

**Martínez Baztán, Alfonso**

Centro de Lenguas Modernas. Universidad de Granada

E-mail: [alfonsom@ugr.es](mailto:alfonsom@ugr.es)**Resumen**

La evaluación es una de las áreas que menos se ha desarrollado en la enseñanza de segundas lenguas a inmigrantes. Sin embargo, se necesita su participación para dar respuesta a las necesidades que se plantean en los programas con jóvenes y adultos inmigrantes. En el primer caso, para diseñar prueba de clasificación y evaluación final. En el segundo, para proponer modelos de certificación general que respondan a las necesidades de los adultos inmigrantes.

**Palabras clave**

L2, pruebas de nivel, certificación

**Abstract**

Evaluation is one of the fields less developed in the teaching of second languages to immigrants. Nevertheless, its presence is necessary to satisfy the needs of the programs integrated by young and adult immigrants. In the first case, we need to design level tests and final evaluations. In the second one, it is necessary to propose certifying models that respond to adult immigrant needs.

**Key words**

Evaluation L2, placement test, second language certification



## Introducción

En el campo de la evaluación para inmigrantes, como en el de la evaluación en general, hay un largo camino que recorrer. Quizá una de las necesidades más urgentes sea probablemente la evaluación de los inmigrantes escolarizados para la clasificación en cursos y niveles (según sus destrezas y la organización de los Centros) y la evaluación cuando terminan los cursos para decidir objetivamente la promoción escolar. Y respecto a los adultos son sin duda necesarios unos planteamientos generales (y probablemente a nivel nacional) tanto sobre el tipo de evaluación que requieren como sobre la certificación a la que aspiren dependiendo de sus necesidades y deseos de integración, trabajo, nacionalización, etc. De algo de todo esto pretende aclarar dudas y marcar caminos esta comunicación que va a desarrollarse en los siguientes apartados:

- 1 - Consideraciones generales.
- 2 - Evaluación de inmigrantes adultos.
- 3 - La evaluación en contextos escolares.
- 4 - Algunas muy necesarias cuestiones técnicas.
- 5 - Preguntas para el debate.

## 1. Consideraciones generales

A pesar de las necesidades detectadas, la situación actual de la enseñanza del español a inmigrantes y en consecuencia la de su evaluación no está suficientemente desarrollada como para poder disponer de planteamientos y teorías con los que argumentar y tomar decisiones de forma suficientemente objetiva y definitiva. Muchos son todavía los análisis e investigaciones que se necesitan para poder dar respuestas satisfactorias –y acordes para cada situación– a los interrogantes propios del proceso educativo y de aprendizaje de lenguas que se plantea el testing. Cosas concretas pero todavía borrosas como las siguientes -siguiendo a García Ramos (1989)-:

- si se han logrado cumplir los objetivos de aprendizaje de un curso,
- si ese aprendizaje ha cambiado la conducta lingüística de los alumnos suficiente y adecuadamente,
- si fueron efectivos los procedimientos empleados y si no lo fueron en qué destrezas se necesita más atención o qué tipo de estudiantes necesita mayor atención;
- en definitiva, si se siguieron los mejores caminos para obtener los objetivos establecidos,
- y si el no lograrlos es déficit de los mecanismos empleados o de los estudiantes.

Entre las funciones de la evaluación está dilucidar objetivamente conceptos tan importantes como los aludidos y hacerlo de forma que sus resultados puedan ser equiparados con los obtenidos en otros Centros, extendidos a otras poblaciones y ofrecidos como generales (cada uno dentro de su campo de dominio y según sus objetivos). Pero para llegar a tener tales datos es necesario diseñar antes unos productos (pruebas,



exámenes y tests) que ayuden a contestar a estas preguntas de forma fiable, objetiva, y justa es decir, sin discriminar o penalizar a determinados grupos de población (por edad, contexto social, formación educativa o cultural, etc.). Y para ello hay que partir de teorías fundamentadas en una práctica que explique los procesos de aprendizaje de la lengua -tanto en marcos de integración informales como académicos o laborales- y dotarse de descripciones de ese desarrollo de los procesos en niños, adolescentes y adultos –e igualmente en ámbitos familiares, sociales o académicos–. A la falta de trabajos, investigaciones, teorías y discusiones en estos sentidos se une la falta de adecuadas programaciones y objetivos debidamente graduados que tengan en cuenta diversificaciones de ritmos, necesidades y logros de aprendizaje procedentes de diferentes motivación, estilo, destrezas, técnicas o bagaje formativo inicial. Por eso este foro debe ser decisivo para sentar las bases de trabajos urgentes a corto, medio y largo plazo; y probablemente todas las comunicaciones y debates serán altamente informativos, aclaratorios y fructíferos en muchos de estos sentidos.

