

El Certificado Internacional de Lengua Española (CILES): análisis psicométrico y adecuación al Marco Común Europeo de Referencia (MCEUR)

Iris Viviana Bosio

irisvivianabosio@uda.edu.ar

Gloria Ángela Ginevra

gloria.ginevra@gmail.com

Hilda Emilia Difabio de Anglat

Celsa Mabel Anta de Bernabé

Carolina Ana Sacerdote

Becaria: Romina Gisela Cangiano

Universidad del Aconcagua

Escuela Superior de Lenguas Extranjeras

En este trabajo presentaremos los resultados del proyecto de investigación ejecutado en el periodo 2008-2010 en el marco del CIUDA (Consejo de Investigaciones de la Universidad del Aconcagua).

1. Objetivos del proyecto:

1) Determinar las cualidades psicométricas del CILES (Certificado Internacional de Lengua Española), certificación de dominio del español como lengua extranjera diseñado y administrado en la Escuela Superior de Lenguas Extranjeras (UDA)

2) Reformular la certificación original para:

a) cumplir con las condiciones de un adecuado proceso de evaluación (validez, confiabilidad y practicidad)

b) adecuar los niveles del examen a los estándares del MCEUR (Marco Común Europeo de Referencia)

2. Hipótesis de trabajo del proyecto:

La determinación de las cualidades psicométricas del CILES y la adecuación de sus niveles a los estándares del MCEUR (Marco Común Europeo de Referencia), daría como resultado la optimización operativa del instrumento.

3. Metodología utilizada en el proyecto:

En la primera etapa de la investigación -determinación de las cualidades psicométricas del CILES (Certificado Internacional de Lengua Española)-, el tipo de diseño adoptado fue cuantitativo, seccional y descriptivo.

Se analizó la totalidad de los exámenes CILES (161) administrados por la ESLE durante el periodo 1998-2000.

Los exámenes están organizados en tres niveles: inicial, intermedio y avanzado. Su objetivo es evaluar a los candidatos para certificar su competencia lingüístico-comunicativa en comprensión lectora y auditiva, por un lado, y en producción escrita y oral, por el otro. En todos los niveles, el CILES evalúa competencia gramatical, competencia textual, competencia ilocutiva y competencia sociolingüística (Bachman, 1990). Asimismo, el examen se organiza en una secuencia de cinco secciones:

- 1) Comprensión auditiva
- 2) Comprensión de textos
- 3) Gramática
- 4) Producción escrita
- 5) Producción oral

Los datos fueron obtenidos de fuentes primarias, dado que provinieron de la resolución de las secciones “comprensión auditiva”, “comprensión lectora” y “gramática” de la muestra. Las secciones “producción escrita” y “producción oral” no fueron tomadas como fuentes de datos, debido a que no se establecieron parámetros de medición de resultados para secciones de respuestas abiertas.

El instrumental técnico empleado fue el programa informático estadístico SPSS (Statistics Package for Social Sciences).

Se calcularon cuatro cualidades psicométricas: 1) **confiabilidad** mediante el *alpha de Cronbach* (estadístico habitual en la investigación contemporánea); 2) homogeneidad de los ítems a través del índice de correlación de Pearson entre cada ítem y la puntuación total de la prueba como índice de la consistencia interna, una aproximación a la **validez de constructo** (Anastasi, 1973: 122); 3) **discriminación de los ítems** por medio de los coeficientes de correlación biserial normalizados; 4) **nivel de dificultad** como el promedio de los porcentajes de respuestas correctas en cada ítem en el grupo superior (por encima de la mediana) y en el inferior (por debajo de la misma).

Respecto del cálculo del nivel de dificultad de cada consigna, se mantuvieron los porcentajes por nivel de dificultad pero se adoptó una categorización distinta de la propuesta en el proyecto, dado que discriminó mejor dichos niveles esperados: 10%: ítems muy fáciles; 20%: fáciles; 40%: moderados; 20%: difíciles; 10%: muy difíciles.

