La creación de una prueba de clasificación estandarizada y automática

Iñaki Tarrés
Instituto Cervantes de Berlín

Este artículo afronta una de las necesidades organizativas más frecuentes de los centros de enseñanza de ELE: la clasificación o nivelación de alumnos con conocimientos previos, y su ubicación dentro del programa de cursos del centro. El punto de vista metodológico desde el que se enfrenta este tipo de evaluación, se relaciona con la lingüística cognitiva. Parte de un análisis de un grupo de pruebas de clasificación existentes, pasa a la propuesta de una serie de criterios para la redacción de un test estandarizado y automático, y termina con la presentación de una propuesta de prueba de clasificación automática que se está usando en la actualidad, desarrollada desde un punto de vista cognitivo. Espero que este artículo pueda contribuir a mejorar el proceso de nivelación de los alumnos que llegan a los centros de enseñanza. Es nuestra responsabilidad como docentes poder ofrecer instrumentos de evaluación tan fiables como se espera de la institución que a sus ojos representamos, y responder adecuadamente a sus expectativas.

Los centros de enseñanza de ELE nos vemos continuamente en la necesidad de ubicar dentro de nuestra estructura de cursos, a los alumnos que llegan con conocimientos previos. El hecho que influyó de forma determinante en la necesidad de establecer un nuevo modelo de prueba de clasificación en mi centro del Instituto Cervantes de Berlín no fue, frente a lo que pudiera esperarse, la situación pedagógica derivada de la pandemia que empezó en 2019, sino el proceso de digitalización de los procedimientos administrativos de matriculación de los alumnos. Este proceso empezó antes, y la pandemia solo hizo que se acelerara. El hecho es que antes de esa fecha cada vez más alumnos se estaban matriculando de forma autónoma en cursos de distintos niveles a través de la plataforma de pago habilitada por la sede del Instituto Cervantes, lo que hacía que cada vez más algunos alumnos debieran cambiar de curso por estar mal situados. Necesitábamos una herramienta digital de acceso directo por parte del alumnado que le diera un valor claro con el que pudiera operar, es decir, seleccionar sin ayuda de un profesor y adecuadamente el curso que debía hacer dentro de itinerario del centro. Por eso, desde antes ya del inicio de la pandemia estábamos trabajando en la dirección que presentamos ahora en este artículo.

En este artículo usaré como equivalentes las expresiones «test de nivelación» y «prueba de clasificación».

ANÁLISIS DE PRUEBAS DE CLASIFICACIÓN

En este apartado voy a presentar el resultado del análisis de las pruebas de clasificación de catorce centros de la red del Instituto Cervantes usadas hasta el año 2019. Deseo agradecer expresamente a los jefes de estudios y responsables académicos de los centros que amablemente quisieron colaborar en este proyecto: José Roberto Ortí Teruel; Ana Isabel Reguillo Pelayo; Penélope Villar; Nuria de la Torre García; Rafael Mellado Jurado; Pablo Martínez Gila; Eva Lloret Ivorra; Patricia Mengual Verdú; Mª Teresa de Santos Borreguero; Carmen Martínez; Luis Martín Carretero; Chelo García Manzano; Mercedes Barrios Pérez; Myriam Pradillo; Manuel Fernández Fernández; María Belén Castillo Sedano; Marina Alonso Abal; Mª Teresa de Santos Borreguero; Rosa María Pérez Bernal.

Las pruebas analizadas, como digo, corresponden a usos anteriores a la pandemia. Muchos de los valores que voy a describir han dejado de ser actuales, y la mayoría de los centros ha revisado el formato de test que utilizan, pero creo que el análisis refleja los usos y creencias de los centros de enseñanza de ELE en relación con la clasificación de alumnos con conocimientos previos. Son descriptores genéricos.

1. Se veía poca variedad de formatos y de tipo de tareas. A grandes rasgos se puede decir que los test consistían en baterías de ítems formados por frases o microdiálogos con huecos y respuesta cerrada de opción múltiple (entre 3 y 4 opciones), o por textos con huecos y el mismo tipo de respuesta, a lo que se añadía en ocasiones una tarea de expresión escrita. Casi no había tareas de comprensión de lectura y en ningún caso tareas de comprensión auditiva. Ningún test utilizaba imágenes.

2. No se veía correspondencia entre el número de horas de clase por nivel (o para decirlo de otra manera, el volumen de contenidos curriculares) y el número de ítems de las pruebas de clasificación. Centros con el mismo número de horas por nivel ofrecían cantidades de ítems muy diferentes.

3. Tampoco se veía una relación uniforme entre el número de ítems y el tiempo disponible para la realización de la prueba. Además, aunque los propios ítems podían tener más o menos la misma extensión en número de palabras, se establecieron tiempos muy diferentes para su realización. No parecía que hubiera algún tipo de fórmula uniforme para calcular el tiempo necesario para la realización del ítem. Tampoco parecía que se tuviera en cuenta el tiempo como factor de dificultad.

Veamos por ejemplo el nivel A2 a partir de la información ofrecida por los propios centros. Teniendo en cuenta que se realiza además una entrevista oral que dura unos 10/15 minutos, podemos calcular el tiempo aproximado disponible para cada ítem a partir de la información proporcionada por los centros para toda la prueba.