En cuanto al testing para inmigrantes en concreto, y a pesar de las carencias dichas, cabe comenzar a dar unos pasos –informativos unos, formativos otros y de trabajo y diseño algunos otros– que permitan sentar algunas bases para realizar trabajos serios en el futuro, trabajos que deberán ser estudiados y sometidos a análisis y discusiones, y que no podremos por el momento considerar definitivos. Por ello –desgraciadamente– desde aquí todavía no podemos dar una propuesta de contenidos ni diseñar un test piloto ya que este debe ser diseñado tras discutir objetivos, contenidos, materiales y necesidades acordes a las diferentes realidades de los Centros y a sus tipos de estudiantes. Incluso el diseño de un test más general que evalúe habilidades genéricas de inmigrantes a nivel regional o nacional necesitará de unos planteamientos y estudios previos. Pero sí podemos y desde aquí vamos a dar unas pautas de los pasos que se deben empezar a dar en estas direcciones esperando que con ello se sienten bases para realizar una labor necesaria, importante y urgente.

El testing se compone de productos (exámenes en general e ítems en particular) pero es en sí un proceso que comienza cuando se discuten y deciden los objetivos de un curso y se establecen los programas y los sistemas de medir si se han alcanzado, y que finaliza cuando se toman las decisiones oportunas basadas en esos datos. Por ello un test no es sólo el trabajo final de la cadena sino que tiene una enorme influencia sobre elementos básicos como la metodología de la educación porque define lo que se espera de un alumno y con ello ayuda a establecer los objetivos e incluso determina en gran medida los medios, el producto y el proceso de la educación. Pero para que funcione hay que hacer averiguaciones y comprobaciones con expertos, profesionales y usuarios (los propios estudiantes, en este caso inmigrantes) y someter los resultados a análisis estadísticos de fiabilidad.

Lo primero que deberíamos plantearnos es los tipos de tests que deseamos y necesitamos para la diversa población inmigrante existente en nuestro país, más allá de los test académicos del ámbito específico del aprendizaje de la lengua como lengua extranjera para estudiantes (generalmente universitarios pero no todos) que vienen de “vacaciones lingüísticas” para aprender en nuestro país durante cortos periodos de tiempo.

En realidad cada test sirve solamente para aquello que se pretende y se ha probado efectivo. Por tanto el paso previo para una adecuada evaluación de inmigrantes es plantearse para qué deseamos medir la lengua de los inmigrantes o extranjeros residentes en nuestro país. Si es para averiguar sus necesidades de formación académica pueden servir los modelos de evaluación propios de escuelas o academias acordes con los



cursos que en ellas se imparten. Si es para planificar una política educativa de apoyo a sus necesidades o para fines laborales se necesitarán otros métodos que analicen las habilidades funcionales reales de los hablantes y averigüen su nivel y su capacidad para enfrentarse a diversas situaciones según las labores sociales y profesionales en que vayan a desenvolverse. Si es para ayudarles en su sociabilización se necesitará un modelo que dé cuenta de sus capacidades de comprensión general, de expresión de ideas, explicación de opiniones e interacción. Si es para el desenvolvimiento en la vida diaria habrá que medir sus capacidades para hacerse entender en temas relacionados con los quehaceres habituales. Y entre todos ellos habrá que considerar casos en los que –aunque el nivel de lengua pueda ser alto– no haya ni se pretenda –en principio– conocimiento de la lengua escrita (Martínez, 2004).

Por ello hay que plantearse antes de nada preguntas como las siguientes:

¿hay que evaluar de la misma manera y con los mismos objetivos a un inmigrante que desee aprender la lengua con intenciones de equipararse a un hablante normal (no desearía que se notara mucho que es extranjero y quiere utilizarla en contextos cultos y formales) y a otro que se conforma con mantener relaciones sociales y laborales básicas y desenvolverse en la sociedad que le rodea (incluso aspirando a obtener la ayuda que sabe que todo hablante le va a brindar cuando sus estrategias de comunicación presenten lagunas o carencias)?, es decir, ¿hay que evaluar igual a un hablante orientado a la norma que a otro orientado a la comunicación?

Y para contestarlas propongo las siguientes reflexiones. Estamos acostumbrados a un tipo de enseñanza que prima lo académico e institucional y busca la excelencia, a través de cursos y diplomas y que tiene unos objetivos determinados propuestos y perseguidos por la institución educativa. Pero en el conjunto de los inmigrantes aprendices de una lengua podemos encontrar a gente que la haya aprendido o esté aprendiendo viviendo, conviviendo o trabajando y que entre sus objetivos no esté seguir aprendiéndola más allá de unos límites en los que sienta que sus necesidades están cubiertas, además de a otros cuyas aptitudes no den para más, junto con los que sí aspiren al más alto nivel de conocimientos y uso. Si a todos ellos quisiéramos medirlos con los mismos objetivos de una institución dedicada a enseñar la lengua formalmente les estaríamos exigiendo lo que no desean ni necesitan y solamente mediríamos el grado en que saben lo que queremos que sepan y no qué es lo que ellos realmente saben. Por lo tanto propongo en primer lugar definir qué se quiere medir, en qué situaciones y a qué personas, además de considerar y respetar lo que quieren estas personas que se les mida y respecto a qué situaciones y contextos de uso de la lengua, atendiendo a los diversos objetivos que tengan tanto del uso de la lengua como a la expectativa de uso de los resultados o certificados que les otorguemos en adelante.