Nivel de dificultad	Categorización inicialmente propuesta	Categorización efectivamente aplicada (Backhoff et al., 2000)
Muy fácil	0,75 a 1	0,87 a 1
Fácil	0,55 a 0,74	0,74 a 0,86
Moderado	0,45 a 0,54	0,53 a 0,73
Difícil	0,25 a 0,44	0,33 a 0,52
Muy difícil	0,00 a 0,24	0,00 a 0,32

En la segunda etapa de la investigación -Estandarización según el Marco Común Europeo de Referencia (MCEUR)-, se realizaron los siguientes pasos metodológicos:

- a) Revisión de las versiones vigentes del CILES en sus tres niveles y adecuación a los estándares del MCEUR, a partir de la comparación de los descriptores de competencias del candidato con las secciones del examen sometidas a revisión (comprensión auditiva, comprensión lectora, gramática y vocabulario).
- b) Revisión, reordenamiento y reformulación por nivel y por sección, según una complejidad creciente, de: la competencia gramatical, la comprensión de diferentes tipos textuales y el dominio de vocabulario.
- c) Consideración y aplicación de los criterios estadísticos de validez y confiabilidad en la producción de consignas y/o estímulos, respuestas correctas y distractores.

4. Resultados del proyecto:

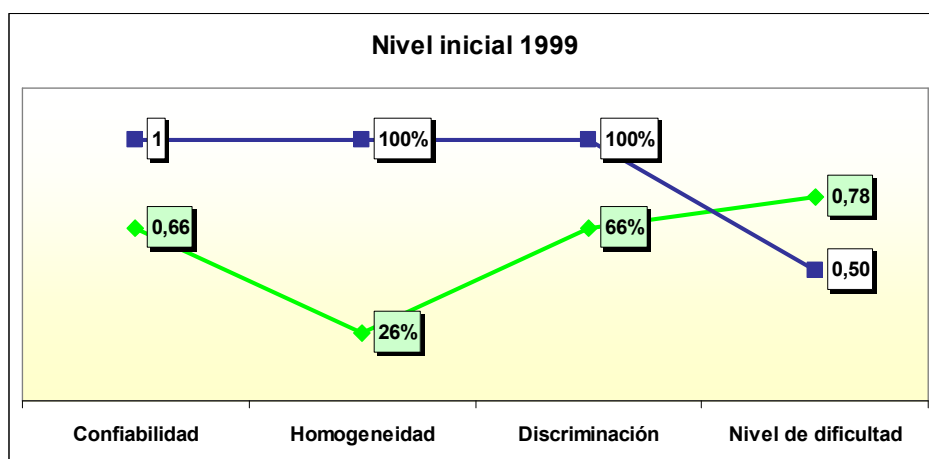
4.1. Primera etapa de la investigación:

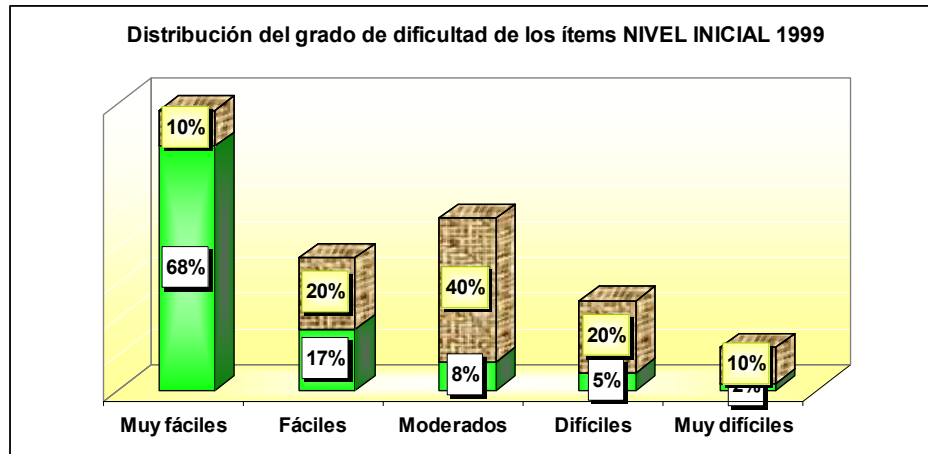
Para presentar los resultados, se incluyen los datos descriptivos, los gráficos de síntesis (por nivel -inicial, intermedio y avanzado- y por año) y la interpretación de los mismos por niveles.