Nivel A2	Centro 1	Centro 2	Centro 3	Centro 4	Centro 5	Centro 6	Centro 7
N.º de horas por cuso	30	30	40	30	30/60	30	40 (20)
N.º de cursos	2	3	2	5 (A1 y A2)	2	2	3 (A1 y A2)
N.º de ítems	51	65	30	20	16	13	70
Tiempo total	30-45 min.	50 min.	30-45 min.	60 min.	15-30 min.	45 min.	45 min.
Tiempo por ítem	35″	32″	1′	2′ y 15 ″	56″	2′ y 18″	25″

Figura 1a

Nivel A2	Centro 8	Centro 9	Centro 10	Centro 11	Centro 12	Centro 13	Centro 14
N.º de horas por cuso	60	30/60	30	60	60	25	30
N.º de cursos	2	4	4	3 (A1 yA2)	2	3	5 (A1 y A2)
N.º de ítems	20	10	3	20	8	32	20
Tiempo total	45 min.	45 min.	¿?	¿?	30 min.	¿?	45 min.
Tiempo por ítem	2′	3′	¿?	¿?	1′ y 56″	¿?	1′ y 30″

Figura 1b

4. Había circunstancias del centro, como el número de horas de los cursos, los manuales o el itinerario, que parecían no influir en la conformación del formato de test. Centros con una L1 común, centros con itinerarios parecidos, o centros que compartían manual, tenían cantidades de ítems muy diferentes.

5. Desde el punto de vista formal, las hojas de respuestas, con alguna excepción, no permitían reunir datos para crear el perfil del candidato como aprendiz de ELE previo al inicio de los cursos en el centro, lo cual podría ser muy útil para el desarrollo del centro, pues hubiera permitido realizar, entre otras cosas, un seguimiento del desarrollo de los candidatos.

6. La gestión del test variaba también bastante, lo cual es lógico pues debían adaptarse a las necesidades del centro y a la idiosincrasia del país o de la ciudad. En el momento del análisis solo un test se podía realizar a distancia o en formato electrónico. Este es el elemento que más ha cambiado en la gestión de la prueba de clasificación.

7. En general, las pruebas de nivel analizadas no evaluaban competencias sino solo conocimientos, y de forma muy sesgada, centradas en los contenidos gramaticales casi exclusivamente y en menor medida en los léxicos. Transmitían, de esa manera, una concepción de la lengua fundamentalmente estructuralista. Este aspecto me parece fundamental, y lógicamente se presta a debate. Además, su posible revisión precisa de una decisión firme por parte de los centros: el test es también la primera carta de presentación del centro, la primera toma de contacto con el concepto de lengua y de enseñanza de la institución. Volveremos más adelante sobre este aspecto.

8. Sorprende la ausencia en muchos casos de un documento de especificaciones del test de nivelación que estableciera los perfiles de los candidatos según los niveles, los criterios con los que se habían construido las tareas y la manera de interpretar los resultados.

LA PRUEBA DE CLASIFICACIÓN COMO INSTRUMENTO DE EVALUACIÓN

Los test de nivelación estandarizados, ya sea automáticos, como en esta propuesta, o gestionados por un profesor, se configuran dentro de los test de diagnóstico, y como cualquier instrumento de evaluación, buscan recabar datos para poder tomar una decisión: la ubicación de un candidato en un curso concreto dentro de una estructura curricular que usa unos materiales concretos, es decir, un manual determinado, y tiene un itinerario específico con su correspondiente distribución de contenidos curriculares. En el concepto que estamos desarrollando, el usuario del test es el propio alumno-candidato, el instrumento lo usa para obtener una medida que le permita por sí mismo y con seguridad elegir el curso del centro más apropiado a su nivel de conocimiento o su nivel de competencia, según se configuren las tareas en una dirección o en otra, o, idealmente, en ambas direcciones al mismo tiempo, aun cuando sea difícil que un candidato tenga un nivel de competencia y un nivel de conocimiento equiparables. El uso de estrategias de compensación, así como el conocimiento de lenguas afines o la práctica en cursos previos y una especial sensibilidad lingüística puede hacer que un candidato o candidata resuelva satisfactoriamente una tarea cuyos conocimientos lingüísticos evaluados ese candidato nunca ha visto. La configuración del test deberá, como veremos más adelante, tener en cuenta este tipo de circunstancias. Sea como sea, se concibe un instrumento de evaluación cuyo usuario es el propio alumno-candidato: es quien administra el test y el beneficiario del valor resultante.

Así, como cualquier tipo de prueba, el test de nivelación estandarizado y automático debe cumplir tres requisitos fundamentales: validez, fiabilidad y viabilidad.

