LAS PRUEBAS DE NIVEL ONLINE MEDIANTE *C-TESTS* EN ELE

Miguel Monreal Azcárate
Universidad de Navarra

RESUMEN: Los *C-Tests* son un tipo de pruebas de idioma relativamente comunes en la enseñanza-aprendizaje de idiomas. Desde su aparición en los años 80 se han ido convirtiendo en un modelo principal de prueba de clasificación en numerosas universidades y centros superiores de idiomas de países como Alemania. Sin embargo, su empleo en ELE es prácticamente inexistente en nuestro país. En esta comunicación mostraré un ejemplo de una prueba de este tipo, y analizaré sus resultados en el centro del que procedo durante el presente curso académico, cotejándolos con los de una prueba «convencional» con el objetivo de cuestionar su fiabilidad y validez.

ORIGEN DE ESTAS PRUEBAS: LOS *CLOZE TESTS*

Los *C-Tests* son la evolución de los *cloze tests*, mucho más comunes en la enseñanza de idiomas, y cuyo origen se sitúa en 1953, cuando Wilson Taylor presentó dicha técnica en su artículo «*Cloze procedure*: a new tool for measuring readability».

Taylor se basó en el llamado *principio de cierre* (*closure*) de la psicología de la Gestalt, que sostiene que el ser humano es capaz de *cerrar* o completar determinadas estructuras o patrones que, debido a la experiencia, le resulten familiares, como ocurre en las imágenes siguientes.



Desde el punto de vista lingüístico, el principio subyacente del procedimiento *cloze* es el mismo que el de otros tipos comunes de evaluación (como los dictados, o los *tests de ruido*): el de la *redundancia reducida*. Es este un concepto acuñado por Spolsky, que defiende que el lenguaje, de por sí, incluye diversos elementos redundantes. El hablante competente, gracias a sus conocimientos del sistema lingüístico, el contexto, etc., será capaz de descifrar aquellos mensajes donde estos elementos redundantes hayan sido reducidos.

Taylor, basándose en estas teorías, creó los *cloze tests,* suprimiendo determinadas palabras de los textos de forma sistemática. De este modo, tomando como referencia el índice de éxito del lector a la hora de «restaurar» el texto, podía medirse la dificultad de lectura del mismo. Aunque inicialmente Taylor propuso este procedimiento como una herramienta para medir la dificultad de lectura (*readability*) de los textos, poco después empezaron a utilizarse los *cloze tests* para evaluar la habilidad lectora de los hablantes nativos.

Más adelante, el procedimiento *cloze* comenzó a aplicarse a la enseñanza de idiomas, siendo hasta nuestros días muy empleado, en sus diversas variantes (*maze, cloze restringido, cloze progresivo, cloze acumulativo*...), en un buen número de pruebas oficiales de idiomas (como los DELE, o las de ESOL de Cambridge) debido a su fiabilidad y objetividad.

Pese a su popularidad, el uso de los *cloze tests* en la enseñanza-aprendizaje de idiomas también ha sido objeto de numerosas críticas, en especial tras la llegada de los enfoques comunicativos.

Alderson y Klein-Braley señalaron algunas de las desventajas que los *cloze tests* tradicionales presentan. Una de ellas es que pueden resultar muy difíciles, o imposibles de completar (incluso para hablantes nativos) si seguimos un criterio exacto o aleatorio a la hora de suprimir las palabras que el estudiante deberá restaurar. Y si seguimos, por lo contrario, un criterio selectivo, la preparación del test resultará costosa y poco auténtica.

Por otro lado, se requieren numerosos huecos (y, por lo tanto, textos muy extensos) para asegurar la fiabilidad de los resultados. Y habitualmente se emplea un solo texto, el cual no supondrá una muestra representativa del idioma en cuestión.

Otra de las críticas comunes hacia el procedimiento *cloze* es el hecho de que pruebas de este tipo son, para muchos, un indicador de determinados conocimientos de la lengua más que de su uso. Además, para algunos especialistas, no está claro que estas pruebas midan las habilidades lingüísticas y comunicativas del usuario, sino más bien otro tipo de destrezas (relacionadas con procesos de la habilidad lectora), algo que puede demostrar el hecho de que, para algunos hablantes nativos no habituados al procedimiento, estas pruebas puedan resultar, como se ha comentado ya, extremadamente difíciles.

