

La Plataforma de exámenes multilingüe PLEVALEX¹: Resultados del diseño y perspectivas de investigación futura de la Plataforma de Exámenes Valenciana de Lenguas Extranjeras

Jesús GARCÍA LABORDA²

Universidad Politécnica de Valencia
jgarcial@idm.upv.es

Recibido: diciembre 2005
Aceptado: mayo 2006

RESUMEN

Este artículo describe cómo está diseñada la Plataforma de Exámenes Valenciana de Lenguas Extranjeras (PLEVALEX) para uso en redes y cómo los ítems de exámenes tradicionales han sido adaptados para esta plataforma informática. Adicionalmente, también justifica los tipos de tareas de exámenes incluidas en la plataforma. Tras una breve revisión de un análisis de necesidades que justificó el desarrollo de esta herramienta, el artículo muestra las tres partes principales de los exámenes distribuidos con PLEVALEX: selección múltiple, redacción escrita y oral. Finalmente, el artículo incluye los temas de investigaciones y desarrollo futuros del proyecto de investigación. Como conclusión, se sugiere que PLEVALEX es una herramienta importante que hará más asequible los exámenes para grandes cantidades de alumnos y de gran efecto social (*backwash*).

Palabras clave: exámenes, redes informáticas, plataformas informáticas, investigación, necesidades.

The multilingual test platform PLEVALEX: results and future development in research

ABSTRACT

This paper addresses how the web based Plataforma de Exámenes Valenciana de Lenguas Extranjeras (PLEVALEX) is designed and how traditional test items have been adapted to the computer-based

¹ La plataforma de examen se llamó HIEO-HIELE originalmente pero, conforme avanzó el proyecto, su nombre cambió a PLEVALEX e integró los dos programas bajo una misma denominación. HIELE es parte de la plataforma de examen PLEVALEX que incluye las herramientas HIELE e HIEO. La Herramienta Informática de Evaluación Ora (HIEO) no se podría haber elaborado de no ser por la subvención de la Generalitat Valenciana (Proyecto GV043/436) de más de 28.000? HIELE (su parte escrita) se ha podido desarrollar gracias al patrocinio del proyecto de la Universidad Politécnica de Valencia (20040941) de 12000?

² Desde estas líneas queremos agradecer su contribución a las siguientes compañeras participantes en el proyecto: Marie France Collart, Josefa Contreras Fernández, Daniela Teresa Gil Salom, Mercedes López Santiago, Amelia Maria Amelia Plasencia Abasolo, Josep Angel Mas Castells, Cristina Navarro Lobulais, Inmaculada Sanz Alava, Coral López Mateo y Lola Cuenca Jaramillo de la Universidad Politécnica de Valencia y a los doctores Luis G. Bejarano (Valdosta State University, EE.UU.), y Pilar García Caricedo (UCM) por su apoyo inicial y, sobre todo, a Emilia V. Enríquez Carrasco.

platform. Additionally, it also justifies the types of testing tasks included in the platform. After a brief review of the needs analysis that justified the computer-based platform development, the paper shows the three main sections in the test prepared through PLEVALEX: multiple choice, written composition and oral. Finally, the article includes the topics of further research and development in the research project. As a conclusion, it is suggested that PLEVALEX is a powerful tool that will facilitate high stakes testing (and backwash).

Key words: testing, computer nets, computer platforms, research, needs.

La plateforme d'examens multilingue PLEVALEX: Resultats de ce logiciel et perspectives de recherche de la plateforme d'examens de Valence pour les langues étrangères

RÉSUMÉ

Cet article décrit comment on a conçu la Plate-forme d'Examens de Valence de Langues Étrangères (PLEVALEX) pour son utilisation dans des réseaux et comment les items d'examens traditionnels ont été adaptés à cette plate-forme informatique. Additionnellement, on y justifie aussi les types de tâches d'examens incluses dans la plate-forme. Après une brève révision d'une analyse de besoins qui a justifié le développement de cet outil, l'article montre les trois parties principales des examens distribués avec PLEVALEX: sélection multiple, rédaction écrite et orale. Finalement, l'article inclut les sujets de recherche et les développement futurs du projet de recherche. Comme conclusion, on suggère que PLEVALEX est un important outil qui rendra plus abordables les examens pour de grandes quantités d'élèves et de grand effet social (backwash).