VALIDEZ

El resultado debe representar una medida que el candidato pueda interpretar de forma clara y limpia. No puede ser demasiado compleja de manera que lleve a la desorientación del candidato, y debe ser suficientemente precisa como para permitirle crear una imagen fiable de su competencia lingüística como hablante de ELE en relación con la institución en la que quiere aprender español, aun cuando esta imagen pueda ser contradictoria con la que tiene previamente. El valor establecido por el test automático debe quedar guardado en el correspondiente perfil, y debe ser recuperable en cualquier momento tanto por el candidato como por la institución. El candidato debe saber desde el principio que la validez del resultado se circunscribe a la realización de un curso en esa institución, con lo que no es un valor absoluto de su competencia lingüística. Es decir, no certifica su competencia, sino que indica una posición en un itinerario.

Otro factor que puede influir en una interpretación fiable del resultado del test es el factor educativo. Me refiero con esto a la manera como ha aprendido español, su currículo como aprendiente de ELE, los cursos que ha realizado anteriormente el candidato, los exámenes que haya podido hacer y los manuales que haya podido utilizar, todo lo cual configura una memoria activa que interviene en la imagen que pueda tener de su nivel de ELE, así como su capacidad para autogestionar el test.

La validez no puede verse afectada por circunstancias como el tiempo disponible para completarlo o las veces que se repita. Dado que el candidato es quien administra la realización del test, muchos profesores ponen instintivamente en duda la validez un test autónomo.

Finalmente, hay una serie de factores afectivos que influyen en el uso o la interpretación que el alumno-candidato vaya a dar del valor resultante de la prueba. En primer lugar, y como señalaba más arriba, la imagen que un candidato pueda tener de su propio nivel de ELE, que en ocasiones difiere de lo que el test refleja. En segundo lugar, los motivos para hacer un curso, que constituyen el mayor generador de especificidad. En tercer lugar, el hecho de que el test de nivelación es el primer paso en la satisfacción de las expectativas y necesidades que llevan a una persona a buscar un curso de ELE. Hay que insistir en el hecho de que no se trata de certificar el nivel del candidato, sino de ofrecer al futuro alumno una herramienta que forma parte del proceso que conduce a la satisfacción de sus necesidades. En este sentido, se puede reforzar en el candidato la conciencia de que realizar el test de forma honesta es garantía de validez, pero a la hora de configurar el test se deberá tener en cuenta los posibles «errores» de administración de la prueba, y establecer determinantes como el tiempo o la aparición aleatoria de respuestas, dentro de lo que la aplicación informática permita, que lleven a minimizar el efecto de una administración desviada por parte del candidato.

FIABILIDAD

A la hora de nivelar a un candidato en una interacción oral, los profesores no solo tenemos en cuenta sus conocimientos lingüísticos, sino que hacemos una suerte de compensación entre sus conocimientos y sus competencias y habilidades, a lo que añadimos toda la información que podemos recabar sobre cómo ha aprendido español, para qué quiere hacer un curso, qué otros idiomas sabe, en qué circunstancias va a tener que aprender español, si esas circunstancias son muy diferentes a las que ha tenido en su aprendizaje previo, cuántos años tiene y cuántos años más quiere seguir aprendiendo. Todo ello nos permite configurar el perfil del alumno y alcanzar una nivelación fiable. En otro tipo de evaluaciones estandarizadas como los exámenes DELE o más recientemente los exámenes del sistema SIELE, en los que nuestra intervención es mucho menor, ese tipo de informaciones explícitamente quedan fuera de evaluación. Para nivelar a un candidato de manera fiable necesitamos saber algo más que sus conocimientos y habilidades lingüísticos, pero el hecho es que estamos diseñando una herramienta de uso autónomo por parte del candidato, y ese tipo de informaciones no son reductibles a valores matemáticos, aunque indirectamente sí se pueden incorporar al test, como veremos más adelante.

Uno de los factores que influyen en la fiabilidad de un test de clasificación es la selección de contenidos evaluados. Ante la imposibilidad, como parece lógico para no extender innecesariamente la prueba, de evaluar todo el contenido curricular de un nivel, se impone la necesidad de establecer una selección de contenidos representativos. Para ello se puede partir de una secuencia común de contenidos presente en la inmensa mayoría de los libros de español, que son los que en última instancia suelen establecer qué se trabaja en cada curso. Si un candidato demuestra tener cierto contenido o cierta habilidad, podemos esperar que tenga otros contenidos u otras habilidades que se suelen trabajar asociados al primero, con lo que no será necesario evaluar los contenidos asociados para mantener la fiabilidad del test. En cualquier caso, la selección y secuencia de contenidos estarán determinadas por los contenidos que se den en los cursos, y no por una secuencia ajena a la propia institución.

En otro orden de cosas, dado que los cursos son el referente, será necesario tener en cuenta el perfil de los grupos que realizan esos cursos. Si los cursos están definidos como cursos para estudiantes universitarios, por ejemplo, o como cursos con fines específicos, estas descripciones deberán verse reflejadas de alguna manera en el test, no pudiéndose aplicar un test a un candidato que no va a realizar el tipo de cursos para los que se ha diseñado el test.

