecer mecanismos para especificar puntos de corte o puntuación mínima necesaria en cada examen para poder ser considerado como *apto*.
8. Redactar especificaciones que marquen criterios claros para la evaluación de la comprensión, incluyendo longitud, tema, tipo y fuente de los textos, características de las tareas y de los ítems.
9. En cualquier caso, creemos que para evaluar de forma fiable el nivel de comprensión de un alumno, la prueba debe incluir —como mínimo— tres tareas con un mínimo total de 25 ítems. De esta forma puede garantizarse la representación del constructo de comprensión escrita u oral, tal y como se describe en las distintas escalas y descriptores que se incluyen en el MCER.
10. Comprobar la posibilidad de reducir las pruebas innecesarias.
11. Acordar si las pruebas deben evaluar, además del dominio de las destrezas comunicativas, el conocimiento gramatical de forma específica, ya que existe diversidad de criterios en los títulos.
12. En el caso de la expresión oral, por ejemplo, el/la candidato/a debe haber generado tanto un discurso monológico (expresión) como dialógico (interacción), y esto supone el que las tareas de evaluación permitan este tipo de actuación. En el caso de la expresión escrita, debe pensarse, a partir del nivel B1, en la inclusión de más de una tarea, de tema y tipología discursiva distinta.
13. Completar las pruebas que posee cada institución allá donde no se mida alguna destreza, y recoger todos los niveles y competencias señalados en el MCER.
14. Establecer un sistema de seguimiento de las adaptaciones propuestas, fijando un calendario de aplicación.
15. Diseñar un estudio que permita validar empíricamente la vinculación de las distintas titulaciones (una vez realizadas las modificaciones propuestas con los niveles del MCER).
16. Crear una comisión mixta que arbitre, coordine y supervise las actuaciones y propuestas relativas a las titulaciones de euskara, garantizando su validez, fiabilidad y equidad.

Por último, se desea destacar el alto nivel general observado en la elaboración de pruebas de idiomas de las diversas instituciones implicadas. Al margen de las sugerencias realizadas, la calidad manifiesta y los procesos seguidos son dignos de destacar en todas las pruebas analizadas.

7. CONVALIDACIONES ENTRE TÍTULOS DE EUSKARA CONFORME AL MCER

En base a las recomendaciones efectuadas por el grupo de expertos, el Gobierno Vasco estudió las diversas propuestas, adoptó las decisiones que estimó oportunas y publicó las convalidaciones entre títulos y certificados de euskara conforme al MCER, cuyos detalles se pueden leer en el Decreto 64/2008, de 8 de abril del 2008, y en el Decreto 48/2009 del 24 de febrero de 2009.

9. BIBLIOGRAFÍA

- Alderson, J.C. et al. (2006). Analysing tests of reading and listening in relation to the CEF: the experience of the Dutch CEF construct project. *Language Assessment Quarterly*, 3(1) 3-30.
- Council of Europe (2001): *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2003): *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): Manual Preliminary Pilot Version*. Strasbourg: Language Policy Division.
- Council of Europe (2004): *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Language.
- Downing, S.M; Haladyna, T.M. (eds.) (2006): *Handbook of test development*. London: Lawrence Erlbaum Associates.
- Europako Kontseilua (2005): *Ikaskuntza, irakaskuntza eta ebaluaziorako Europako Erreferentzia Markoa*. Donostia: Eusko Jaurlaritza.
- Figueras, N. (2005) "El Marco Europeo Común de Referencia, de la teoría a la práctica". *Revista Carabela* n. 57: 5-23. SGEL.
- Lizasoain, L; Joaristi, L; Santiago, C., J.F. Lukas; Moyano, N.; Sedano, M.; Munarriz, B. (2003) "El uso de las Técnicas de Segmentación en la evaluación del rendimiento en lenguas. Un estudio en la comunidad autónoma vasca", *Revista de Investigación Educativa*, Vol. 21, nº 1: 93-111.
- Pikabea, I; Joaristi, L; Lizasoain, L. (2004), "Uso social de un segundo idioma y estrategias de aprendizaje", *Revista de Investigación Educativa*, Vol. 22, nº 2: 459-472.
- Verdía, E.; Conde, G.; De Samblanc; Cassany, D. (2002): "Marco común europeo de referencia y Portfolio de las lenguas.", *Mosaico*, 9, monográfico. Revista para la Promoción y Apoyo a la Enseñanza del Español.

Recibido: 05 de mayo de 2009.

Aceptado: 14 de octubre de 2009.