

LA VALIDEZ DE IMAGEN DEL EXAMEN CELU

Este trabajo se realiza en el marco de la validación del examen CELU. Recordando, por un lado, los conceptos que, en el Segundo Coloquio Celu, nos dejara la Dra. Lorena Llosa sobre la consideración de la validación de un examen “como un proceso continuo que comienza con el argumento del uso y diseño de la evaluación y continúa con la recopilación de evidencia pertinente todo el tiempo que se use el examen para tomar decisiones”¹; y recordando también, por otro lado, mi aporte en el Primer Coloquio Celu, sobre los diferentes tipos de validez y en especial, al que me referí en aquella ocasión: la validez concurrente del Celu². Quiero presentarles hoy los resultados de la recopilación de datos obtenidos a través de encuestas realizadas a candidatos que rindieron el Celu en diferentes tomas con relación a la determinación de la Validez de Imagen o Validez Aparente (Face Validity), porque consideramos que este tipo de validez tiene una importancia significativa ya que la aceptación pública de una prueba podría modificar positiva o negativamente el impacto que ésta produce no sólo en los candidatos, sino también en la comunidad educativa en general y en los diferentes organismos que presentan alguna conexión con esta certificación.

Como una breve introducción quiero recordar que, siguiendo a Alderson³, la validez está relacionada con el grado de adecuación de un examen a lo que se considera que mide. En el capítulo 8, Alderson señala que existen diferentes métodos para determinar la validez de un examen y agrega que “Lo mejor es llevar a cabo la validación de una prueba en el mayor número de formas posibles”. En otras palabras, cuantos más ‘tipos’ de validez puedan establecerse, mejor; y cuanta más evidencia pueda recogerse para cada ‘tipo’ de validez, mejor”. Esta puede establecerse básicamente de tres maneras: mediante la demostración de a) la validez interna (estudia el contenido del examen y su efecto), b) la validez externa (hace referencia a la comparación de los resultados del examen con los de otras pruebas) y c) la validez de constructo de un examen (implica la valoración de hasta qué punto la prueba está basada de forma clara en la teoría que la subyace).

Voy a enfocarme, entonces, en uno de los aspectos de la validez interna del examen CELU, el de su imagen. Cabe aclarar que la validación interna se realiza mediante cuestionarios de diversos tipos con el fin de estimar la validez:

¹ Llosa, L. 2006. Estado actual de las investigaciones sobre validación de exámenes de lengua extranjera. <http://www.celu.edu.ar/archivos/pp.pps>

² Krsul, L. 2005. Para una validación concurrente del examen Celu. [http://www.celu.edu.ar/coloquios/Primer Coloquio. Reseñas y Ponencias.](http://www.celu.edu.ar/coloquios/Primer_Coloquio_Resenas_y_Ponencias)

³ Alderson, J.Ch. Clapham, C. Wall, D. Exámenes de idiomas. Elaboración y evaluación. Cambridge University Press. 1998.

a) del contenido: mediante cuestionarios realizados a expertos en evaluación y en enseñanza de lenguas se compara el contenido de los distintos ítems del examen con las especificaciones y el programa.

b) de la respuesta: también, a través de preguntas, pero esta vez a los candidatos para que opinen, retrospectivamente, su proceso de respuesta a las consignas del examen.

c) de la imagen que se tiene del examen: se realizan encuestas para obtener la opinión de los candidatos y de otras personas las cuales, sin ser expertas en el examen, están relacionadas con él, por ejemplo, docentes de lengua, administrativos, funcionarios, directores de institutos o personal de recursos humanos de empresas.

Mi intención es precisar el tipo c) de validez. Quiero aclarar, sin embargo, que por razones de tiempo dejo para un segundo paso de este trabajo el análisis de los datos obtenidos de las encuestas realizadas al segundo grupo mencionado, es decir, el de los no candidatos, quienes aportaron una opinión intuitiva del examen, por razones de tiempo.

Alderson sostiene que la validez de imagen obtenida a través de la opinión, tanto de candidatos como de no expertos, es, en general, global y se refiere a la prueba como un conjunto, aunque aclara que también los cuestionarios pueden incluir ítems concretos referidos, por ejemplo a las instrucciones en el examen o al tiempo asignado para realizar las consignas, entre otros. Y, agrega que si el término validez de imagen (o aparente) es usado en forma positiva equivaldría a "aceptable" para el usuario de un examen y también para la comunidad en la que se mueve. Y concluye diciendo que las pruebas que no les parecen válidas a los usuarios (es decir: las que ellos consideran que no miden lo que afirman medir) no se tomarán en serio para su propósito y, por el contrario, si los candidatos consideran que una prueba tiene un aspecto válido, creerán que pueden hacerla mejor y responderán a los ítems de forma apropiada⁴.