VIABILIDAD

El test tiene que ser realizable dentro de los límites cognitivos y afectivos del candidato tipo que se presenta para hacer un curso en nuestra institución, con unos

límites de tiempo adecuados, y una aplicación informática adecuada. Es necesario que haya una relación equitativa entre el número de ítems, y la dificultad cognitiva de los mismos dentro del entorno pedagógico, pero también del tecnológico, en el que se lleve a cabo el test. No se debe dejar de tener en cuenta qué aplicación o qué herramienta digital va a actuar de soporte de los ítems y va a permitir su administración autónoma por parte del candidato. Por ejemplo, la presencia de textos demasiado largos puede ser un inconveniente en aplicaciones que puedan abrirse con un teléfono móvil. Otros ejemplos serían una cantidad excesiva de ítems o la necesidad de saltar a otras aplicaciones o usar recursos tecnológicos auxiliares como audios o imágenes, todo lo cual puede influir en la viabilidad del test, pues la capacidad tecnológica del usuario puede acabar influyendo en cómo se hace el test, y finalmente en su resultado. Además, el test debe ser viable dentro de los límites impuestos por la propia institución educativa.

OBJETIVIDAD Y SUBJETIVIDAD EN LA PRUEBA DE CLASIFICACIÓN

En general es creencia bastante extendida entre los profesores de ELE que el hecho de que un candidato acierte en la selección de la opción correcta de respuesta de un ítem formado por una frase o un microdiálogo con un hueco no significa que sepa realmente qué está eligiendo, y mucho menos representa su nivel de adquisición o de asimilación del contenido lingüístico evaluado en ese ítem. Y más aún: no da una información fiable sobre su capacidad para usar ese contenido en un texto articulado o en una interacción oral. Cualquier profesor de ELE firmará sin dudar la afirmación de que solo una prueba oral es capaz de medir las capacidades lingüísticas de un candidato con conocimientos o habilidades previos a la hora de establecer qué curso debe hacer dentro del itinerario de su centro. Lo normal era, hasta el advenimiento de la automatización de los procesos de matrícula, que se evaluara al candidato en función de esa conversación con un profesor del centro preparado para esa tarea, y no tanto en función de los resultados del test aparentemente objetivo, aunque no por ello se dejaba de usar también este instrumento de evaluación. Aunque las tareas de selección múltiple vienen acompañadas del adjetivo "objetivas" porque solo hay una respuesta posible, la mayoría de los profesores pondrá en entredicho que estas sean fiables a la hora de establecer el nivel de un alumno-candidato. En este sentido, ¿para qué hacerle pasar por una prueba de este tipo si finalmente lo importante son las capacidades medidas en la prueba oral, y no el conocimiento que pueda demostrar por otras vías?

Por un lado, creo firmemente que un test de nivelación es una herramienta útil para establecer la posición de un candidato en el itinerario de un centro si este test cumple ciertos requisitos ajustados a las circunstancias actuales, es decir, ajustado a su viabilidad, entendiendo esta como el factor que garantiza los otros dos criterios, validez y fiabilidad; por otro lado, las circunstancias nos obligan a hacer el esfuerzo en confeccionar una herramienta válida y fiable, cuyo resultado puede eventualmente verse refrendado por una prueba de interacción oral con un

profesor, pero que en principio no deberá necesitar esa confirmación más que para evaluar la eficacia, validez y fiabilidad del instrumento. Sostengo que la clave está en la concepción y configuración del test entendido como un todo, no solo en la de los ítems de forma individual, y en el hecho de que se ha de realizar de manera autónoma por parte de un alumno-candidato del que podemos presuponer ciertos hábitos y cierta trayectoria.

Por otro lado, la nivelación de candidatos por parte de los profesores de un centro tampoco ha demostrado una objetividad «infalible», pues no se realizan con unos mínimos de uniformidad. Los perfiles ideales del candidato de cada nivel que tienen los profesores no son compartidos por los propios profesores pues en pocos casos se hace en el centro educativo un trabajo interno de uniformización. Por ejemplo, los contenidos mínimos de un nivel exigibles a un candidato para entrar en un curso, base de cualquier documento de especificaciones de una prueba de nivel, pocas veces se ven establecidos por el claustro de profesores del centro. Todo ello lleva a que se encuentren a diario casos de candidatos mal nivelados por los profesores a través de una prueba oral. La nivelación con una prueba estandarizada puede permitir obtener un resultado suficientemente fiable, objetivo y, sobre todo, uniforme: no es admisible que el mismo candidato sea nivelado con diferentes criterios y con diferentes instrumentos, o que dos candidatos con perfiles y motivaciones idénticos acaben realizando cursos de niveles diferentes por haber sido nivelados por profesores diferentes. Ello tendrá como consecuencia diferencias sustanciales en los resultados. El objetivo prioritario es siempre una nivelación uniforme y consistente.

ESPECIFICACIONES DE EXAMEN

Como cualquier instrumento de evaluación, la configuración de una prueba de clasificación deberá empezar en la redacción de un documento de especificaciones de examen. El CVC establece qué debe cubrir este documento. Aunque la verdadera prueba de validez y fiabilidad del test es que el candidato esté bien nivelado, el documento de especificaciones sigue siendo un instrumento necesario para medir esos dos criterios. Como decía más arriba, debe tener como base el programa de cursos, establecido probablemente a partir de los manuales de clase, y más lejanamente a partir del plan curricular decidido por la institución, en nuestro caso el PCIC, pero creo que dicho plan curricular no puede ser la referencia primera. La prueba de clasificación tiene una utilidad eminentemente práctica, y está orientada hacia su usuario primordial, el futuro alumno. También debe establecerse de común acuerdo por los profesores del centro, aunque directamente no intervengan en la gestión de la prueba.