LOS *C-TESTS*

En los años 80, Christine Klein-Braley y Ulrich Raatz presentaron los *C-Tests* (o *pruebas tipo C*) como una variante mejorada de los *cloze tests*, junto con unos criterios de calidad que aseguraban su validez. Al igual que en el caso de los *cloze tests*, se emplean textos auténticos; pero en esta nueva modalidad se escogen normalmente cuatro o cinco textos breves (o fragmentos de temática diversa, pero siempre *cerrados* o completos), que se ordenarán de menor a mayor dificultad.

El procedimiento (denominado *C-Principle*) es el siguiente: la primera y la última frase de cada texto se dejan intactas; a partir de la segunda frase se elimina la segunda mitad de las palabras pares hasta obtener 20 o 25 huecos por texto. Las palabras de una letra y los nombres propios se excluyen de la omisión. Para garantizar un resultado válido, el número total de huecos de la prueba deberá ser de unos 100 (es decir: cinco textos de 20 huecos cada uno, o cuatro textos de 25 huecos cada uno).

Véase un ejemplo:

LOS PAJAROS PERCIBEN MUCHOS MAS COLORES QUE EL HOMBRE

La naturaleza no sólo ha dotado a los pájaros de una exuberante gama de colores en su plumaje que no supera ningún otro grupo de vertebrados. Su desarr[ ] sentido d[ ] la vi[ ] les per[ ] distinguir mud[ ] más col[ ] de l[ ] que pod[ ] percibir l[ ] humanos, se[ ] asegura u[ ] estudio publ[ ] esta sem[ ] en l[ ] revista 'Behavioral Ecology'.

S[ ] retina e[ ] sensible a l[ ] gama ultrav[ ], por l[ ] que pue[ ] ver colores invisibles para los humanos.

Desde hace tiempo, los ornitólogos creen que los colores de las plumas tienen diversas funciones. Por ejemplo, les sirven para camuflarse, emitir señales a otros animales o en el cortejo para conseguir pareja.

www.elmundo.es

Como se puede observar, para restaurar adecuadamente los huecos es necesaria determinada competencia lectora, además de unos conocimientos léxicos, gramaticales, sintácticos y de construcción textual, y ciertas estrategias de deducción e interpolación. En menor grado, pueden requerirse también algunas habilidades sociolingüísticas y pragmáticas. «No es un simple ejercicio de llenar huecos en el que solo se necesitan habilidades lectoras mínimas, sino que está presente un análisis discursivo que implica el uso de la lengua como un todo integrado» (Linares i Zapater, 2009, pág. 54).

Las ventajas sobre los *cloze tests* tradicionales, según Klein-Braley (1997), son las siguientes:

- Se obtiene un número mayor de ítems mediante textos mucho más cortos y diversos, por lo que la temática es más variada; y no se favorece a quien, en el caso del texto único, pudiera estar familiarizado con el tema.
- La evaluación y corrección son más rápidas y objetivas (puesto que normalmente hay solo una respuesta posible).
- Son pruebas que los nativos pueden resolver con facilidad, pero que resultan imposibles de completar para quienes no tienen conocimientos del idioma.
- Los ítems recogen una muestra muy representativa de todas las clases de palabras.

Klein-Braley y Raatz mencionan diversas investigaciones previas que demuestran que los segmentos omitidos en un texto de este tipo se corresponden con una muestra al azar representativa de las diversas partes de la lengua. De esta forma, la prueba resultará auténtica y completa, ya que el candidato se enfrentará a una muestra real de lengua en la que los elementos que se van a evaluar no han sido seleccionados ni manipulados de antemano por el examinador.

Por este motivo, los defensores del *C-Test* sostienen que este es un excelente instrumento de medida de la competencia lingüística general que resultará muy útil como herramienta de selección y nivelación. Fuera de estos usos, su alcance es, sin embargo, limitado, ya que no resultan suficientes, por ejemplo, como pruebas de rendimiento; ni para determinar el nivel específico del MCER que posee el usuario.

