

## Confiabilidad y propiedades psicométricas de los ítemes: prueba CODICE-Derecho Universidad de Chile

### *Reliability and psychometric properties of the items: CODICE-Derecho test, University of Chile*

Jorge Hernández Muñoz<sup>1</sup>

**Resumen:** la prueba CODICE-Derecho es un instrumento de evaluación diseñado en conjunto entre el DEMRE y la Facultad de Derecho, ambos de la Universidad de Chile, con el objeto de precisar las competencias de lectura y escritura de sus estudiantes. El artículo ofrece un análisis comparativo de los resultados obtenidos en las aplicaciones del año 2009 y del 2011, el cual considera la confiabilidad del instrumento y dos indicadores psicométricos (la dificultad y el índice de discriminación). Por último, se destaca el carácter piloto del instrumento y se plantean algunos análisis estadísticos futuros con los resultados obtenidos.

**Palabras clave:** *confiabilidad, análisis psicométrico, instrumento de evaluación.*

**Abstract:** CODICE-Derecho test is an assessment tool designed jointly by the DEMRE and the Faculty of Law -both are departments of Universidad de Chile- in order to clarify the reading and writing skills of their students. The paper provides a comparative analysis of the results obtained from tests conducted in 2009 and 2011, which considers the reliability of the instrument and two psychometric indicators (difficulty and discrimination index). Finally, it highlights the pilot character of the instrument and suggest some future statistical analysis based on results described.

**Key words:** *Reliability, psychometric analysis, instrument of assessment.*

## Presentación

Las sociedades actuales reconocen la importancia de un adecuado desarrollo de competencias asociadas a la lectura y la escritura para que las personas puedan insertarse y participar de manera activa y responsable en el mundo laboral y profesional<sup>2</sup>.

De hecho, para que las interacciones de la vida cotidiana sean satisfactorias, se demanda de las personas competencias comunicativas, tanto en el plano escrito como en el oral, para el procesamiento y construcción de conocimientos<sup>3</sup>.

---

<sup>1</sup> Profesor de Castellano, Universidad Metropolitana de Ciencias de la Educación. Magíster en Lingüística Hispánica, Universidad de Santiago de Chile. Jefe (S) Unidad de Estudios e Investigación del Departamento de Evaluación, Medición y Registro Educacional (DEMRE), Universidad de Chile. Correo electrónico: jorherna@uchile.cl.

<sup>2</sup> Ver más en: PERONARD, Marianne; GÓMEZ MACKER, Luis; PARODI, Giovanni y NÚÑEZ, Paulina. *Comprensión de textos escritos: De la teoría a la sala de clases*. Santiago: Editorial Andrés Bello, 1998; y CARLINO, P. *Escribir, leer y aprender en la universidad. Una introducción a la alfabetización académica*. México: Fondo de Cultura Económica, 2005.

Las experiencias cotidianas de contacto con la lectura y escritura en una sociedad alfabetizada durante los primeros años de vida<sup>4</sup> se sistematizan con el ingreso a la escuela con el objetivo de que continúen enriqueciéndose de manera continua en la educación superior y el mundo académico, así como el laboral y profesional<sup>5</sup>.

Por esta razón, los procesos de admisión a la educación superior incorporan instrumentos de evaluación orientados a la evaluación de las capacidades de lectura y de escritura. Particularmente en Chile, la batería de pruebas de selección universitaria PSU®, incluye la prueba de Lenguaje y Comunicación, en la cual se evalúan las capacidades lectoras.

Sin embargo, la información entregada por esta medición está orientada a la Educación Superior en general, y no responde exclusivamente a una carrera en particular. Esto hace que exista la necesidad por parte de muchas instituciones, especialmente en aquellas carreras pertenecientes al ámbito de las Ciencias Sociales, de contar con información complementaria respecto a estas competencias. Sobre todo, si se considera que cada carrera universitaria tiene sus propias maneras de “leer” y “escribir” como parte de su quehacer académico<sup>6</sup>.

En este contexto, con el propósito de contar con información para potenciar la toma de decisiones académicas y pedagógicas, las autoridades de la Facultad de Derecho de la Universidad de Chile en conjunto con el Departamento de Evaluación, Medición y Registro Educativo (DEMRE) desarrollaron el año 2009 en forma piloto un instrumento estandarizado, denominado CODICE–Derecho, orientado a la medición de los niveles de desarrollo de las competencias discursivas en lectura y en escritura de sus estudiantes de primer año.

El proyecto CODICE–Derecho desarrolló aplicaciones piloto los años 2009 y 2011. En ambos casos, la prueba estuvo compuesta de preguntas de selección múltiple (cerradas) para la medición de las competencias lectoras y preguntas de desarrollo (abiertas) para la evaluación de las habilidades escriturales.