Es mucha la bibliografía especializada en evaluación que sostiene la importancia de este tipo de validez. Puedo citar, entre otros, a Hughes quien afirma que un examen que no tiene validez de imagen no puede ser aceptado por los candidatos, maestros, autoridades de educación o empleadores⁵. También Palencia de Burgo se refiere al tema diciendo que la prueba tiene que parecer correcta a simple vista, y agrega que, en definitiva, la validez de imagen resulta muy importante en evaluación porque una prueba sólo es tomada en serio por los usuarios si es percibida como válida⁶. Alan Davis va un poco más allá y señala que la falta de medición de validez de imagen de un examen puede poner en peligro la credibilidad pública del test y el currículo en el que el test está basado.⁷

⁴ Alderson. Op.cit. Capítulo 8

⁵ Hughes, A. *Testing for language teachers*. Cambridge University Press. Cambridge. 1994.

⁶ Palencia del Burgo, R. (1990) La evaluación como diagnóstico y control. En: *Didáctica de las segundas lenguas. Estrategias y recursos básicos*. Santillana, Aula XXI.

⁷ Alan Davies, Mc Namara. (1990) *Principles of Language Testing*. Oxford: Basil Blackwell.

Douglas Brown en su libro *Language Assessment*⁸ sostiene que un examen tiene validez de imagen positiva si los candidatos y los no expertos perciben que el test es válido y que, desde la perspectiva de ambos, evalúa lo que se propone evaluar. Agrega en su libro que, como ésta es una opinión intuitiva sobre cómo se puede percibir un test, algunos expertos en evaluación como, por ejemplo, Stevenson (1985) consideran que la validez de imagen es un factor superficial que sólo depende del capricho de quien lo percibe. Con respecto a lo señalado por Stevenson, Douglas Brown afirma que este tipo de validez es un concepto complejo pero indispensable para que los elaboradores de exámenes entiendan qué es lo que el público valora en un test. Y agrega que los estudiantes juzgarán, generalmente, que un examen es válido (y así tendrá una validez de imagen positiva) si:

- 1) Las instrucciones son claras.
- 2) La estructura del test está organizada lógicamente.
- 3) Su nivel de dificultad es apropiado.
- 4) El test no presenta imprevistos.
- 5) El tiempo es adecuado.

Teniendo en cuenta los conceptos expuestos por Douglas Brown, a continuación expondré los resultados de la información obtenida mediante las encuestas realizadas a los candidatos que se presentaron en las tomas 206 y 107, a fin de aportar datos que permitan considerar la Validez de Imagen del examen como un aporte positivo para la demostración de una validez global del CELU.

La encuesta que fue solicitada con carácter voluntario y anónimo a los candidatos una vez que finalizaron el examen escrito, fue completada por 87 alumnos en la toma 206 y 74 en la toma 107. En ella fueron incluidos los siguientes ítems en los cuales se debía optar por una respuesta. Los resultados fueron los siguientes:

a) Opinión sobre el examen

(en porcentajes)

Opiniones*	Toma 206	Toma 207
Difíciles	25	21
Aceptable	70	72
Fáciles	5	6
Total	100	100
Índice	5.0	3.5

* Se agruparon las opciones "Muy difícil" y "Difícil" en Difíciles; y las opiniones "Muy fácil" y "Fácil", en fáciles.

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

⁸ Douglas Brown, H. (2004) *Language Assessment: Principles and Classroom Practices*. Pearson Education, Longman.

La mayoría de las opiniones señala que el examen fue aceptable, es decir que el examen propone al candidato un desafío que se adecua a lo que él esperaba. Se destaca que, por un lado, un examen considerado fácil estaría señalando que el candidato queda insatisfecho ya que no puede demostrar todo lo que sabe. Por otro lado, un examen que es considerado difícil parecería ser más valorado por los candidatos ya que requeriría de ellos un mayor esfuerzo y un mayor dominio de la lengua.

b) Opinión sobre la cantidad de tiempo disponible para completar el examen
(en porcentajes)

Opiniones*	Toma 206	Toma 207
Mucho	12	4
El justo	82	86
Poco	6	10
Total	100	100
Índice	2.0	0.4

* Se agruparon las opciones "Excesivo" y "Muy largo", constituyendo el grupo de opiniones "mucho"; y las opiniones "Escaso" y "Muy corto", conformando el grupo "poco".