Diversos estudios (Klein-Braley, Cantera Adrados y Khodadady, entre otros) han defendido la validez de los *C-Tests* como pruebas de nivel al comparar sus resultados con los obtenidos por el estudiante mediante otro tipo de pruebas *convencionales*. A lo largo de este trabajo, y tal como proponían los creadores de los *C-Tests*, procuraré aportar algo más de información a ese respecto mediante un estudio comparativo realizado en la Universidad de Navarra en agosto de 2011.

DESCRIPCIÓN DEL PROCEDIMIENTO

MÉTODO

Este estudio parte de unas pruebas mediante *C-Tests* diseñadas por mí y realizadas *online*, a modo de pruebas de clasificación, por los estudiantes de ELE de la Universidad de Navarra del curso académico 2011-2012 junto con otro tipo de pruebas.

El objetivo de dichas pruebas era conocer de antemano el nivel de los estudiantes para poder clasificarlos en las diversas asignaturas del Instituto de Lengua y Cultura Españolas (ILCE), responsable de la docencia de ELE en la Universidad de Navarra. Dicha docencia está separada en materias diversas, divididas en grupos según el nivel de los estudiantes; por ejemplo: *Español [Gramática] A1-C1, Cultura y sociedad hispánicas A / B, Expresión escrita en español A / B, Grandes figuras de la literatura en español A / B, Historia medieval y moderna de España A / B,* etc. Por ello, se requería un conocimiento bastante preciso de las diversas destrezas de los estudiantes.

Las pruebas fueron realizadas *online* por los estudiantes durante la segunda quincena de agosto (es decir: previamente al inicio del curso), y, en la mayoría de casos, desde sus países de origen, mediante la plataforma *Moodle*. Se trata de un sistema gratuito de gestión del aprendizaje (*Learning Management System*, LMS) o Entorno de Aprendizaje Virtual (*Virtual Learning Environment*, VLE) muy empleado en los diversos ámbitos de la educación (primaria, secundaria y superior) de todo el mundo. Según indica la propia web de *Moodle*, se emplea en 215 países y cuenta con más de 45 millones de usuarios por todo el mundo, siendo especialmente popular en EE.UU., España, Brasil, Reino Unido y Alemania.

Esta plataforma ofrece múltiples posibilidades (formación *en línea*, complemento a la formación presencial, autoaprendizaje...) y permite llevar a cabo diversos modelos pedagógicos (como el aprendizaje colaborativo).

PARTICIPANTES

Los tests fueron realizados por 80 estudiantes universitarios de perfil diverso

Los tests fueron realizados por 80 estudiantes universitarios de perfil diverso

Nacionalidades:

China / Taiwán: 32,5%
EE.UU.: 12,5%
Italia: 7,5 / %
Francia: 6,25%
Reino Unido: 5%
Noruega: 5%
Canadá: 5%
Japón: 3,75%
Irlanda: 3,75%
República Checa: 2,5%
Polonia: 2,5%
Pakistán: 2,5%
Dinamarca: 2,5%
Otras nacionalidades: 8,75%

Lengua materna:

Chino mandarín / taiwanés: 32,5%
Inglés: 26,25%
Italiano: 7.5%
Francés: 6.25%
Noruego: 6%
Japonés: 3,75%
Otras (checo, polaco, urdu, danés): 15,75%

Programa o plan de estudios.

Alumnos del ILCE (*independientes*): 30%.
Programas de intercambio: 30%
Erasmus: 20%
Bridge Year (*Año puente*): 10%
Filosofía y Letras: 10%

Mujeres: 71,25%. Hombres: 28,75%

Edades: 17-25 años.

PROCEDIMIENTO

La prueba de nivel consistió en tres partes separadas (cuyos resultados se analizarán y contrastarán posteriormente): un test de opción múltiple (*multiple choice*), un *C-Test*; y la redacción de dos pequeñas composiciones.

El test de opción múltiple contenía 70 ítems, consistentes en preguntas pertenecientes a niveles léxicos y gramaticales diversos.