Y respecto a los adolescentes y a la evaluación para clasificación y promoción escolar propongo plantearse los niveles de dominio que pudieran ser exigibles en cada edad y etapa escolar, la formación específica del profesorado, y la necesidad de (además del derecho a) una educación bilingüe, verdadera puerta al éxito en el aprendizaje de una segunda lengua si no también en el de otras muchas facetas educativas (Krashen, 1996).

## 2. Evaluación de inmigrantes adultos

La evaluación de la lengua y del desarrollo del aprendizaje del español por inmigrantes adultos no tiene –en principio– unas características distintas del resto de la evaluación para los estudiantes de



Español como lengua extranjera. Pero en el proceso de uso y aprendizaje de la lengua de los inmigrantes hay ciertos elementos que son de mayor relevancia y que permiten acercarnos a su evaluación desde perspectivas diferentes (Martínez, 2003). El fundamental es, ya lo hemos apuntado arriba, que no están en nuestro país aprendiendo la lengua, estudiando en cursos y con el objetivo de lograr el éxito académico, sino que la usan cuando la necesitan y principalmente para comunicarse, pasando por diferentes grados de interés y necesidad. Así en muchos casos habrá que tener en cuenta que la evaluación se hará sobre personas o grupos que usen la lengua con diferente preocupación de formalidad y corrección según la situación en que se encuentren, dependiendo de sus interlocutores o de la importancia que le den a cada interacción o situación.

En todo el arco de inmigrantes que están siendo y van a ser usuarios de nuestra lengua en los próximos años vamos a encontrarnos con muchos modelos y a todos ellos deberemos prestar atención. Desde quienes no tengan intención de aprenderla o usarla y necesiten siempre de un mediador traductor (caso que no nos ocupará por no existir ninguna L2 que analizar o medir), hasta quienes quieran usarla en contextos académicos o políticos abstractos de alto nivel e incluso de más alto nivel que la media nativa, pasando por otros que se limiten a usar su L1 mezclada con la L2, o quienes limiten el uso y aprendizaje de la L2 a contextos laborales o a determinados contextos sociales. Así que podemos y debemos acercarnos a la lengua, su aprendizaje y uso también en ámbitos no académicos, menos, poco o nada cultos y más variados; con todas las limitaciones de falta de excelencia, falta de ambición o fallo en el logro de objetivos que conlleva.

Pensando en estas cosas podemos preguntarnos -tomando como ejemplo la población de Norteamérica donde no hay un solo modelo de inglés hablado sino muchas variedades habladas por diferentes grupos de población (afroamericana, latinos, chicanos, spanish creole (Phillips, 1999), ya que cada comunidad indígena adapta y adquiere su propia versión de la lengua extranjera (Winford, 2003)- ¿cuántos modelos de la lengua española hay ahora en España?, ¿cómo es el español hablado en nuestro país por los diferentes grupos de inmigrantes que lo usan en diferentes contextos y áreas de la península y sus islas?, ¿qué grado de aprendizaje poseen estos?, ¿qué características se dan en sus variedades?, ¿en qué estadio del proceso de aprendizaje o adquisición se encuentran?

Evaluar la lengua en estos contextos de inmigración, donde el uso es diferente según múltiples factores sociales o económicos, y donde el principal objetivo es la comunicación y no el conocimiento de la lengua ni su aprendizaje (por lo que los aprendices se sienten completamente libres de adaptar la L2 de la manera que mejor se amolde a sus necesidades -Street y Giles, 1982-) es, antes que dar un certificado de nivel de conocimiento de la lengua (como los que se dan en escuelas o academias de lenguas o universidades) especificar la variedad de registros existentes y evaluar (o considerar) la adecuación de estos a los diferentes contextos de uso, respetando tanto la consideración que los hablantes tengan de su interlengua (sin duda deudora de factores internos de contacto lingüístico y de factores sociales externos socioeconómicos, familiares, etc.) como el modelo adoptado para su intercomunicación en cada momento del día, e incluso su derecho a mantener y usar en determinados contextos su propia lengua nativa.

Para nuestra tarea de definir una evaluación de la lengua de estas personas inmigradas necesitaremos un marco de situaciones y descripciones lingüísticas que incluya toda la variedad de



componentes y permita establecer diferentes usos y niveles interrelacionados. Cada situación puede también requerir de una batería de test diferente. Porque no es lo mismo medir globalmente la lengua que hacer medidas de la comprensión general, o de la mera comprensión de números, de la interacción, de la comunicación para desenvolverse en la vida cotidiana, de la comunicación en contextos laborales, de la escritura creativa o de la escritura de informes, etc. Sin duda son muy útiles a este respecto las descripciones generales de dominio de la lengua y las escalas ilustrativas de cada una de las habilidades del Marco de Referencia Europeo, (MRE 2002), así como las de ALTE (1991, ver [www.alte.org](http://www.alte.org)), DIALANG (2003, ver [www.dialang.org](http://www.dialang.org)) o ACTFL (1989, ver [www.actfl.org](http://www.actfl.org)). (Más adelante, en los apartados de la evaluación de adolescentes y de cuestiones técnicas se dan unas pautas de diseño de test que también pueden servir para adultos).