Mots clé: examens, réseaux informatisés, plateformes informatiques, recherche, besoins.

SUMARIO: 1. Introducción; 2. Análisis de necesidades; 3. PLEVALEX: Una plataforma para evaluar todas las destrezas; 4. Futura agenda de investigación; 5. Conclusiones; 6. Referencias Bibliográficas

1. INTRODUCCIÓN

Hace ya dos años García Laborda (2004) anunciaba el inicio de los trabajos en la primera Plataforma de evaluación oral y escrita en Valencia (y, probablemente, en España) que permitiese realizar una evaluación completa de los estudiantes internacionales tanto españoles como extranjeros en movimiento pertenecientes a la Universidad Politécnica de Valencia. La plataforma, llamada PLEVALEX, está ya en su fase avanzada de prueba en sus fases de redacción y selección múltiple así como el fichero de gestión de ítems, alumnos y exámenes y la sección oral que mencionamos brevemente estará a mediados de 2006 (García Laborda, 2006a). En este trabajo se presenta el diseño especificando sus posibilidades y agenda de investigación futura.

La plataforma PLEVALEX puede considerarse como un avance significativo en España en el diseño y desarrollo de plataformas multilingües por varias razones: primero, es una pionera en España de un mercado que está llamado a transformar el campo de los exámenes; y segundo, porque aportará valiosa información sobre la posibilidad futura de realizar exámenes asistidos por ordenador como parte de otros mucho más exigentes y que, hasta la fecha, no han podido incluir secciones orales

por lo difícil de su organización y por su coste elevado y propiciando indicadores, no estudiados en la actualidad, de preparación de exámenes. PLEVALEX aparece en un momento muy especial porque son varias las empresas de importancia internacional que está experimentando plataformas similares en varias partes del mundo. Probablemente la más desarrollada en este momento sea la ya operativa TOEFL (desde 2004) pero hay otras en experimentación como la última versión de IELTS y otras en varias partes del mundo (como el *LARC project*, en California). Antes de adentrarnos en la descripción de la herramienta diremos que, en contra de lo que pueda parecer, en este caso la investigación española no ha seguido a la americana ya que los primeros estudios de PLEVALEX son anteriores a la presentación de la versión nueva de TOEFL y fue precisamente la revista *Didáctica* la que tuvo a bien publicar el primer artículo sobre esta nueva plataforma. De cualquier manera, este artículo presenta el diseño final de la misma con la ilusión de que pronto haya más en nuestro país y, asimismo, se mencionan ciertas cuestiones asociadas al uso de las tecnologías de la información que, probablemente, adquirirán una repercusión que, hasta ahora, se había limitado en el campo los exámenes oficiales.

2. ANÁLISIS DE NECESIDADES

En 2004, García Laborda y Bejarano (2005) analizaron las necesidades de los estudiantes internacionales de la Universidad Politécnica de Valencia y Valdosta State University (Georgia, EEUU) y observaron ciertas características comunes que requerían el desarrollo de una herramienta informática de exámenes. Entre estas necesidades estaban:

- Diagnosticar inicialmente a los estudiantes internacionales en movimiento incluso antes de su llegada a la universidad anfitriona,
- Evaluar las deficiencias no detectadas en sus cursos reglados de lenguas extranjeras,
- Describir los conocimientos y actitudes ante el comienzo de cursos en un país extranjero en una segunda lengua,
- Detectar los estudiantes con una motivación instrumental en su aprendizaje de L2,
- Diseñar un curso inicial para adaptarse a su nuevo entorno lingüístico (y no tanto cultural), y
- Programar cursos a la medida que favorezcan una progresión rápida para adaptarse a las asignaturas curriculares en la lengua del país destino.