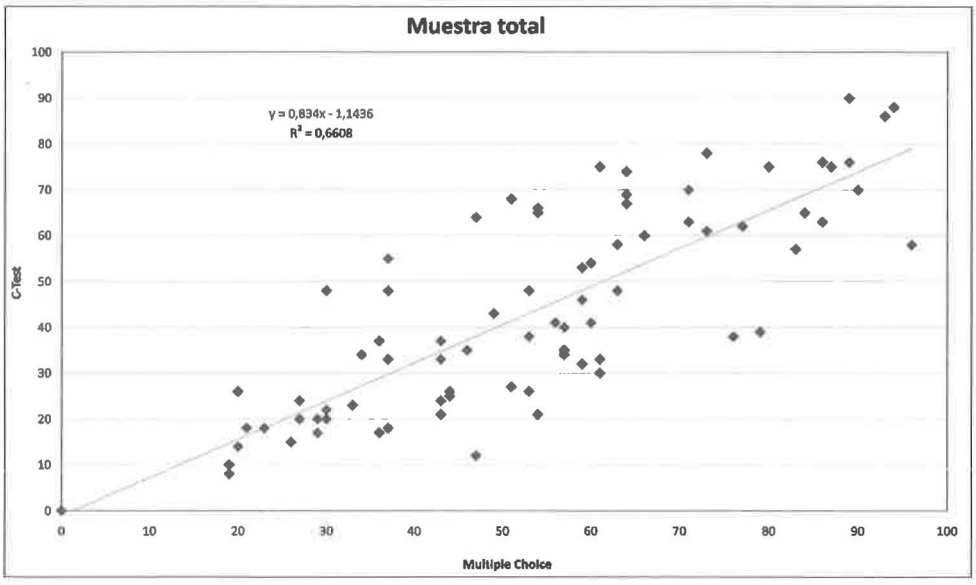
El *C-Test* consistía en cinco textos variados (periodísticos y narrativos) con un total de cien huecos.

La parte de redacción requería dos pequeñas composiciones (50 palabras cada una) de carácter descriptivo («*¿Cómo es tu mejor amigo? Descríbelo*») y narrativo («*¿Cuál es el recuerdo más feliz de tu infancia?*»)

Adicionalmente, los estudiantes que deseaban matricularse en las clases de conversación que ofrece el Instituto realizaron también una prueba oral presencial, a modo de pequeña entrevista. Aunque inicialmente dicha prueba no iba a formar parte del presente estudio (por no haber sido realizada por todos los estudiantes que completaron las pruebas escritas), algunos de sus resultados sí que se cotejaron con los de las pruebas escritas a la hora de elaborar conclusiones.