Es necesario seguir los pasos habituales en la redacción de cualquier procedimiento evaluador (según Ortega 1995):

1.º establecer qué evaluaremos;

2.º establecer cuál será el procedimiento para recoger datos y evaluar lo observado;

3.º determinar con qué escalas asignaremos valores a lo observado.

En cuanto al primer punto, creo necesario insistir en que los contenidos evaluados deben definirse con claridad y en común acuerdo por todos los responsables académicos del centro. También deben ser coherentes con la prueba oral, en el caso de que la haya, de manera que no se evalúen dos veces los mismos contenidos y destrezas. Es fundamental someter todo el proceso a revisión por parte de los profesores del centro. Este proceso debe incluir la validación del test.

Volviendo a algo que apuntaba más arriba, tanto el formato como los contenidos evaluados suponen una primera presentación de la institución frente al candidato. El futuro alumno está infiriendo la metodología y la imagen de la lengua que tiene el centro a través del tipo de test de nivelación que hace. En este sentido, en la propuesta que comentaremos más adelante se han incluido junto a los ítems imágenes de los espacios del centro (aulas, secretaría, cafetería, etc.) para que el alumno se familiarice con el centro.

CRÍTICA DEL EJERCICIO DE HUECOS

En relación con el procedimiento evaluador, entiendo que el formato de ítems de respuesta múltiple como único procedimiento es insuficiente, aunque práctico a la hora de obtener un valor numérico, y se impone al menos un formato con variación de tipos de tareas para que los cambios cognitivos de una tarea a otra lleven al candidato a usar realmente sus conocimientos para resolver dichas tareas. También se impone la necesidad de que los ítems estén relacionados entre sí para que la atención del alumno-candidato se dirija al contenido y no a la tarea, pero de tal manera que solo el conocimiento que queremos comprobar permita realizar dicha tarea.

Los estudios pioneros de percepción de los años veinte del siglo xx descubrieron que cuando se nos presenta una serie de puntos, vemos instintivamente la línea que forman, aunque los puntos no estén unidos entre sí. Al mismo tiempo, es un hecho conocido que el cerebro tiene la capacidad de terminar las frases. De esta manera, los patrones de percepción sensorial podrían corresponderse con patrones de percepción lingüística, es decir, tenemos almacenados en la memoria lingüística esquemas sintácticos generales, así como combinaciones léxicas frecuentes que conforman esos patrones que a su vez nos permiten, por ejemplo, anticipar palabras. Si se nos presenta un fragmento de discurso (una frase, un microdiálogo) y a su vez distintas opciones para completar ese fragmento, seremos capaces de seleccionar el que completa adecuadamente el fragmento haciendo uso de nuestro conocimiento natural de la lengua en la que está codificado el fragmento. Esa capacidad mental sería la que fundamenta los ejercicios llamados

coloquialmente «rellena huecos». Si un alumno ha adquirido una estructura gramatical concreta, será capaz, tal y como haría un nativo, de seleccionar la opción que corresponde a esa estructura. Lo mismo vale para ejercicios que activan el conocimiento léxico. Sin embargo, en la práctica no se comprueba que suceda así.

Este tipo de ejercicios ha sido justamente denostado en los últimos años, y de hecho los exámenes para la obtención del DELE hace muchas convocatorias que dejaron que usarlos para evaluar la competencia de un candidato en un nivel concreto, pues probablemente no evalúan competencias, sino solo conocimiento declarativo. Desconozco los argumentos que los redactores de exámenes DELE usaron para tal decisión. Presentaré los míos.

1. Los famosos «rellena huecos» corresponden a una concepción estructuralista de la lengua en la que el sistema está formado por piezas que encajan perfectamente unas en otras. Puedo sacar una pieza de una unidad mayor y comprobar si el alumno reconoce la que encaja entre varias opciones, o incluso generarla, en ambos casos a partir de su conocimiento de la lengua, puesto que conocer un idioma significa encajar piezas de distintas unidades para generar muestras de nivel superior. Esta concepción de la lengua no la comparten los profesores, pero tampoco instituciones como el Instituto Cervantes o el Consejo de Europa (léase, el MCER). Sin embargo, como comprobamos más arriba, una gran mayoría de pruebas de clasificación analizadas siguen usando este tipo de ejercicios.

2. Se trata de un tipo de ejercicio muy ritualizable. Es decir, de forma mecánica los candidatos seleccionan las opciones o escriben las palabras que faltan sin pensar realmente en lo que están haciendo, sin necesidad de activar todo su conocimiento lingüístico. Es más, al no haber vínculo entre los ítems y el candidato, y entre la propia tarea y el candidato, el ejercicio se realiza de una forma desligada. No es una tarea significativa, no hay implicación por parte del candidato, su conocimiento lingüístico no es realmente necesario, o no se activa, para la realización de la tarea.