Como en otros estudios previos, esto requería el desarrollo de herramientas de evaluación rápida, accesible a larga distancia y eficiente. Para ello, García Laborda (2004) sugería el desarrollo de la herramienta informática que presenta la siguiente sección de este artículo. Esta herramienta debería tener ciertos efectos sociales asociados (García Laborda y Enríquez Carrasco, 2005) u otros efectos denominados

efectos de los exámenes de criterio y que se pueden asociar a exámenes de acceso a instituciones públicas como la Selectividad (García Laborda, 2006b). Estos efectos estudian cuál es la influencia de la realización positiva de ciertos exámenes que permiten, en el caso de la Selectividad, el acceso a unos estudios deseados frente a otros estudiantes competidores basándose en los resultados globales obtenidos (lo que en España se ha venido denominando coloquialmente «nota de corte»). En el caso de la plataforma PLEVALEX pensamos que tiene un futuro inmediato de cara a la realización de exámenes de acceso de estudiantes internacionales que agrupen y cubran las necesidades iniciales planteadas más arriba.

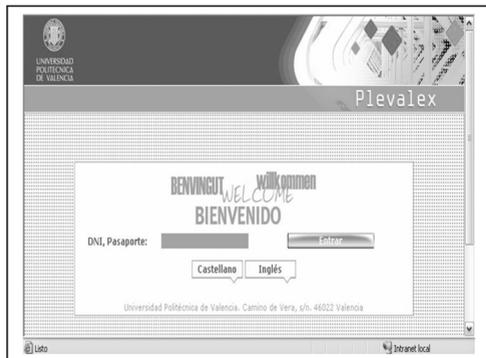
3. PLEVALEX: UNA PLATAFORMA PARA EVALUAR TODAS LAS DESTREZAS

Si bien es verdad que los ordenadores han resultado muy beneficiosos en el campo de realización de exámenes desde el principio de los años 80 (Chalhoub-Deville, 2001), no es menos cierto que existían ciertas limitaciones como la necesidad de realizar cualquier examen oral por videoconferencia (Luecht, 2001; Roever, 2001), las actitudes de interacción oral y global con ordenadores (Kenyon y Malabonga, 2001) asegurar la identidad del examinando y evitar el plagio (Roever, 2001), diseño de interfaces (Brown, 1997), velocidad de transferencia (Holland, 1995; Warschauer, 1996) o diferencias de actuación entre los tests escritos y los basados en ordenador (Sawaki, 2001)). Como quiera que este es ya un tema recurrente y visto por estos y otros autores, en esta publicación nos centraremos solamente en decir que, sin duda alguna, la evaluación de destrezas orales en exámenes asistidos por ordenador había sido un problema irresoluble hasta el principio de este siglo (Norris, 2001). En esa época comenzaron las pruebas con la nueva versión del *Test of English as a Foreign Language* (TOEFL) (Fraser *et al*, 1997; Lee, 2001; Lee, 2002; Cumming *et al*, 2004a y b). Hasta la fecha de hoy, no tenemos noticia alguna en España de otra plataforma similar aunque no dudamos que surjan pronto. A continuación mostramos cómo es la plataforma, sus secciones y algunos interfaces de la misma.

3.1. Presentación

Al acceder a la plataforma observaremos la clave de acceso seguido de la identificación del alumno y el examen que está haciendo así como el tiempo que éste ha de durar y una muestra del mismo.

Gráficos 1 y 2. Interfaces de acceso



A priori, una cosa que llama la atención ha sido nuestra intención de que estos interfaces fueran lo más sencillos posible (Fulcher, 2003). Ya que demasiadas imágenes o colores tienden a distraer la atención del examinando. Además, debemos tener en cuenta que este examinando será probablemente extranjero y que su natural ansiedad debe ser correspondida con facilidad de identificación de los elementos en pantalla. Tras estas pantallas el alumno verá las distintas secciones del examen.

3.2. Secciones del examen

En cuanto al diseño de las secciones, el Grupo de Investigación en Lenguas para Fines Específicos (GILFE) debatió casi desde el principio el tipo de preguntas. Así se plantearon los siguientes requisitos sobre las características de las tareas de la nueva herramienta:

- Debería ser capaz de observar el cumplimiento de los criterios exigidos en cada nivel del Marco de Referencia Común Europeo,
- Incluiría preguntas cortas y largas,
- Contaría con pruebas orales y escritas,
- Tendría que incluir unas comunicativas y otras más formales, y
- Tendría registros informales y académicos.