El marco que proponemos nosotros ahora y para adultos específicamente es más descriptivo que normativo, y principalmente oral; diciendo en todo momento cuáles son los logros y cuáles las limitaciones de cada caso. En este análisis deberemos primeramente describir el contexto social en que se encuentran los contactos lingüísticos y las características de la lengua que se produce o utiliza en ellos. Para eso debemos tener en cuenta los escenarios de uso de la lengua (personal, público, laboral, educativo), y los modelos de interacción social y de actuación empleados (productivo, receptivo, interactivo o mediador). Todos los modelos de evaluación existentes (de expresión y comprensión oral, de expresión y comprensión escrita, de clasificación para formar cursos, de rendimiento, etc.) son válidos para medir la lengua de estos extranjeros e inmigrantes en marcos de enseñanza formal en escuelas o academias. Pero para los casos no académicos nuestra propuesta es definir un tipo de test que responda a las necesidades tanto de los propios inmigrantes como de la sociedad de acogida. Por ello apostamos principalmente por test orales que midan el comportamiento lingüístico real en términos de conducta observable y que atiendan al interés principal que es la comunicación: test de entrevistas diseñados y dirigidos de forma que sirvan para obtener las mejores muestras posibles de la lengua hablada de cada candidato (Breiner-Sanders, 2000, Martínez, 2001). Estas pruebas abiertas orales –si con ellas pretendemos medir algo más que la habilidad para hacer frases gramaticalmente correctas contestando a diferentes temas– son las más difíciles de objetivar y de realizar con criterios homogéneos y las más expuestas a variaciones de resultados según sea la forma de realizarlas de cada evaluador, o la suerte de los candidatos. Por ello deben tener bien descritos los objetivos, los protocolos de corrección (claros y compartidos por todos los evaluadores) y el método de realización de entrevistas y elicitación de respuestas. Y los evaluadores deben compartir el sistema de hacer preguntas para la obtención de los datos e intentar no premiar con él ni la orientación sicolingüística del candidato (si es normativa o integrativa), ni cosas ajenas al verdadero nivel de lengua como conocimientos culturales, inteligencia u opiniones personales. Por lo tanto es fundamental pasar por procesos de formación y certificación que los homologue como evaluadores fiables.

Sería –de paso– muy interesante definir homogéneamente los niveles de acceso a universidades, cursos y puestos de trabajo desde una perspectiva de comunicación y comprensión oral (definiendo qué se puede hacer con la lengua y no qué se sabe de ella) tal y como se está determinando en otros países (Swender 2003). Se lograrían protocolos no discriminatorios y se daría un importante paso en la objetividad de medición de la lengua española que usan diferentes personas que además tienen variados y distintos objetivos. Cosa que hasta ahora no ha sido apreciada ni debidamente tenida en cuenta, ni siquiera en los cursos formales.



### 3. La evaluación en contextos escolares

Desde la publicación del MRE (aunque antes hemos citado tablas de referencia lingüística como las de ALTE o ACTFL) este es la referencia obligada para todas las prácticas pedagógicas oficiales de la Unión Europea. En lo que se refiere a la evaluación dedica todo un capítulo a ella aclarando las pautas a seguir y los principales conceptos para poder referirnos a ellos entre todos los profesionales de forma inteligible. Y establece unas escalas de dominio de la lengua en sus diversas habilidades (comprensión escrita y lectora y expresión oral y escrita). El marco es por lo tanto una herramienta de trabajo y un compendio de conocimientos para servirse de ellos al hablar y usarlo en discusiones y trabajos sin pretender ser el final de nada sino una ayuda al proceso. Por ello en el caso que nos ocupa, evaluación de inmigrantes adolescentes para clasificación y promoción escolar, lo tomaremos para plantearnos cómo evaluar a inmigrantes escolarizados (y a adultos) en los primeros niveles de aprendizaje y uso de la lengua llamados A1 y A2.

Se puede realizar un test de fin de curso en cada Centro para promoción de estudiantes (que debe ser sometido a unas comprobaciones técnicas) el cual –aún siendo diferente de otros test del mismo nivel diseñados en otros centros (con materiales y tareas diferentes)– sea homologable y pueda comparar sus resultados. El MRE precisamente está pensado para estas cosas, pues aunque no prefiere ningún modelo pretende que se puedan comparar los resultados obtenidos en cada lugar y momento gracias al uso de criterios iguales establecidos en sus escalas de dominio.

Igualmente se pueden diseñar test de otras habilidades atendiendo a las necesidades de cada Centro y a los diferentes intereses o características de sus estudiantes. Cada test deberá adecuarse (en las tareas, actividades y preguntas con que se componga) a los niveles descritos en el Marco con el objeto de poder comparar sus resultados con los de otros centros.