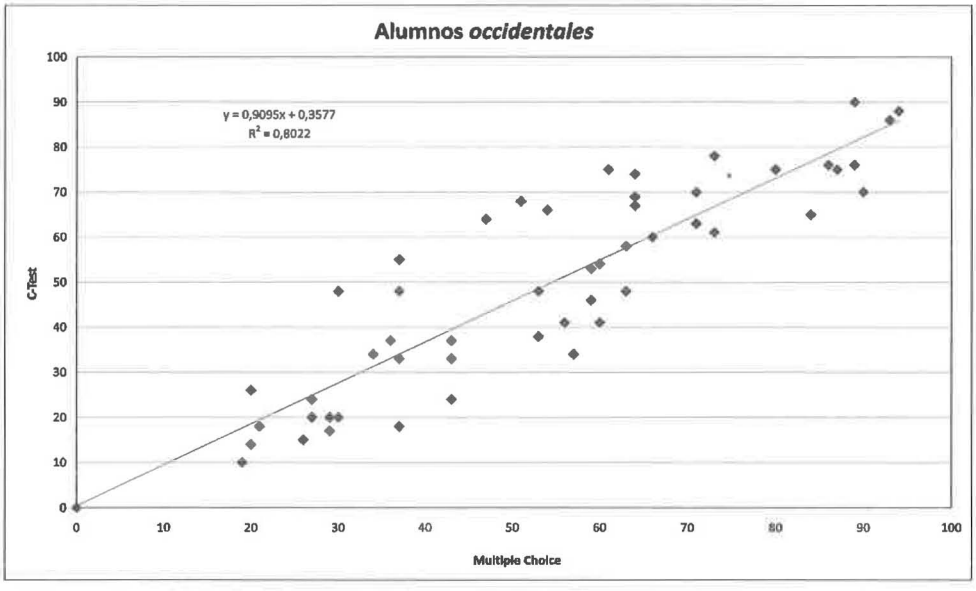
ANÁLISIS DE RESULTADOS

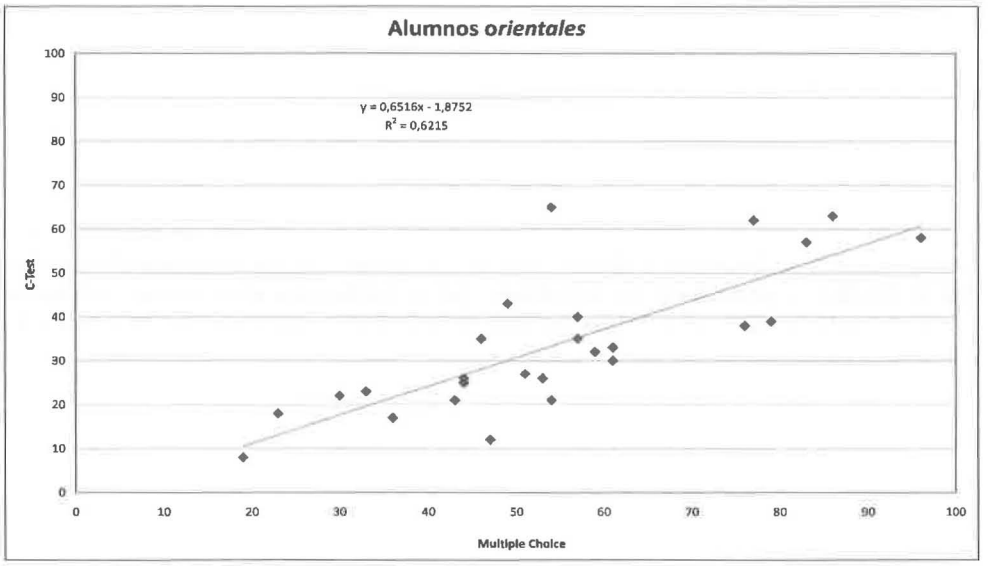
El análisis comparativo de los resultados obtenidos en el test de opción múltiple y en el *C-Test* mostró que, en esta segunda prueba, las calificaciones fueron aproximadamente un 17% más bajas, y la correlación de un 0,66, tal como se puede apreciar en el gráfico siguiente:



Puesto que la correlación no es excesivamente alta, podríamos pensar que no existe una relación directa y exacta entre las capacidades que una y otra prueba miden.

Sin embargo, un análisis pormenorizado de los casos particulares de cada estudiante mostró que los alumnos de origen oriental (taiwaneses, japoneses y chinos), que suponen una parte importante del alumnado del ILCE, obtuvieron resultados muy inferiores en la prueba de *C-Test*. Obsérvese, a continuación, un análisis contrastado de unos y otros alumnos:





Como se observará, la disparidad entre la correlación de ambas pruebas en los alumnos *occidentales* (0,8022) y *orientales* (0,6215) resulta muy llamativa.

El análisis de las estadísticas obtenidas, junto con un estudio de las pruebas adicionales que se llevaron a cabo (expresión escrita y expresión oral) nos permite emitir una serie de conclusiones:

1. Los resultados en el *C-Test* fueron, en este caso, inferiores a los de una prueba con ítems de opción múltiple. Esto puede deberse a varias causas. Una de ellas es el número de *aciertos aleatorios* que los ejercicios de opción múltiple permiten. Pero también debemos tener en cuenta que los ejercicios de opción múltiple resultan más conocidos y menos tediosos que los *C-Tests* (los cuales pueden resultar frustrantes cuando el alumno no está acostumbrado a este tipo de pruebas, o cuando, debido a una competencia baja, este no es capaz de deducir el sentido general del texto).
2. Existe una correlación relativamente alta entre los resultados de la prueba de opción múltiple y el *C-Test*, pero no a todos los niveles:
   2.1. Los resultados del *C-Test* fueron muy inferiores a los del test de opción múltiple en el caso de nuestros estudiantes orientales, como se ha comentado ya. Este hecho podría deberse a que su lengua materna resulta muy diferente del español (lo que, en niveles bajos de competencia, dificulta la comprensión global del texto); pero también puede atribuirse a que muchos de estos estudiantes están habituados a un aprendizaje excesivamente gramatical y poco comunicativo, y a una evaluación poco *productiva* mediante tests de opción múltiple y ejercicios similares.
   2.2. El resultado del *C-Test* fue más satisfactorio en aquellos estudiantes que demostraron una mayor competencia en las destrezas productivas o activas (expresión oral y escrita). Una vez más, el modelo de aprendizaje pudo haber jugado un papel fundamental (puesto que los estudiantes que demostraron una mayor habilidad comunicativa en las pruebas orales también obtuvieron mejores resultados en el *C-Test*).