1. Nivel inicial

NIVEL INICIAL AÑO 1999

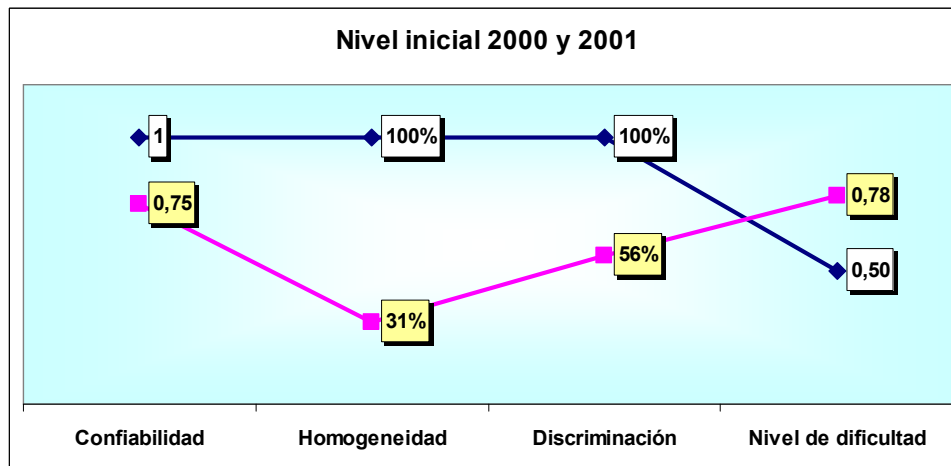
CASOS	38
ÍTEMES	54
MEDIA	49.92
D.S.	3.026
MEDIANA	51
PJE. MINIMO	42 (el 78% del instrumento)
PJE. MÁXIMO	54 (el total posible)

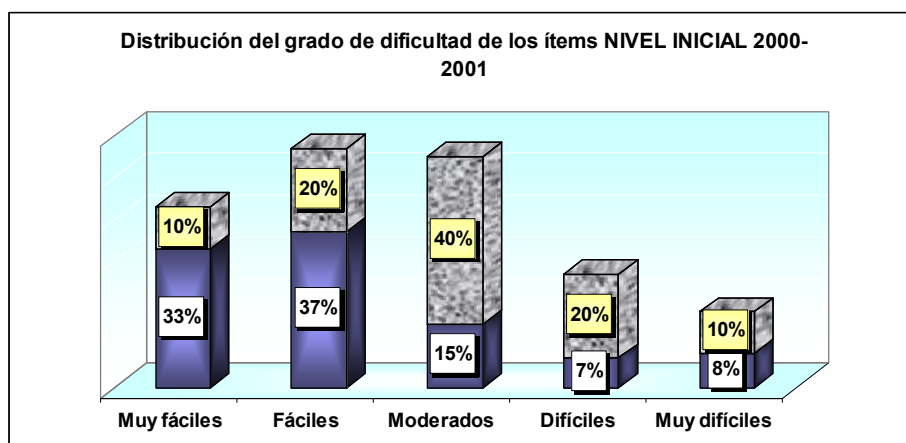




NIVEL INICIAL AÑO 2000-2001

CASOS	25
ÍTEMES	55
MEDIA	47.16
D.S.	4.432
MEDIANA	48
PJE. MINIMO	34 (el 62%)
PJE. MÁXIMO	52 (el 94% -total posible: 55-)





En suma, en el **nivel inicial**:

a) Es necesario incrementar la *confiabilidad* a un índice de 0.80¹.

b) Resulta muy acertada la modalidad de evaluación de la *comprensión auditiva*, pero:

- En la versión de 1999 ocho ítems de 15 (el 53%) tienen índice de discriminación nulo porque todos los sujetos los responden correctamente (Nº 1, 3, 6, 7, 8, 10, 11 y 14); 3 ítems -el Nº 9, 13 y 15- evidencian un error; sólo tres alcanzan validez significativa de construcción.
- En la versión de 2000-2001, si bien hay variabilidad, cuatro ítems -el Nº 2, 9, 10 y 12- manifiestan un solo error.
- Luego, a fin de aumentar la homogeneidad de los ítems y su nivel de dificultad, se podrían presentar las opciones de respuesta simultáneamente a la audición.