Lo mismo podemos decir respecto a pruebas de nivel para clasificación: cada una deberá diseñarse en función de los contenidos y de las programaciones del Centro, de las compartimentaciones de niveles y subniveles que se hayan establecido y de los objetivos que se hayan decidido en cada curso de acuerdo a sus especiales circunstancias, todo ello previamente definido partiendo del MRE como base general.

El grueso del trabajo para hacer cada uno de estos test está en estos pasos:

- a. selección de la habilidad o subhabilidades a medir y de las escalas a seguir,
- b. diseño de actividades y tareas,
- c. valoración de resultados mediante análisis de las respuestas dadas por voluntarios,
- d. comprobar mediante técnicas estadísticas y análisis de resultados y consultando tanto a expertos como a los propios candidatos en las fases de pilotaje del test (previas al uso definitivo del test) que las preguntas producen los resultados que se pretendía y no otros. Es decir, que se contesta de la manera que se esperaba y usando las destrezas que se pretende medir y no mediante otros caminos que nada dicen respecto a si se sabe o no lo que pretendemos. Por poner un ejemplo, en una pregunta de comprensión lectora puede haber quienes habiendo comprendido el texto no acierten la respuesta y otros



que sí contesten bien sin haberlo comprendido. Si sucede esto (y cosas así es muy fácil que sucedan si no se plantea bien un test y no se analizan sus resultados) la pregunta no servirá para nada y habrá que comprobar qué mecanismos de razonamiento están utilizando para responder (o qué trucos) y después rediseñar la pregunta hasta que efectivamente sirva para encontrar a aquellos candidatos que mejor hayan entendido el texto, frente a los que no lo entiendan bien, muy poco o nada. Esta es la labor técnica del testing que se ocupa de validar (dar certificado de validez, de que algo sirve para lo que se pretende) las preguntas y le da fiabilidad y objetividad.

Vamos a ver brevemente cómo se realizan estos pasos. Respecto al contenido de los test, y a los materiales y funciones o tareas de que se componga, no creo que sea tarea nuestra ahora la de proponer nada concreto (actividades, tareas, preguntas, etc). Sean las que sean, deben adaptarse a muchos factores (edad, escolarización o no, materiales trabajados durante el curso, contextos y funciones estudiadas y practicadas en el curso, programaciones, etc.). Cada profesor o cada Centro debe diseñarlas acordes a sus cursos y objetivos. También se puede plantear un test regional o nacional general pero no es momento ahora para darlo por hecho pues requiere previos estudios de cómo se produce la educación y a qué ritmos, con que contenidos se van a abordar, qué funciones se desean enseñar y que objetivos se van a primar sobre todos los demás cuando unos presenten más facilidad de aprendizaje que otros. Y es tarea larga que debe ser sometida a numerosas comprobaciones de validación técnica y estadística, etc. (las mismas que para obtener cualquier test fiable), además de interdisciplinar y probablemente cara (se necesitan equipos de diseño y discusión en los que participen profesores, técnicos, candidatos en prueba que representen a toda la población con la que vaya a usarse el test, expertos, etc.)

Por ejemplo, para un test del nivel A2 habrá que acudir a los cuadros del MRE y buscar entre ellos (ofrece cuadros de habilidades de cosas tan diferentes como expresión oral, expresión escrita, comprensión oral, comprensión escrita, dominio y riqueza de vocabulario, coherencia y cohesión, corrección, inteligibilidad, etc.) lo deseados para el test. A saber:

Dominio de léxico A2: **“Domina un limitado repertorio relativo a necesidades concretas y cotidianas”**.

Riqueza de vocabulario A2: **“Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia”**

Coherencia y cohesión A2: **“Enlaza grupos de palabras con conectores sencillos, como, por ejemplo, “y”, “pero”, y “porque”**.

Escritura general A2: **“Escribe una serie de frases y oraciones sencillas enlazadas con conectores sencillos tales como y, pero y porque.”**



Evaluar con referencia a estas escalas supone comenzar por escribir ítemes o preguntas y exámenes que mediante tareas, actividades y materiales (diferentes según diferentes grupos de población, mujeres, niños, adolescentes, adultos, en contexto laboral o académico, etc.) provoquen un discurso o una respuesta que sea prueba de la existencia de un conocimiento o de una habilidad determinados que se desea evaluar (de acuerdo con la escala –en este caso A2). El camino que proponemos para diseñar estos ítemes o preguntas y tareas y con ellos un examen consiste en leer e interpretar las escalas de cada nivel y proceder a pensar mediante qué preguntas o tareas se va a provocar una respuesta o discurso que queremos medir/evaluar. Luego habrá que pasar a la fase de validación del test, a comprobar si este es objetivo y fiable.