Esta exposición presenta los temas técnicos más relevantes de ambas aplicaciones de la prueba CODICE–Derecho. En primer lugar, se señalarán las características técnicas de los instrumentos aplicados. En segundo lugar, se mostrarán los resultados de ambas aplicaciones: las características del grupo de estudiantes que participó, la confiabilidad estimada del instrumento y, finalmente, las propiedades psicométricas de los ítems.

---

<sup>3</sup> Ver más en: ATORRESI, A. *Competencias para la vida en las evaluaciones de lectura y escritura* (SERCE-LLECE, LABORATORIO LATINOAMERICANO DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN), UNESCO-OREALC, 2005. Disponible en web: [http://www.oei.es/evaluacioneducativa/competencias\\_vida\\_evaluaciones\\_lectura\\_escritura\\_atorresi.pdf](http://www.oei.es/evaluacioneducativa/competencias_vida_evaluaciones_lectura_escritura_atorresi.pdf) [consultado el 15 de noviembre de 2011]; SOLÉ, I. *Estrategias de Lectura*. Barcelona: Editorial Graó, 2006; y PARODI, G., PERONARD, M. e IBÁÑEZ, R. *Saber leer*. Madrid: Editorial Aguilar, 2010.

<sup>4</sup> GOODMAN, Y. “Las raíces de la Alfabetización”. En: *Revista Infancia y aprendizaje*. 1992, núm. 58. P. 29 - 42.

<sup>5</sup> Ver más en BRAVO, D. y CONTRERAS, D. *Competencias y destrezas básicas de la población adulta, Chile*. Santiago: Departamento de Economía, Universidad de Chile, 2001; ARNOUX, E. *La lectura y la escritura en la universidad*. Buenos Aires: EUDEBA, 2002; y CARLINO, P. *Ibíd.*

<sup>6</sup> CARLINO, P., *Ibíd.*

## Características técnicas de CODICE–Derecho

La prueba CODICE–Derecho experimentó algunos ajustes entre las aplicaciones 2009 y 2011, especialmente, en el tratamiento de las preguntas cerradas. Las diferencias más relevantes se centraron en la cantidad de niveles de desarrollo, en la asignación de puntaje a los ítemes y en la cantidad de preguntas.

El año 2009, el instrumento estaba compuesto por cincuenta (50) preguntas de selección múltiple (cerradas) de cinco opciones y tres preguntas de desarrollo (abiertas). En el año 2011, la cantidad de preguntas cerradas se redujo a cuarenta preguntas (40). Por otra parte, los puntajes a las preguntas cerradas se asignaron el año 2009 según la dificultad estimada; en cambio, el año 2011, según competencias. En el caso de las preguntas abiertas, se utilizó una rúbrica de cinco indicadores que contemplaban un rango de puntaje de 0 a 5 puntos por cada indicador. Finalmente, mientras el año 2009 se definió 6 competencias, con seis niveles; el año 2011, se agruparon en tres competencias, manteniendo los seis niveles generales. Las preguntas abiertas en esta aplicación mantuvieron el mismo tratamiento (ver tablas 1 y 2).

**Tabla 1. Características técnicas de las preguntas de selección múltiple de las pruebas CODICE DERECHO años 2009-2011**

	<b>CODICE-Derecho 2009</b>	<b>CODICE-Derecho 2011</b>
Cantidad de preguntas cerradas	50	40
Asignación de puntaje	Por dificultad estimada: Fácil: 1 Mediana: 2 Difícil: 3	Por competencia: Competencia 1: 1 Competencia 2: 2 Competencia 3: 3
Competencias lectoras definidas	6	3
Niveles puntuación total	6	6

**Tabla 2. Características técnicas de las preguntas de selección múltiple de las pruebas CODICE DERECHO años 2009-2011**

	<b>CODICE-Derecho 2009</b>	<b>CODICE-Derecho 2011</b>
Cantidad de preguntas cerradas	50	40
Asignación de puntaje	Por dificultad estimada: Fácil: 1 Mediana: 2 Difícil: 3	Por competencia: Competencia 1: 1 Competencia 2: 2 Competencia 3: 3
Competencias lectoras definidas	6	3
Niveles puntuación total	6	6

## Resultados de la aplicación CODICE-Derecho

Los resultados que se presentan a continuación corresponden a las aplicaciones 2009 y 2011 de la prueba CODICE a los estudiantes de primer año de la carrera de Derecho de la Universidad de Chile. Para ello, se señalan inicialmente las características del grupo de participantes y posteriormente, la confiabilidad y las propiedades psicométricas de los ítems según Teoría Clásica de Test (TCT), correspondientes a los instrumentos ya señalados.