3. Las baterías de frases con huecos suponen una actividad mental agotadora para el mejor de los candidatos si intenta hacer cada ítem atendiendo a lo que se le pide, pues supone un prolongado esfuerzo de abstracción, de análisis y de memorización de reglas que requiere una concentración sostenida, lo que lleva a muchos profesores a reclamar ejercicios de duración apropiada, es decir, corta, lo que reduce las posibilidades de evaluación de suficientes contenidos representativos del nivel. El resultado del test automático, si se prolonga excesivamente, pierde entonces su fiabilidad a causa del cansancio que la prueba genera, llevando al participante a seleccionar opciones sin atender demasiado a lo que pone en el ítem.

4. Como comentábamos antes, la prueba de clasificación es el primer contacto del alumno con la metodología de la institución en la que va a aprender español: infiere el concepto de lengua y experimenta el sistema de evaluación, y de alguna manera, de enseñanza, y por tanto, percibe algunos conceptos metodológicos fundamentales de dicha institución. Esta función de presentación de la

institución frente al alumno se puede aprovechar para dar informaciones y para presentar una imagen concreta que le interese proyectar a la institución.

Todos estos factores hacen que, como digo, las baterías de ítems consistentes en frases con huecos y una única tarea repetida, no sean, en principio, apropiadas para prueba de clasificación de gestión autónoma. Sin embargo, y en contra de los argumentos expuestos, hay que admitir que hoy por hoy resulta poco factible desarrollar una prueba de clasificación automática a través de una aplicación informática online que los excluya completamente. Leemos, por ejemplo, en un artículo al respecto del profesor Foncubierta, que este reclama creatividad por parte del profesor a la hora de diseñar este tipo de ejercicios, sin llegar a rechazarlos de plano. Por ello, en la propuesta que presentamos a continuación hemos seguido usando parcialmente ejercicios de «rellena huecos» para algunos aspectos de la lengua que por su carácter estructural sí pueden prestarse a esta tipología. Además, otro argumento a su favor es que muchos alumnos los conocen y los reconocen, por lo que su presencia puede tener un efecto psicológico positivo en la realización de la prueba.

La mayor ventaja de los ejercicios de respuesta cerrada en una prueba de clasificación automática en la que no interviene un profesor, es su facilidad a la hora de establecer el resultado de la misma, pues se genera fácilmente un valor numérico, no descriptivo, que permite al alumno satisfacer el objetivo de la prueba, es decir, seleccionar el curso más apropiado simplemente operando con ese valor sobre una escala de cursos y valores numéricos asociados: «Si usted ha obtenido x puntos puede realizar un curso de nivel x.x...». La confección de la prueba de clasificación tiene que conseguir un equilibrio entre distintas características aparentemente contradictorias:

- unos mínimos de fiabilidad y validez, determinadas tanto por la tipología y confección de las tareas como por la duración de la prueba;
- un nivel de exigencia cognitiva que mantenga la atención y no genere cansancio;
- un procedimiento fácil de administrar por parte del alumno que genere un valor que le permita alcanzar el objetivo de la prueba;
- la representación de un concepto de lengua acorde con el posicionamiento al respecto de la institución;
- una presentación de la metodología y de la imagen de la institución.

BREVE RESUMEN

El establecimiento del curso que acaba realizando un candidato responde muchas veces a circunstancias en la práctica tanto o más poderosas que los requerimientos puramente metodológicos. Su impacto distingue esta evaluación de la realizada por, por ejemplo, las pruebas de nivel tipo DELE o los exámenes de

rendimiento, en los que, como decíamos más arriba, se entiende que la evaluación no atenderá a factores individuales. Por el contrario, los test de nivelación sí deben atender, además de a criterios objetivables (contenidos y competencias lingüísticos), a factores individuales. El buen test de clasificación deberá ser capaz de combinar todos estos factores a la hora de establecer el nivel que le corresponde al alumno-candidato dentro de la estructura de su centro. Al hacer un test de nivelación se genera, como en cualquier proceso de evaluación, una tensión emocional y cognitiva entre el centro que necesita evaluar al candidato, y el candidato al que no le hace ninguna gracia ser evaluado de un idioma en el que se siente más o menos inseguro, y del que tiene, además, una imagen concreta que puede discrepar de la del centro donde quiere realizar el curso.

PROPUESTA DE PRUEBA DE CLASIFICACIÓN DEL INSTITUTO CERVANTES DE BERLÍN

Paso ahora a presentar y explicar la prueba de clasificación del nivel A1 desarrollada para el centro del Instituto Cervantes de Berlín que está actualmente en uso. En las siguientes tablas se presentan los 12 ítems correspondientes al nivel A1 junto a los contenidos evaluados. Se han seguido, además de los criterios expuestos más arriba, estos otros:

1.º Cada ítem propone un tipo de tarea diferente, representada en la instrucción. El objetivo es que el alumno tenga que prestar atención cada vez a lo que tiene que hacer y no ritualice la realización del test. Las instrucciones están en el idioma del alumno. Hay cuatro tareas diferentes: rellenar huecos, interpretar fragmentos, buscar equivalentes e identificar errores.