Por ejemplo, para medir la destreza de escritura del nivel A2 (que ya hemos visto que se definen en el MRE como: **“Escribe una serie de frases y oraciones sencillas enlazadas con conectores sencillos tales como y, pero y porque”** y **“Escribe cartas personales muy sencillas en las que da las gracias o se disculpa”**) unas tareas podrían ser pedir al candidato que escriba respecto a los siguientes temas:

- Has perdido tu cartera, unos libros o a tu perro (selecciona una cosa). Después de poner carteles por el barrio explicando el problema, una persona lo encuentra y te lo entrega. ¿Qué le escribirías para darle las gracias?
- Un amigo te ha ayudado mucho para buscar un piso, trabajo, etc. y se ha ido de la ciudad. ¿Qué le escribirías para darle las gracias?
- Describe cómo es tu país/ la gente de tu país/ tu ciudad (selecciona un tema).

Igualmente habrá que pensar en tareas para otras destrezas que quieran evaluar los conceptos de corrección, inteligibilidad, etc.

Tras administrar estas preguntas de test escrito en fase de pruebas entre una población de candidatos de similares características a los que finalmente deberán examinarse, se recogen las respuestas escritas (o discurso) producidas por el candidato (sea estudiante o adulto) y se procede a valorar en qué medida contienen los elementos que se buscaban y a juzgarlos con un sistema claro y compartido por todos los correctores. En la corrección hay que usar igualmente otras escalas relacionadas con la expresión escrita, que en este caso podrían ser las de “Dominio de vocabulario” (“Domina un limitado repertorio relativo a necesidades concretas y cotidianas”) y “Coherencia y cohesión” (“Enlaza grupos de palabras con conectores sencillos, como, por ejemplo, “y”, “pero”, y “porque”) entre otras. Habrá que llegar a un acuerdo sobre la manera en que vamos a analizar el discurso (las respuestas de los estudiantes) de acuerdo al cumplimiento de la tarea propuesta, según los criterios y escalas del MRE y sin dejarnos llevar por otros criterios no escritos o no considerados ni en el Marco ni en el diseño y objetivos del test.

Antes y hasta que aparecieron los enfoques comunicativos y las escalas de criterios como el Marco, se daban notas según el resultado del conjunto de todo un grupo de estudiantes de mejor a peor, según un ranking. Ahora se tiende a pensar que también el peor del grupo podría cumplir las escalas y por ello superar los objetivos del curso. Por lo tanto ahora se tiende a aprobar a quienes cumplan con las escalas o criterios externos



con independencia del lugar que ocupe en el ránking interno del grupo. Se mide por lo tanto con relación a unos criterios externos que pueden cumplir o no todos los estudiantes y no sólo con relación a un ránking que establece el mejor o el peor de cada curso. Es la única manera de que los resultados puedan ser homologables y comparables con los de cualquier otro candidato de otro Centro que haya realizado un examen similar aunque con tareas, contenidos o materiales diferentes.

Como propuesta general pensamos que cada profesor en su Centro y en reuniones con sus colegas debería intentar diseñar test de promoción basados en actividades de los diferentes niveles que imparta, adecuadas a la realidad social y cultural de sus alumnos y a lo trabajado en las clases y otros test de clasificación basados en lo que se haya planeado trabajar durante el próximo curso. Y con ello proceder a discusiones sobre si los contenidos y la dificultad de las tareas son apropiados y posteriormente a probarlo con poblaciones de estudiantes (en fase de pruebas) similares a las que luego lo usarán en serio como examen.

Estos test pueden ser tanto de gramática como de otro tipo, de preguntas cerradas o abiertas, siempre que cumplan con el objetivo de adecuarse a un nivel de dificultad claro y preestablecido. Luego si un estudiante es más habilidoso y aprende más rápido se le pasan los test de mayor nivel para averiguar si ha alcanzado o no sus objetivos o en cuál de ellos se ha quedado.

Espero que puedan surgir propuestas más concretas en los diferentes Centros según múltiples y variadas necesidades. Seguro que durante el taller se podrán aportar ideas sobre tareas, actividades, contenidos y materiales específicos para cada caso y discutirlos con profesionalidad.

#### **4. Algunos pasos previos -técnicos- para un diseño de test fiables y para la correcta formación de examinadores de pruebas abiertas de hablar y escribir.**

Al enfrentarnos al diseño de una buena prueba de evaluación nos vamos a encontrar con numerosos problemas debidos tanto a la naturaleza compleja de la materia a medir –la lengua– como a la de las actuaciones y discursos de los candidatos o a la forma y método de administrar las pruebas. La evaluación en sí requiere muchos análisis y planteamientos y se puede decir que en sí no es un producto sino un proceso. El proceso de este diseño es igual para todos los test. Hay que tener en cuenta una serie de importantes conceptos (que enseguida comentamos someramente) y dar una serie de pasos. Entre estos pasos destacan:

1. Contar con una serie de criterios iniciales y especificaciones de la prueba y de sus usuarios y objetivos.
2. Discusión de objetivos y habilidades a medir, prestando atención a las opiniones de profesores y expertos.
3. Planteamiento de tareas, actividades y materiales mediante los que se pretende medir.
4. Pruebas con nativos.
5. Pruebas con una población significativa de candidatos de las mismas características que aquellos para los que se haya pensado el test (edad, conocimientos culturales, nivel de educación, sexo, nacionalidad, etc.).