De aquí, se decidió hacer tres secciones:

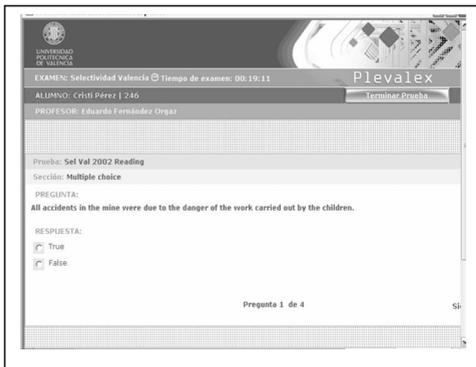
- Sección de selección múltiple,
- Sección de redacción con preguntas más largas y otras más breves, y
- Sección oral con vídeos informativos formales e informales o bien otros soportes digitales como fotos o grabaciones de audio (aún en primera fase de experimentación).

La inclusión de tareas de selección múltiple se debe a la posibilidad de transformar prácticamente cualquier tipo de pregunta en selección múltiple. Esta opinión es también compartida por Herrera Soler (2005) quien, basándose en el hecho de que los exámenes miden criterios de valoración sobre los conocimientos de los alumnos (no los conocimientos mismos), afirma que las tareas de selección múltiple son necesarias en exámenes normativos como la Selectividad. Las otras secciones se decidieron ajustándose a la necesidad comunicativa oral y escrita.

3.2.1. La sección de selección múltiple

Para esta sección también convenía plantearse si sería mejor que todas las preguntas apareciesen seguidas (lo cual sería especialmente útil) en casos de lecturas largas o bien, de cara a aspectos relacionados con el diseño web, preguntas presentadas de una en una para una mayor claridad y navegación. Se optó por esta segunda opción (gráfico 3) a pesar de que, en textos largos, el estudiante quizás tendría que volver de las preguntas al enunciado inicial, por cuestión de diseño aunque parece necesario que cuando esta herramienta evolucione se pueda ver en un cuadro tal y como veremos que sucede con aquellas preguntas que incluyen archivos gráficos (Gráfico 4).

Gráficos 3 y 4. Sección de selección múltiple con y sin vídeo.



En cuanto a la sección con vídeo se observó que el problema mayor provenía del tamaño de los vídeos. Las opciones era si crear un marco flotante de mayor tamaño y trabajar en la pantalla de fondo o crear una imagen menor (como la del gráfico 4) que permitiese trabajar con todos los datos a la vista. Sin duda, la segunda prevaleció para simplificar los formatos de cara a los alumnos.

Para la transición entre secciones, se utilizó una pantalla que permitiese ver las partes del examen realizadas y también que indicase cuando el alumno podía aban-

donar. Como se ve en el diagrama 5, todo ello queda recogido en una sola pantalla y es bastante claro. Además, cada botón tiene un color claramente distintivo que permiten que incluso los alumnos con más dificultades encuentren una pantalla fácil de seguir y en la que se reconocen mecanismos como cierre, comienzo, tiempos y otras necesidades de manera muy sencilla.

Gráfico 5. Pantalla de transición entre secciones



3.2.2. La sección de redacción

Quizás sea esta la que tiene un formato más tradicional ya que, como veremos en esta sección, lo verdaderamente importante es la transmisión de datos y la corrección. Así, en el gráfico 6 observamos cómo quedaría la misma.

Gráfico 6. Sección escrita en PLEVALEX



En el interface podemos observar prácticamente el mismo diseño que en la sección anterior, cosa ciertamente natural, pero, a diferencia de otros exámenes informatizados, esta sección tiene la posibilidad de incorporar elementos audiovisuales que creemos que mejoran la actuación de los estudiantes (Ghinter, 2001). Por tanto

podríamos incluir respuestas a sonidos, descripciones de fotografías o reacciones ante vídeos. De hecho, en uno de los estudios en curso se trata de ver las distintas reacciones ante estos tres tipos de elementos basándonos en la cuestión de la relación entre la ansiedad de la realización de un examen y la dinámica que toma cada uno de estos apoyos audiovisuales. Asimismo, también se podría incluir un texto largo y pedir distintos tipos de tareas como tareas basadas en input (por ejemplo, le pediríamos al estudiante que tomase notas durante una miniconferencia y que las integrase en su respuesta o resúmenes de pruebas de comprensión auditiva) o tareas de respuesta abierta (como expresar opiniones). En cuanto a los géneros serían tan variados como los reales: artículos de opinión, historias, completar folletos y hojas informativas, cartas, notas personales, ensayos y otros muchos géneros exigidos por los exámenes de Cambridge o el Cervantes.