En la evaluación de las tres competencias, es necesario reformular los distractores sin elección -esto es, que convierten al ítem en dicotómico o variable con dos valores solamente- porque, evidentemente, su contribución es nula.

La combinación de ambas muestras en los apartados comunes *-lectura y comprensión de textos y gramática-* evidencia:

- Cuatro ítems sin variabilidad: el Nº 4 y 9; el Nº 1 y 7, respectivamente.
- El 46% arroja adecuada validez de construcción, porque 18 ítems de 39 (seis en *lectura y comprensión de textos*: Nº 2, 3, 5, 6, 13 y 14; doce en *gramática*: Nº 2, 3, 5, 8, 10, 12, 13, 15, 16, 17, 19 y 20) correlacionan significativamente con el puntaje total del *test* (entendido en este contexto como "apartados comunes"). El estadístico utilizado nos dice si las personas "adecuadas" son las que responden correctamente el ítem e incluso informa sobre la validez predictiva del instrumento.

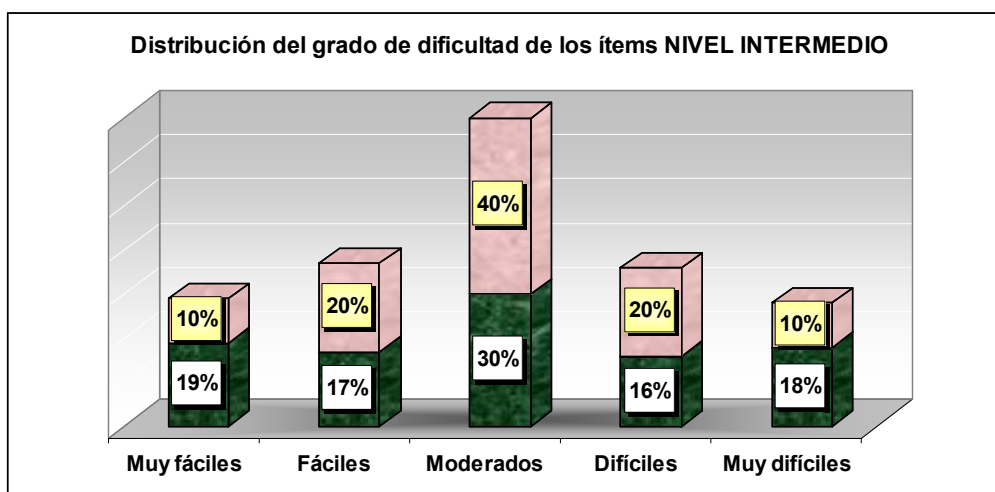
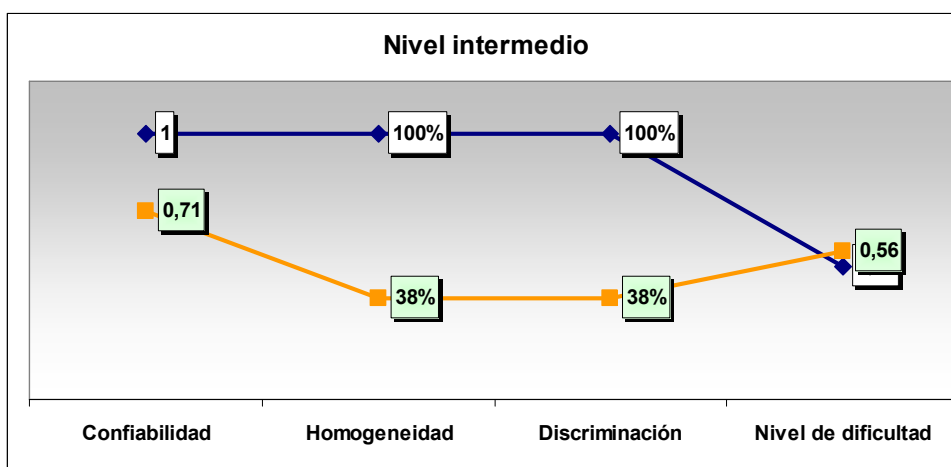
El *test* resulta demasiado sencillo: con un grado de dificultad de 0.78 o más. La teoría psicométrica aconseja que el nivel medio de dificultad de un examen oscile entre 0.5 y 0.6 (Backhoff et al, 2000: 4). La proporción de examinados que responden correctamente un ítem afecta su potencialidad discriminativa: si es respondido correctamente (o incorrectamente -aunque este caso sólo se verifica en el ítem Nº 3 de comprensión auditiva, versión 2000/2001-) por más del 85% tiene poder nulo de discriminación. Conviene que el 90% de los examinados responda correctamente un rango entre 30% y 80% de los ítems.