6. Discusión con usuarios (inmigrantes) respecto a cada ítem y test para intentar comprobar si se responde de la manera que pretende el diseñador del test o se ha respondido utilizando otros trucos que desvirtúen el resultado, así como sobre el formato, tiempos, estructura, etc. para ver si son aceptables por los usuarios.
7. Someter los datos o resultados de las pruebas a análisis estadísticos: análisis de las respuestas para ver si son las esperadas, si las preguntas (tareas y actividades) han provocado efectivamente el tipo de respuesta esperada. Y para probar que la pregunta la contestan efectivamente quienes más saben y no al revés o aleatoriamente: pues si la responde igual de bien gente con nivel y gente sin nivel no es buena para encontrar a quien sabe y diferenciarlo de quien no sabe). Y mediante comprobaciones estadísticas de esos resultados (sobre el grado de dificultad de cada pregunta).
8. Rediseño y posterior discusión de los nuevos ítems que haya sido necesario variar o reformar.
9. Comparaciones de los resultados del test con otras medidas conocidas de los estudiantes (por ejemplo preguntando a sus propios profesores si creen que el ranking de mejores a peores notas que haya resultado de un test es efectivamente la realidad de sus estudiantes o el test falla y no da resultados fiables).
10. Establecer adecuados criterios de corrección.
11. Formar adecuada y sistemáticamente a los evaluadores.

Y los principales conceptos a que aludíamos y que no se pueden olvidar son:

- Asegurarse de que el test sirve efectivamente para medir aquello que se pretende (Validación general del test y Validez de constructo) y de que va a provocar y a obtener las actuaciones y respuestas esperadas de los candidatos (Validez de apariencia y Validez de respuesta).
- Asegurarse de que los resultados son acordes con otras medidas conocidas de los candidatos y con valor predictivo (Validez concurrente y Validez predictiva).
- Asegurarse de que los ítems o preguntas de que constan se corresponden con lo que se desea medir (Validez de contenido) y están diseñados con la debida dificultad (Coeficiente de discriminación e Índice de dificultad) de forma que solo los puedan realizar bien las personas que tengan ese nivel y de forma sistemática (Fiabilidad).

Por ser una práctica común y muy mal realizada entre los profesores de lenguas en general la del diseño de ítems para test de gramática, quiero detenerme en explicar los principales problemas generales del diseño de ítems cerrados de respuesta de elección múltiple. Los demás conceptos técnicos no son en realidad difíciles de comprender ni de averiguar y requerirían simplemente unas cuantas horas de formación teórica y práctica (se intentarán explicar, aunque también brevemente, en el artículo final que se ha escrito para este encuentro. Ver para una explicación sencilla pero imprescindible el libro de Alderson, 1995).

El primer problema es que estas preguntas no suelen probarse con nativos que las validen ni mucho menos con estudiantes que prueben si efectivamente son lo difíciles que se pretende al diseñarlas. Luego pasa que muchas de estas preguntas pensadas como difíciles las logran responder estudiantes que apenas saben ni entienden la lengua o que hay otras preguntas que las logran responder estudiantes que apenas saben



mientras otros de más nivel de conocimientos las fallan. Los resultados así obtenidos— no hace falta mucha cabeza para darse cuenta— sirven de bien poco.

Las cuatro cosas más corrientes que suceden son:

- a. que un ítem aparentemente fácil sea en realidad muy difícil y sólo lo respondan bien personas con mucho conocimiento de la lengua;
- b. que el ítem no aporte información sobre el nivel de los estudiantes que lo contestan bien, pues casi todos o casi ninguno lo logren acertar (en este caso el ítem no discrimina, no diferencia a los que ya saben de los que saben poco o saben muy poco);
- c. también puede pasar que contesten mejor a un ítem los estudiantes que menos saben y no sepan contestarlo bien otros que tengan más nivel;
- d. y por último que entre quienes lo contestan bien haya gente de poco nivel, gente de nivel medio y gente de nivel alto, con lo que tampoco esa pregunta ayuda a diferenciar entre personas con conocimientos y otras sin ellos.

Ya vamos a terminar. Pero antes queremos hacer hincapié en otro aspecto siempre descuidado: el de la adecuada y sistemática formación de evaluadores. El examinador es la persona responsable de juzgar la actuación de un candidato en una prueba, sea objetiva o subjetiva. Si la puntuación obtenida en una prueba no es válida, el trabajo de diseño y administración de pruebas no vale para nada. Hay que poder confiar en los resultados de la pruebas y para ello hay que poder confiar en la manera en que el evaluador decide esos resultados, sobre todo cuando se basan en pruebas subjetivas (de hablar o escribir) que dependen de él totalmente. En una prueba de este tipo no hay una máquina que dé la puntuación final; mejor dicho la máquina es el evaluador y hay que conseguir mediante la adecuada formación, que su forma de dar los resultados sea siempre la misma y siempre igual a la del resto de sus compañeros evaluadores (que estén realizando la misma tarea). La inexactitud en la puntuación y la variedad en la interpretación de los datos son dos peligros importantes. La formación debe encaminarse, por lo tanto a:

- disminuir la variedad de interpretación de datos
- disminuir la inexactitud en la puntuación
- comprender las escalas e interpretar los descriptores de forma coherente.