Conclusiones generales

La fiabilidad, validez y objetividad de este tipo de pruebas han sido defendidas por diversos especialistas y estudios, aunque también criticadas por otros. Pese a ello, y tal como se ha querido demostrar, los *C-Tests* pueden ser un procedimiento útil y preciso para determinados tipos de evaluación (evaluación diagnóstica, pruebas de clasificación...) y en determinados contextos de aprendizaje.

En el caso que nos ocupa, hemos podido verificar que este tipo de pruebas resultan válidas para el fin previsto (puesto que nuestro propósito era determinar la competencia lingüística general de los estudiantes para clasificarlos en el grupo correspondiente de cada asignatura) y fiables (ya que sus resultados guardan relación con los de las otras pruebas que se llevaron a cabo).

Por otro lado, puede constatarse que el grado de satisfacción de los estudiantes respecto al grupo asignado fue alto. Como suele ser habitual en nuestro centro cada curso, algunos de los estudiantes (un 7,5%) solicitaron un cambio de nivel, arguyendo que el que se les había asignado era demasiado bajo. Se decidió que estos alumnos demostrasen su competencia mediante la prueba de clasificación del Aula Virtual de Español (AVE) del Instituto Cervantes. El nivel asignado según esta prueba coincidió, en el 83% de los casos, con el del *C-Test*.

El uso de los *C-Tests* ofrece, pues, diversas ventajas para el profesor y el alumno de lenguas extranjeras. Se trata de pruebas cuyo diseño es sencillo, y cuya corrección y resultados son inmediatos, exactos y objetivos. Gracias a las nuevas tecnologías, son, como se ha querido demostrar, un tipo adecuado de prueba a distancia.

Bibliografía

ARTOLA GONZÁLEZ, Teresa (1991): «El procedimiento cloze: una revisión general», *Revista Complutense de Educación.*, vol. 2, número 1, pp. 69-81.

CANTERA ADRADOS, Secundina (2011): *Modelos del* C-Test *y su correlación con los tests de la comprensión lectora de la E.O.I.* Tesis doctoral. Madrid: Universidad Complutense.

FARHADY, Hossein, y Ferdos JAMALI (1999): «Varieties of *C-Test* as measures of general proficiency». *Journal of the Faculty of Foreign Languages* (Tehran University Press) 3, pp. 23-42.

FISCHER, J.: «Das C-Test-Projekt zur Kalibrierung der UNIcert® –Abschlussniveaus – Zwischenergebnisse der Untersuchung», en O'DUILL, M. et al. *Zusammenarbeiten: eine Festschrift für Bernd Voss.* Bochum: AKS Verlag, 2005, pp. 102-103 (*Fremdsprache in Lehre und Forschung,* 37)

KHODADADY, Ebrahim: «C-Tests: Method Specific Measures of Language Proficiency», *Iranian Journal of Applied Linguistics,* vol 10, n.º 2, septiembre de 2007, pp. 1-26.

KLEIN-BRALEY, Christine: (1997): «*C-Tests* in the context of reduced redundancy testing: an appraisal», *Language Testing,* 14/1, 47-84.

LINARES I ZAPATER, Gemma (2009): *Los C-Tests (pruebas tipo C) en ELE: método de nivelación para conocimientos de lengua extranjera.* Memoria de Máster de ELE. Tarragona: Universitat Rovira i Virgili.

RAATZ, Ulrich y Christine KLEIN-BRALEY (1981): «The C-Test – a modification of the cloze procedure», *Practice and problems in language testing. University of Essex Occasional Papers.* Colchester, Essex: Department of Language and Linguistics, University of Essex, pp. 113-148.

SPOLSKY, Bernard (1971): «Reduced redundancy as a language testing tool», *Applications of linguistics.* Cambridge: Cambridge University Press, 383-390.

TAYLOR, Wilson (1953): «Cloze Procedure: A New Tool For Measuring Readability», *Journalism Quarterly* 30, 415-433.