2.º Los microdiálogos tienen palabras marcadas en rojo que corresponden al contenido evaluado, y hacia el que dirige la atención el alumno.

3.º El conjunto de ítems corresponde a experiencias e informaciones reconocibles por el alumno.

4.º La sucesión de microdiálogos presentes en los ítems conforman una situación social concreta. Hay personas concretas con nombres concretos que representan situaciones vividas por el candidato. De esta manera, por un lado, se reduce el cansancio emocional, y por otro, se desvía la atención hacia el contenido del test alejándose de la realización de la tarea.

Ver esta prueba de clasificación.

Test N.º 1		Se evalúna el nivel A1. Cursos: A1.1 y A1.2	
N.º de ítem	**Contenidos Manual: Bitácora 1 (Ed. Difusión)**	**Unidad del libro**	**Tipo de ítem**

01	Vocabulario: saludos y despedidas	Unidad 00	Selección de la interpretación correcta
INSTRUCCIÓN: Selecciona la interpretación correcta del fragmento en rojo.			
Diálogo en una escuela de idiomas. —**Buenos días,** ¿eres Gloria Ruiz? —Sí, **buenos días, ¿qué tal?** Hoy es mi primer día de clase.			
a) La chica llega b) La chica se va c) No se sabe Opción correcta: a)			

02	Gramática: diferencias *ser/tener/estar*	Unidad 01	Completar frases con huecos
INSTRUCCIÓN: Completa el texto con una de las opciones.			
El primer día de clase, la profesora se presenta. —Hola. Me llamo Gloria y _____vuestra profesora. Vivo en Berlín, tengo un hijo de 16 años que se llama Luis y _____ un perro. Su perro se llama Darko. Soy profesora de español y de yoga. Mi familia _____ en Almería. ¿Y vosotros? ¿Podéis presentaros?			
a) es/tiene/estoy b) soy/tiene/está c) estoy/tengo/es Opción correcta: b)			

03	Gramática: concordancia de género y número	Unidad 03	Completar frases con huecos
INSTRUCCIÓN: Completa el texto con una de las opciones.			
—Tú, por ejemplo, ¿cómo te llamas? —Hola, Gloria. Yo me llamo Gustav y estudio español porque mi mujer es_____, pero su papá es _____. Tengo un hermano en Roma, su hija habla _____ y _____. Queremos ir a vivir a Oaxaca el próximo año.			
a) mexicano/japonés/italiano/alemán b) mexicano/japoneses/italiana/alemana c) mexicano/japoneses/italiana/alemana Opción correcta: c)			

04	Vocabulario: miembros de la familia	Unidad 03	Selección de la interpretación correcta
INSTRUCCIÓN: Selecciona la interpretación correcta de los fragmentos en rojo.			
—Vale, gracias, Gustav. ¿Quién quiere continuar? —Bueno, ahora yo. Me llamo Ivonne, soy francesa pero vivo en Berlín. Trabajo en la embajada de Francia. **Mi marido** se llama Carlo y es italiano, pero **el padre de mi marido** es tailandés y **la madre** es escocesa como **mi madre y su hermana.** En casa hablamos muchos idiomas.			
a) La suegra de Ivonne habla inglés como lengua materna. b) La tía de Ivonne habla italiano como lengua materna. c) El cuñado de Carlo habla tailandés como lengua materna. Opción correcta: a)			

05	Gramática: uso del presente y marcadores de frecuencia	Unidad 04	Selección de la interpretación correcta
INSTRUCCIÓN: Selecciona la interpretación correcta de los fragmentos en rojo.			
—Hola, tú eres Silke, ¿no? ¿Quieres presentarte? —No, yo me llamo Sandra, Silke es ella. Es que somos gemelas. Bueno, soy de Hamburgo y soy médica. Me gusta mucho el deporte, **practico** casi **todos los fines de semana,** juego al fútbol y al rugby. **Ahora empiezo** un nuevo deporte, gimnasia rítmica. Estudio español porque **viajo cada año** a Barcelona para competir.			
a) Silke va a Barcelona una vez al año. b) Silke de lunes a viernes no hace deporte. c) Silke quiere ser gimnasta profesional. Opción correcta: b)			

06	Vocabulario: verbos de acciones cotidianas	Unidad 04	Selección de una forma equivalente
INSTRUCCIÓN: Selecciona una opción equivalente a los fragmentos en rojo.			
—Gracias, Silke. Entonces, tú eres Sara y eres la hermana de Silke, ¿no? —Sí, bueno, somos gemelas pero no nos parecemos, a mí no me gusta nada el deporte y no soy de Hamburgo. Nací en Budapest. **Empiezo** este curso porque me gusta **conocer** gente que habla español y **quedar** con ellos para tomar cervezas. **Me acuesto** siempre muy tarde, por eso me levanto siempre muy tarde. No cocino y nunca **hablo** de política.			
a) Comienzo / contactar con / encontrarme / Me voy a dormir / converso b) Hago / encontrarme con / ver / Me duermo / digo c) Visito / llamar / encontrar / Duermo / discuto Opción correcta: a)			