¹ "(...) tests with over 50 items should yield KR-20 values of .8 or higher. Unsatisfactorily low KR-20s are usually due to an excess of very easy (or hard) items, poorly written items that do not discriminate, or violation of the precondition that the items test a unified body of content." (Kehoe, 1995: 2)

2. Nivel intermedio

NIVEL INTERMEDIO

CASOS	41	
ÍTEMES	55	
MEDIA		41.05
D.S.	4.50	
MEDIANA	42	
PJE. MÍNIMO		30 (54% del test)
PJE. MÁXIMO		49 (89% -total posible: 55-)



El análisis de las cualidades psicométricas del nivel intermedio del *CILES* (versión 1999-2001) en sus ítems objetivos manifiesta:

a) Es el nivel que evidencia mejor índice de confiabilidad -0.71- (es decir, que podemos asegurar que no nos equivocamos en el 70% de la medición), aunque, como señaláramos, para *tests* con más de 50 ítems conviene llevar este coeficiente a 0.80.

b) Respecto de la validez de construcción del instrumento:

- En *comprensión auditiva*, el 53% de los ítems alcanza índices

significativos (Nº 1, 2, 3, 7, 10, 11 y 12) o muy cercano (el Nº 5).

- En *lectura y comprensión de textos*, tres ítems -Nº 14, 15 y 18- resultan sin variabilidad (todos los sujetos los responden correctamente), cinco evidencian adecuada validez de construcción (Nº 5, 9, 13, 16 y 17) y dos se acercan a un índice significativo (Nº 4 y 11); luego, se podría replicar el 35%.
- En *gramática*, dos son los ítems sin variación (Nº 4 y 6), también son cinco los que muestran índices significativos (Nº 7, 10, 14, 17 y 19).

Es necesario reformular los distractores sin elección.

c) Los ítems que *discriminan* adecuadamente entre los sujetos de mejor y peor rendimiento son, en su mayoría, los mismos que alcanzan coeficientes significativos o cercanos de validez de construcción:

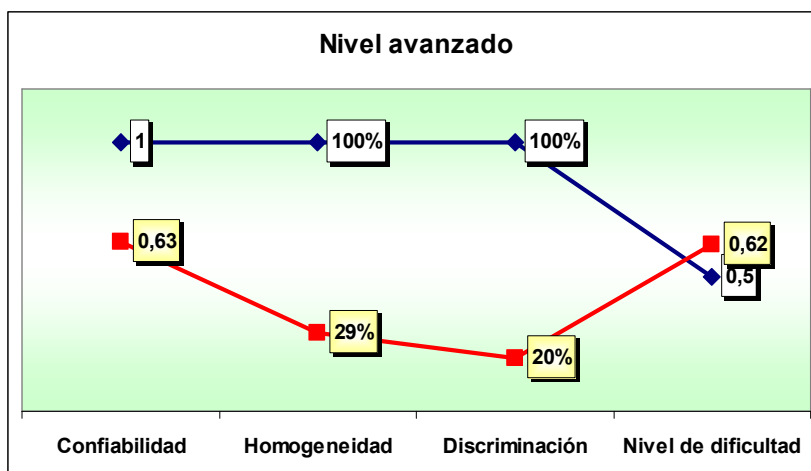
- En *comprensión auditiva*, el Nº 1, 2, 3, 5, 7, 10, 11 y 12 (8 ítems).
- En *lectura y comprensión de textos*, el Nº 4, 5, 9, 13, 16 y 17; a estos, se suma el ítem 8 (7 ítems).
- En *gramática*, el Nº 7, 10, 14, 17 y 19; se agrega el Nº 8 (5 ítems).