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

El tiempo del que dispusieron los candidatos resultó justo, en su percepción. De todas formas, mientras en la Toma 206 fueron más los que consideraron que el tiempo fue mucho antes que poco, en la Toma 107 la percepción cambia.

c) Opinión sobre las instrucciones recibidas
(en porcentajes)

Opiniones*	Toma 206	Toma 207
Precisas	86	96
Algunas con imprecisiones y otras no	11	4
Imprecisas	1	-
Ns/Nc	2	-
Total	100	100

* Se agruparon las opciones "Todas muy precisas" y "Todas precisas", constituyendo el grupo de opiniones "precisas"; y las opiniones "Todas muy imprecisas" y "Todas imprecisas", conformando el grupo "Imprecisas".

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

En opinión de los respondentes, las instrucciones recibidas fueron precisas, y prácticamente no hay nadie que haya contestado que fueron imprecisas. La contundencia de los números habla por sí sola, especialmente en la Toma 107.

d) Opinión sobre las actividades propuestas
(en porcentajes)

Opiniones*	Toma 206	Toma 207
Interesantes	92	84
Indiferentes	5	7
Aburridas	3	5
Ns/Nc	-	4
Total	100	100

* Se agruparon las opciones “muy interesantes” y “interesantes”, constituyendo el grupo de opiniones “interesantes”; y las opiniones “Muy aburridas” y “Aburridas”, conformando el grupo “aburridas”.

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

Las actividades propuestas resultaron muy interesantes, sobre todo en la Toma 206.

e) En las clases para aprender español, ¿cómo eran las actividades respecto del examen CELU?

(en porcentajes)

Opiniones*	Toma 206	Toma 207
Similares	49	46
Distintas	38	51
Nunca cursé español	9	1
Ns/Nc	4	2
Total	100	100

* Se agruparon las opciones “totalmente similares” y “algo parecidas”, constituyendo el grupo de opiniones “similares”; y las opiniones “totalmente distintas” y “distintas”, conformando el grupo “distintas”.

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

Aunque en este punto las opiniones se encuentran muy divididas, para los candidatos de la Toma 206 las actividades del examen fueron parecidas a las de sus cursos de español.

La diferencia porcentual (9 puntos porcentuales) se amplía si se excluyen las opiniones de los que no cursaron español (54% considera que fueron parecidas, 42% que fueron distintas).

Sin embargo, dada la escasa cantidad de respondientes, no conviene sacar conclusiones apresuradas. De todos modos, se destaca que la tendencia se invierte en la Toma 107.

f) Respecto de la vida real, ¿cómo fueron las situaciones propuestas en el examen CELU?

(en porcentajes)

Opiniones*	Toma 206	Toma 207
Parecidas	67	68
En parte parecidas y en parte distintas	30	31
Distintas	3	-

Ns/Nc	-	1
Total	100	100

* Se agruparon las opciones “totalmente similares” y “algo parecidas”, constituyendo el grupo de opiniones “parecidas”; y las opiniones “totalmente distintas” y “distintas”, conformando el grupo “distintas”.

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

En relación con esta pregunta, los resultados son muy significativos: alrededor de dos tercios opina que las situaciones propuestas por el CELU fueron similares a las de la vida real, mientras el tercio restante considera que solo lo fueron en parte.

g) ¿Cómo fue la organización del examen CELU?

(en porcentajes)

Opiniones*	Toma 206	Toma 207
Positivas	90	92
Ni buena ni mala	7	8
Negativas	2	-
Ns/Nc	1	-
Total	100	100

* Se agruparon las opciones “muy buena” y “buena”, constituyendo el grupo de opiniones “positivas”; y las opiniones “muy mala” y “mala”, conformando el grupo “negativas”.

Base: total entrevistados (87 en la toma 206 y 74 en la toma 107)

La organización del examen fue, en opinión de los consultados, excelente: así lo consideran 9 de cada 10.

Los resultados obtenidos por medio de esta encuesta son susceptibles de nuevos análisis y otras comparaciones, por ejemplo con estadísticas referidas al perfil de los alumnos que rindieron en cada fecha. Esto va a permitir, en el futuro, tener una visión más detallada de lo que los candidatos nos dicen con respecto al examen y así considerar la Validez de Imagen del Celu como positiva o negativa. Por ahora, solamente puedo señalar que los resultados obtenidos en las dos tomas, en porcentajes, son bastante coincidentes y tienden a demostrar que los candidatos consideran positivamente el examen.

De todos modos, me gustaría señalar que el punto e) relacionado con las actividades que realizan en los cursos de español y las actividades propuestas en el examen, es el que presenta una mayor diferencia entre las dos tomas. En el caso de la toma 107 presenta un mayor reparto en la elección. Desde el punto de vista del Impacto, que podría tener el examen Celu en la enseñanza general del español.

Será importante seguir proponiendo esta encuesta y comparar estos resultados con las tomas futuras del examen para ver si se concreta una modificación en

estos valores que represente la demostración de impacto del Celu en la enseñanza de español.