Aunque en el futuro de esta plataforma se podrían encontrar la realización de estudios de facilidad y rapidez lectora comparada entre lápiz y papel, posibilidades de corrección automática, etc. Por ahora, se han dejado de lado publicaciones adicionales que, sin duda PLEVALEX puede promover. También existe la posibilidad de ver si esta herramienta sería especialmente propicia para desarrollar una corrección automática. Sin embargo, en la actualidad, la corrección se hace manualmente. Para ello, cuando el estudiante termina el examen este es enviado al corrector. Su archivo escrito entra en la aplicación de correctores-administradores (*backoffice*) y corrige el examen a partir de siete tipos de criterios genéricos (gráfico 7): concordancia, mensaje, ortografía, gramática, vocabulario, coherencia interna y sintaxis. Estos códigos no son sino un resumen de una taxonomía inicial de unos 40 tipos de errores que se fueron reduciendo progresivamente.

Gráfico 7. Backoffice de la sección escrita



Aunque, en este caso, el estudio inicial había planteado un sistema de baremación automática con una validación del profesor, finalmente se optó por un sistema

que diese un informe gráfico para el alumno (disponible a finales de 2006) pero en el cual la baremación la hace el profesor de acuerdo a unos criterios especificados en el constructo del examen.

4. FUTURA AGENDA DE INVESTIGACIÓN

Una herramienta del carácter e interés que tiene esta no deja de ofrecer un amplio abanico de posibilidades de futura investigación. De cualquier manera, hasta el momento, PLEVALEX o sus fases han visto como he hacían un total de 12 estudios. Algunos de ellos han sido mencionados convenientemente en este artículo pero aún existen algunas líneas de investigación abiertas hacia el futuro³:

- Diseño: ¿Cómo mejorar la herramienta informática? ¿Cómo transferir y almacenar óptimamente los datos obtenidos de los exámenes? Creación de bancos de ítems.
- Constructo: ¿Cómo aumentar la validez de los exámenes realizados con PLEVALEX? Valor diagnóstico del mismo, adaptación a la normativa del espacio superior de educación europeo y adaptabilidad a exámenes oficiales de nivel, grados de confianza del examen, análisis y aplicación de la teoría general de tests y teoría de respuesta al ítem.
- Uso y actuación de los estudiantes: condiciones de realización y mejora en los exámenes, impacto social, estrategias de examen (lápiz y papel o informáticas).

5. CONCLUSIONES

La comunidad investigadora que trabaja en la aplicación de nuevas tecnologías en exámenes vive gozosa los últimos avances en el diseño de plataformas informáticas que puedan facilitar la realización de exámenes orales y escritos asistidos por ordenador. El trabajo que acabamos de presentar es posiblemente el primero en España diseñado para exámenes distribuidos a un gran número de alumnos y con importantes efectos sociales (*high stakes exams*). El proyecto, realizado en la Universidad Politécnica de Valencia por un equipo de 22 investigadores de distintas especialidades, ha tenido cierta relevancia en el campo profesional en nuestro país. La mayoría de los estudios realizados tienen un carácter prospectivo pero no dejan de mostrar grandes avances en áreas casi inexploradas en España donde, aparentemente, escasean los investigadores en el área de la evaluación (especialmente la rela-

³ Desde estas páginas aprovechamos para invitar a investigadores interesados en el proyecto a unirse a un equipo de 4 grupos de investigación y un total de 23 investigadores de, al menos, 6 especialidades distintas a unirse. Para ello, simplemente es necesario que contacten con nosotros en el correo electrónico jgarcial@upvnet.upv.es

cionada con exámenes de gran efecto social). Este trabajo ha presentado la plataforma PLEVALEX y sus fases de selección múltiple, redacción y oral y algunas de las alternativas y motivaciones que llevaron a su diseño actual. También ha mostrado los interfaces y tipos de preguntas que se pueden ofrecer en la plataforma. Finalmente se ha añadido los focos de interés de futuros estudios e investigaciones.