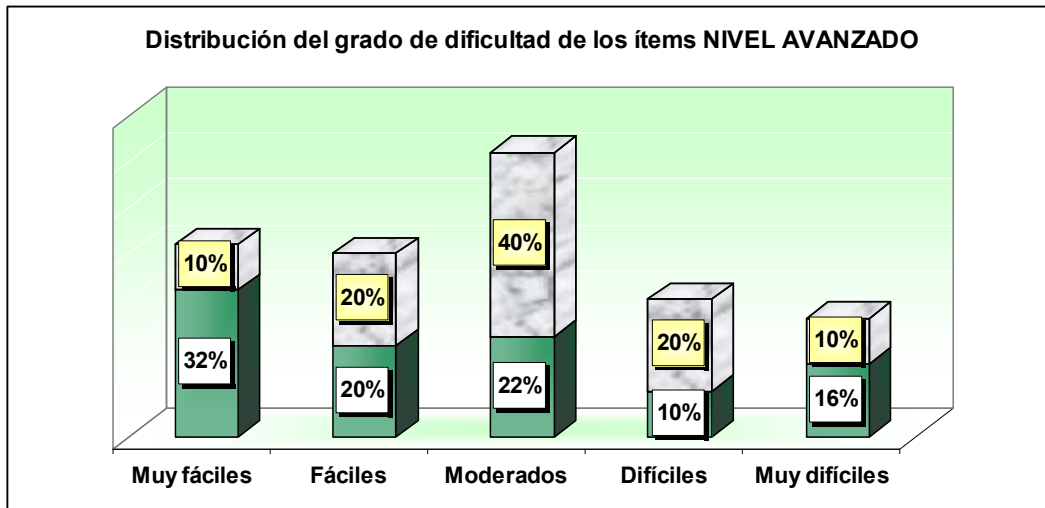
d) Los tres subtests manifiestan un promedio del nivel de dificultad bastante próximo al índice esperado para un *test* de opción múltiple; respecto de su distribución, en *comprensión auditiva* se alcanza el porcentaje más ajustado -en comparación con los otros niveles- en la categoría “moderado” (el 34%) y las otras categorías se acercan bastante a los coeficientes ideales; en *lectura y comprensión de textos*, aparecen demasiados ítems muy fáciles (30%); por el contrario, en *gramática* sobre-abundan los muy difíciles (25%).

3. Nivel avanzado

NIVEL AVANZADO

CASOS	57
ÍTEMES	55
MEDIA	44.05
D.S.	4.16
MEDIANA	45
PJE. MINIMO	35
PJE. MÁXIMO	52





El análisis de las cualidades psicométricas de los ítems objetivos del nivel avanzado del CILES versión 1999-2002 evidencia:

a) El coeficiente de *confiabilidad* -0.63- resulta escaso para un *test* de respuesta cerrada de más de 50 ítems.

b) Respecto de la validez de construcción del instrumento:

- En *comprensión auditiva*, cuatro ítems (Nº 2, 4, 12 y 15) -el 27%- alcanzan índices significativos.
- En *lectura y comprensión de textos*, dos ítems -Nº 1 y 5- resultan sin variabilidad (todos los sujetos los responden correctamente) y nueve evidencian adecuada validez de construcción (Nº 3, 7, 9, 11, 12, 13, 16, 19 y 20) -el 45%-.
- En *gramática*, dos son los ítems sin variación (Nº 8 y 16) y sólo tres (Nº 11, 13 y 17), que representan el 15% del apartado, muestran índices significativos

Nuevamente, destacamos que es necesario reformular los distractores sin elección.

c) Respecto de la *potencialidad discriminativa* del instrumento:

- En *comprensión auditiva*, ningún ítem alcanza el índice de discriminación buena o muy buena y solamente dos (Nº 3 y 12), categoría regular -los demás resultan defectuosos-.
- El 35% de *lectura y comprensión de textos* es discriminativo (se trata de siete de los ítems con adecuada validez de construcción -Nº 7, 9, 11, 12, 13, 16 y 19-); los otros dos ítems válidos -el Nº 3 y 20- logran un índice regular en discriminación.
- En *gramática*, el 25% resulta discriminativo: los tres ítems válidos -Nº 11, 13 y 17- más el Nº 14 y el Nº 20.

d) Desde su *nivel de dificultad*, el promedio de las tres sub-escalas se acerca al nivel esperado, si bien el 73% de los ítems de *comprensión auditiva* se ubica en la categoría “muy fácil” o “fácil”; en *lectura y comprensión*, a excepción de la categoría moderado (que manifiesta la mitad del porcentaje esperado), la distribución es bastante ajustada, patrón que se repite en *gramática*.

4.2. Segunda etapa de la investigación:

Dado que se ha cumplido el objetivo de adecuar los niveles del examen a los estándares del MCEUR (Marco Común Europeo de Referencia), en el siguiente cuadro comparativo se presentará la relación entre los niveles inicial, intermedio y avanzado del CILES y los niveles A2+, B1 y C1 del MCEUR respectivamente:

NIVELES CILES	NIVELES MCEUR	Descriptor de competencias del candidato
Nivel inicial	Nivel Plataforma (A2+)	Comprensión auditiva
		<i>Comprende frases y vocabulario habitual sobre temas de interés personal (información personal y familiar muy básica, compras, lugar de residencia, empleo). Es capaz de captar la idea principal de avisos y mensajes breves, claros y sencillos.</i>
		Comprensión lectora
		<i>Es capaz de leer textos breves y sencillos; encontrar información específica y predecible en escritos sencillos y cotidianos como anuncios publicitarios, prospectos, menus, horarios. Comprende cartas personales, breves y sencillas.</i>
		Gramática
		<i>Utiliza correctamente descriptores sencillos de la organización gramatical: categorías (tiempos simples del modo indicativo); relaciones (concordancia verbal y de género/ número en sustantivos y adjetivos); clases (pronombres y adverbios en función deíctica).</i>
		Vocabulario
		<i>Manifiesta un dominio limitado relativo a necesidades concretas y cotidianas y temas conocidos</i>
Nivel Intermedio	B1	Comprensión auditiva
		<i>Comprende información concreta relativa a temas cotidianos o al trabajo e identifica tanto el mensaje general como detalles específicos, cuando el discurso está articulado con claridad.</i>
		Comprensión lectora
		<i>Es capaz de comprender los puntos principales de textos claros y en lengua estándar si tratan sobre cuestiones que le son conocidas, ya sea en situaciones de trabajo, de estudio o de ocio. Comprende la descripción de acontecimientos, sentimientos y deseos en textos periodísticos expositivos y narrativos.</i>
		Gramática
		<i>Manifiesta tener un buen control gramatical de estructuras (cláusulas coordinadas y subordinadas, conectores y marcadores del discurso); de categorías (tiempos verbales del modo indicativo y subjuntivo, verbos irregulares, grados de significación del adjetivo); de clases (pronombres objeto, adjetivos y pronombres indefinidos y posesivos, pronombres relativos); de relaciones (concordancia gramatical).</i>
		Vocabulario
		<i>Manifiesta un buen dominio del vocabulario elemental en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos.</i>
Nivel Avanzado	C1	Comprensión auditiva
		<i>Es capaz de reconocer información específica,</i>

Avanzado		<i>realizar inferencias lógicas y pragmáticas en conferencias y presentaciones en público</i>
		Comprensión lectora
		<i>Es capaz de entender las ideas principales de textos extensos y complejos, especialmente de tipo argumentativo, que traten de temas tanto concretos como abstractos, incluso textos de divulgación técnico-científica.</i>
		Gramática
		<i>Manifiesta tener un buen control de descriptores complejos de la organización gramatical: categorías (tiempos condicionales y modo subjuntivo); clases (usos de “se”, pronombres objeto, gerundio, verbos impersonales); procesos de transformación (pronominalizaciones); estructuras (oraciones condicionales y concesivas, cláusulas subordinadas sustantivas, adjetivas y adverbiales, queísmo y dequeísmo).</i>
		Vocabulario
	<i>Manifiesta dominio de un amplio y variado repertorio léxico adecuado a distintos registros (+/- especializados; +/- formales; escritos y orales).</i>	