07	Gramática: preposiciones en/a/ para/con	Unidad 04	Selección de una palabra incorrecta
INSTRUCCIÓN: Selecciona la palabra incorrecta.			
—¿Tú eres Robert? ¿Eres alemán? —Sí, me llamo Robert, soy el novio de Antoine. No nos sentamos juntos **porque** no hablar juntos, así aprendemos más. Soy austríaco pero nací en Vladivostok y hablo ruso. Me gusta mucho ir **con** bicicleta y aprendo español **para** ir **de** Perú **a** Argentina, bueno, si Antoine quiere, claro.			
a) porque b) en c) para d) de e) a f)con Opción correcta: a)			

08	Vocabulario: actividades de ocio	Unidad 05	Selección de la interpretación correcta
INSTRUCCIÓN: Selecciona la interpretación correcta de los fragmentos en rojo.			
—Muchas gracias. Tú debes de ser Antoine. ¿Nos hablas un poco de lo que haces en tu tiempo libre? —Bueno, la idea de Robert es un poco loca, ¿no? Vale, el tiempo libre. Voy mucho a restaurantes, soy de París y allí es normal. Al **teatro** no voy con frecuencia, y **al cine** tampoco, creo que en Berlín solo hay **películas** comerciales. **Salgo** mucho con amigos, y otra cosa que me encanta es **comprar** en **mercadillos**. —¡Y el zoológico! A Antoine le encantan los **parques,** en especial el **zoológico,** ¿no? —Sí, Robert, tienes razón.			
a) La vida de Antoine en Berlín es como en París. b) A Antoine le encanta salir de la ciudad. c) Antoine pocas veces se queda en casa. Opción correcta: c)			

09	Gramática: pronombres personales (sujeto, OD / OI)	Unidad 05	Selección de una palabra incorrecta
INSTRUCCIÓN: Selecciona la palabra incorrecta.			
—Gracias, Antoine. ¿Y tú? ¿Eres Fredericke? ¿Quieres presentarte tú también? —Sí, claro. Me llamo Fredericke y soy berlinesa. Tengo una hija que se llama como **mí,** pero **la** llamamos Dolores porque no **le** gusta su nombre. ¡Ah! Y quiero decir**te**, Antoine, que en Berlín no solo hay películas comerciales. Hay muchos cines donde puedes ver películas interesantes.			
a) mí	b) la	c) le	Opción correcta: a)

10	Gramática: la comparación (vocabulario de la ciudad)		Selección de la interpretación correcta
INSTRUCCIÓN: Selecciona la interpretación correcta del fragmento en rojo.			
—¿Tú eres Olga? ¿Eres ruso, ucraniano...? —Soy de San Petersburgo. ¿Me presento? Vale. Pues soy Vladimir y estudio español para vivir un rato en España, por ejemplo en Sevilla, o La Coruña o Bilbao, no sé aún cuál **prefiero.** Mi ciudad, San Petersburgo, es **más** bonita, tiene **más** canales y puentes y los parques son **muy** grandes, pero ahora vivo en Berlín y está bien. En el futuro quiero ir a España y luego a Latinoamérica. Quiero conocer el mundo. Más tarde, Asia, y luego África…			
a) Olga prefiere Berlín a Sevilla o Bilbao. b) A Olga no le gusta Berlín tanto como su ciudad. c) A Olga quiere viajar porque no le gusta Rusia. Opción correcta: b)			

11	Vocabulario: verbos de movimiento	Unidad 09	Completar frases con huecos.
INSTRUCCIÓN: Completa el texto con una de las opciones.			
—Vale, gracias, Vladimir. Y gracias a todos. Dime, Gustav, ¿usas mucho el transporte público? —Bueno, sí. Vivo en Spandau pero trabajo en Köpenick y paso dos horas en el metro: una hora para ____ otra hora para ____. El metro ____ por la Friedrichstr. ____ a las seis y ____ a las siete. Es mucho tiempo. O cambio de trabajo, o mi mujer se divorcia. ¡No nos vemos nunca!			
a) andar / retornar / va / voy / vengo b) ir / venir / viaja / subo / bajo c) ir / volver / pasa / salgo / llego Opción correcta: c)			

12	Gramática: uso del perfecto (marcadores temporales)	Unidad 09	Completar frases con huecos.
INSTRUCCIÓN: Completa el texto con una de las opciones.			
—¿Y tú, Ivonne? ¿Usas el transporte público con frecuencia? —Yo tengo suerte. ____ **voy** a trabajar a pie porque vivo muy cerca del trabajo. ____ **he tardado** un poco más porque ____ no **he dormido** bien y me he levantado tarde. ____ **necesito** quince minutos. Soy traductora para el parlamento alemán, vivo al lado de la Hauptbahhof, en unas casa nuevas. —¡Qué bien! Bueno, muchas gracias a todos por las presentaciones. Vamos a empezar el curso.			
a) Ahora / Esta semana / por la noche / Hoy b) Actualmente / Mañana / cada noche / Esta mañana c) Siempre / Frecuentemente / todas las noches / Raramente d) A menudo / Hoy / esta noche / Normalmente Opción correcta: d)			