Llegados a este punto, quizás los estudios más interesantes por su escasez en España son sobre la creación de bancos de ítems para exámenes, efectos de los exámenes en los estudiantes y modernización de redes para la ampliación del uso de plataformas similares en el territorio nacional. Para concluir, simplemente nos queda mostrar nuestra más profunda confianza en que esta plataforma será de gran importancia para exámenes internos de la universidad como externos en aquellos que se necesite baremar o diagnosticar a un gran número de alumnos simultáneamente. En este sentido, este artículo no es sino una primera aproximación a un futuro al alcance de nuestros dedos en la que las plataformas informáticas están llamadas a transformar los parámetros de diseño de exámenes.

6. REFERENCIAS BIBLIOGRÁFICAS

- CHALHOUB-DEVILLE, M. (2001): «Language Testing and Technology: Past and Future». en *Language Learning & Technology*; 5/2: 95-98.
- CUMMING, A., GRANT, L., MULCAHY-ERNT, P. y D. E. POWERS (2004): «A Teacher-Verification Study of Speaking and Writing Prototype Tasks for a New TOEFL» en *Language Testing*; 21/1: 107-145.
- FRASE, L. T., GONG, B., HANSEN, E. G., KAPLAN, R., KATZ, I. R. y K. SINGLEY (1997): *Technologies for Language Testing*; TOEFL Monograph No. MS-11.
- FULCHER, G. (2003): «Interface design in computer-based language testing» en *Language Testing*, 20/4: 384-408.
- GARCÍA LABORDA, J. (2004): «HIEO: Investigación y Desarrollo de una Herramienta Informática de Evaluación Oral multilingüe» en *Didáctica*, 16: 77-88.
- GARCÍA LABORDA, J. (2006): «PLEVALEX: A new platform for Oral Testing in Spanish» en *Eurocall Review*, 9 (in press).
- GARCÍA LABORDA, J y L. G. BEJARANO (2005): «Análisis de la necesidad de creación de páginas web para la evaluación y baremación de estudiantes internacionales: Una experiencia internacional», en M. L. CARRIÓ PASTOR, «*Perspectivas interdisciplinarias de la Lingüística Aplicada (Tomo II)*» Valencia: U.P.V.: 399-404.
- GINTHER, A. (2001): *Effects of the Presence and Absence of Visuals on Performance on TOEFL CBT Listening-Comprehensive Stimuli*. TOEFL Research Report No. RR-66.
- KENYON, D. M. y V. MALABONGA (2001): «Comparing Examinee Attitudes Toward Computer-Assisted and Other Oral Proficiency Assessments» en *Language Learning & Technology*; 5/2: 60-83.
- LEE, Y.-W. (2001): *The Essay Scoring and Scorer Reliability in TOEFL CBT*. East Lansing, MI: National Center for Research on Teacher Learning. (descargado de Internet febrero 24, 2006 procedente del Servicio de Reproducción ERIC, ERIC No. ED455253).
- LEE, Y.-W., KANTOR, R. y P. MOLLAUN (2002): *Score Dependability of the Writing and Speaking Sections of New TOEFL*. Princeton, NJ: Educational Testing Service.

- LUECHT, R. M. (2001): *Challenges of Web-Based Assessment*. East Lansing, MI: National Center for Research on Teacher Learning. East Lansing, MI: National Center for Research on Teacher Learning (descargado de Internet febrero 24, 2006 procedente del Servicio de Reproducción ERIC, ERIC No. ED458294).
- SAWAKI, Y. (2001): «Comparability of Conventional and Computerized Tests of Reading in a Second Language» *Language Learning & Technology*; 5/2: 38-59.
- SUNG, Y.-T., CHANG, K.-E., CHIOU, S.-K. y H.-T. HOU (2005): «The Design and Application of a Web-Based Self- And Peer-Assessment System» *Computers and Education*; 45/2: 187-202.