5. Importancia del proyecto:

Nuestro trabajo sobre el CILES intenta realizar un aporte significativo al campo de la evaluación en Español como Lengua Segunda y Extranjera (ELSE), ya que, por una parte, contribuye a la construcción del conocimiento del objeto de estudio y por otra parte, socializa este conocimiento para el enriquecimiento de la comunidad científica. Además, esta investigación intenta sentar precedentes en el ámbito local en la estandarización de certificaciones de ELSE.

Habiendo obtenido una certificación en ELSE confiable, válida y adecuada al MCEUR, las universidades argentinas pueden contar con un examen internacional de calidad, que mida la competencia en lengua española de aquellos candidatos extranjeros que deseen estudiar en las mismas. De esta manera, el examen se constituiría en una de las herramientas disponibles por parte de las universidades, para lograr el objetivo de una efectiva integración idiomática en el proceso de internacionalización de los claustros docentes y del alumnado.

6. BIBLIOGRAFÍA

- Anastasi, A. (1973). *Tests psicológicos*. Madrid: Aguilar.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. and A. Palmer (1996). *Language Testing In Practice*. Oxford: Oxford University Press.
- Backhoff, E. et al. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de investigación Educativa*, 2(1). Obtenido el 8 de junio de 2004 desde <http://redie.ens.uabc.mx/vol2no1/contenido-backhoff.html>
- Congresos internacionales de la lengua española. Obtenido el 12 de junio de 2009 desde <http://congresosdelalengua.es>
- Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Trad. en español del Instituto

- Cervantes. Obtenido el 15 de febrero de 2009 desde http://cvc.cervantes.es/obref/marco/cvc_mer.pdf
- Cronbach, L. (1963). *Fundamentos de la exploración psicológica*. Madrid: Biblioteca Nueva
- English Testing Services (2009). *Research Portfolio. Research & Development Division*. Obtenido 12 de agosto de 2009 desde <http://www.ets.org>
- Fox, D. (1981). *El proceso de investigación en educación*. Pamplona: Eunsa.
- Garret, H. (1974). *Estadística en Psicología y Educación*. 3º ed. Buenos Aires: Paidós.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kehoe, J. (1995). *Basic Item Analysis for Multiple-Choice Tests*. Eric/Digest. ERIC Clearing House on Assessment and Evaluation: Washington DC.
- Krsul, L. (2005). *Para una validación concurrente del examen CELU*. Ponencia presentada en el Primer Coloquio CELU "Investigaciones en evaluación de Español Lengua Segunda y Extranjera". Obtenido el 15 de mayo de 2009 desde http://www.celu.edu.ar/images/stories/pdf/coloquios/1_coloquio/cu_p_validez_concurrente_celu_krsul.pdf
- Palacios Martínez, I. (dir.) 2007. *Diccionario de enseñanza y aprendizaje de lenguas*. Madrid: enCLAVE-ELE.
- Palmer, A. y otros (2000). Tutorial sobre coeficientes de correlación con una o dos variables categóricas. *Revista Electrónica de Psicología*, 4(2).
- Prati, S. (2007). *La evaluación en español lengua extranjera: elaboración de exámenes*. Buenos Aires: Libros de la Araucaria.
- Rafaelli, G. y M. Vázquez (2008). *La evaluación de la comprensión y expresión escrita en los exámenes CELU Y DELE*. Ponencia presentada en el Cuarto coloquio CELU "Interculturalidad e Internacionalización en las prácticas evaluadoras: del trabajo en el aula a la certificación". Obtenido el 15 de mayo de 2009 desde http://www.celu.edu.ar/images/stories/pdf/coloquios/4_coloquio/cc_p_la_evaluacion_de_la_comprension_y_expresion_escrita_en_celu_y_dele_rafaelliVa.pdf
- Vianna, H. (1983). *Los tests en educación*. Pamplona: EUSA.