Incluso para esto tan subjetivo existen formas de averiguar si dos evaluadores se están comportando de la misma manera entre ellos e incluso si la forma de evaluar de uno mismo es sistemática o errática. Como para todo lo apuntado aquí existen formas y fórmulas estadísticas fáciles de averiguarlo (remito nuevamente al libro de Alderson).

**Martínez Baztán, Alfonso**

Centro de Lenguas Modernas. Universidad de Granada

E-mail: [alfonsom@ugr.es](mailto:alfonsom@ugr.es)



**Cuestiones para el debate.**

1. ¿Cuál es la tipología de personas inmigradas que con mayor urgencia solicita o requiere ser evaluada? ¿Con qué fines?
2. ¿Por qué y con qué objetivos y necesidades queremos evaluarles?
3. ¿Deben ser evaluadas todas las personas inmigradas o solamente las que lo deseen por cuestiones personales o que deseen superar unos determinados objetivos?
4. ¿Deben ser iguales las evaluaciones para quienes aspiran a un desarrollo académico alto y a una integración sociolaboral, que para quienes solamente aspiran a desenvolverse en determinados contextos de su vida social o personal?
5. ¿Debemos evaluarles con nuestros objetivos o debemos respetar sus intenciones y necesidades de aprendizaje y sociabilización?
6. ¿A quien compete diseñar los test y certificar su validez y fiabilidad?
7. ¿Qué tipo de certificaciones debemos otorgar y qué homologaciones serían convenientes?
8. ¿Debemos integrar estos elementos: a. una evaluación formativa, social o educativa (que conlleva unas exigencias hacia el candidato); b. una evaluación descriptiva (deudora de su modelo de participación social)?
9. ¿Cómo debe integrarse la evaluación en la capacitación profesional, la enseñanza y otros contextos sociales y jurídicos?
10. ¿En qué medida la evaluación puede servir de estímulo a la sociabilización?
11. ¿Debe existir un centro de datos que recoja lo que se está realizando en este campo (diseño de test y certificaciones) a nivel autonómico y/o nacional?
12. ¿Qué instituciones deben encargarse del desarrollo de los diferentes modelos de evaluación con fines específicos y otros fines diversos?
13. ¿Quién debe formar a los evaluadores y cómo deben formarse?
14. ¿Qué perfil debe poseer un buen evaluador de la lengua de personas inmigradas?
15. ¿Cuál es la función social de la evaluación de la lengua?
16. ¿Cómo es el español hablado en nuestro país por los diferentes grupos de inmigrantes que lo usan en diferentes contextos y áreas de la península y sus islas?



**Referencias bibliográficas**

- ACTFL, (1989). *The Oral Proficiency Interview Tester Training Manual*. Kathryn Buck (Ed.). Inc, Hastings-on-Hudson, New York.
- Alderson et all, (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- ALTE: [www.alte.org](http://www.alte.org)
- ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (1994), *The ALTE Code of Practice*. Cambridge, ALTE.
- Breiner-Sanders, K.E., Lowe, Jr, P., Miles, J. Swender. E. (2000). *ACTFL proficiency guidelines-speaking, revised*. *Foreign Language Annals*, 33, 13-18.
- DIALANG: [www.dialang.org](http://www.dialang.org)
- García Ramos, J. M. (1989). *Bases pedagógicas de la evaluación*. Síntesis. Madrid.
- Krashen, S. (1996). *Under attack: The case against bilingual education*. Culver City, CA: Language Education Associates.
- Martínez Baztán, A. (2001). Evaluación Oral de la Competencia Lingüística. En: *Actas de las XI Jornadas de Lingüística Aplicada*. Universidad de Granada.
- Martínez Baztán, A. (2003). Claves en la enseñanza de español a inmigrantes. En: *El aula multicultural. La enseñanza del español como lengua extranjera*. Instituto de Estudios Ceutíes, 15-23.
- Martínez Baztán, A. (2004). Consideraciones sobre la evaluación del español como lengua de personas inmigradas adultas. En: *Glosas didácticas, Revista electrónica de la universidad de Murcia*, núm 15, Diciembre 2004.
- MRE. (2002). *Marco de referencia europeo para el aprendizaje, la enseñanza y la evaluación de lenguas*. Instituto Cervantes. Madrid.
- Phillips, J. K. (1999). *Standard for Foreign Language Learning in the 21st Century*. Allen Press Inc. Lawrence, KS.
- Street, R. L. y Howard Giles, (1982). Speech accommodation theory: a social cognitive approach to language and speech behaviour. En: Michel E. Roloff y C. R. Berger Eds. *Social Cognition and Communication*, 193-226. Beverley Hills: Sage.
- Swender, E. (2003). *Oral proficiency testing in the real world*. En: *Foreign Language Annals*. 36. num. 4, p- 520-527.
- Winford, D. (2003). *An introduction to contact linguistics*, p. 224. Blacwell. London.