### *Participantes: características generales*

La prueba CODICE DERECHO, aplicada el año 2009, la rindieron 337 estudiantes, correspondientes al 85% del total de matriculados. De estos, el 59,9% correspondió a hombres y el 48%, provenían de dependencia “particular pagado”. El año 2011, en cambio, 242 estudiantes rindieron la prueba (62,05% del total de matriculados), de los cuales el 40% correspondió a mujeres y el 41% a la dependencia particular pagado (ver tabla 3).

**Tabla 3. Características del grupo poblacional Prueba CODICE-Derecho años 2009 y 2011**

Variables		2009			2011		
		Inscritos	Rinden	% Rinden	Inscritos	Rinden	% Rinden
GÉNERO	Masculino	244	202	82,79	240	145	60,42
	Femenino	152	135	88,82	150	97	64,67
	TOTAL	396	337	85,10	390	242	62,05
DEPENDENCIA EDUCACIONAL	Municipal	88	72	81,82	86	56	65,12
	Subvenc.	106	97	91,51	117	81	69,23
	Pagado	198	164	82,83	178	101	56,74
	Sin inform.	4	4	100,00	9	4	44,44
	TOTAL	396	337	85,10	390	242	62,05

### *Propiedades psicométricas del instrumento: la confiabilidad*

La confiabilidad<sup>7</sup> es una propiedad psicométrica fundamental en todo instrumento de medición. En este caso, dado que el instrumento fue aplicado sólo en una oportunidad, se decidió utilizar el procedimiento estadístico coeficiente Alfa de Cronbach (1951), que es un indicador de consistencia interna del test<sup>8</sup>. De esta manera, se establece la precisión con que la prueba mide la característica del individuo en un momento determinado y estima si muestras diferentes de preguntas producirían las mismas diferencias individuales.

Al observar la confiabilidad<sup>9</sup>, (las preguntas cerradas del instrumento aplicado el año 2009), presentó un valor estimado en alfa de Cronbach de 0,88 para el total de la prueba y entre 0,86 y 0,89 entre los distintos subgrupos, con una variación de error de medida entre 1,29 y 1,64 (Ver Tabla 4). El año 2011, el valor estimado de confiabilidad fue de 0,90, con una variación entre 0,89 y 0,92 entre los distintos subgrupos.

**Tabla 4. Estimación de la confiabilidad preguntas cerradas prueba CODICE-Derecho 2009, según prueba total y subgrupos**

Confiabilidad preguntas cerradas prueba CODICE-Derecho 2009						
	Grupo total	Género		Dependencia		
		Masc.	Fem.	Mu	PS	PP
Nº	337	202	135	72	97	164
Confiabilidad	0,88	0,89	0,86	0,86	0,86	0,89
Error de medida	1,44	1,31	1,64	1,29	1,52	1,34

**Tabla 4. Estimación de la confiabilidad preguntas cerradas prueba CODICE-Derecho 2011, según prueba total y subgrupos**

Confiabilidad preguntas cerradas prueba CODICE-Derecho 2011						
	Grupo total	Género		Dependencia		
		Masc.	Fem.	Mu	PS	PP
Nº	242	145	97	56	81	101
Confiabilidad	0,90	0,91	0,89	0,89	0,90	0,92
Error de medida	1,41	1,38	1,41	1,33	1,46	1,25

**Nota:** Masc=grupo masculino; Fem=grupo femenino; Mu=Municipal; PS=Particular Subvencionado; PP=Particular Pagado.

<sup>7</sup> Los Standars (1999) definen la confiabilidad como el proceso a través del cual se comprueba que las medidas ofrecidas por un test son consistentes, tras aplicar el mismo en distintas ocasiones sobre una muestra de individuos o grupos.

<sup>8</sup> MUÑIZ, J. *Teoría Clásica de los Tests*. Madrid: Pirámide, 2003.

<sup>9</sup> Para estimar la confiabilidad se excluyó al grupo de estudiantes correspondientes al grupo Sin Información de dependencia.

Por otro lado, las preguntas abiertas aplicadas el año 2009 presentaron una confiabilidad estimada de 0,88 para el grupo total y entre 0,87 y 0,88 entre los distintos subgrupos. El año 2011, para el grupo total se estimó una confiabilidad de 0,84 y entre 0,79 y 0,88 entre los distintos subgrupos.

**Tabla 6. Estimación de la confiabilidad preguntas prueba CODICE-Derecho 2009, según prueba total y subgrupos**

Confiabilidad preguntas cerradas prueba CODICE-Derecho 2011						
	Grupo total	Género		Dependencia		
		Masc.	Fem.	Mu	PS	PP
Nº	337	202	135	72	97	164
Confiabilidad	0,88	0,88	0,88	0,88	0,88	0,87
Error de medida	4,59	4,53	4,69	4,68	4,70	4,47

**Tabla 7. Estimación de la confiabilidad preguntas ABIERTAS prueba CODICE-Derecho 2011, según prueba total y subgrupos**

Confiabilidad preguntas cerradas prueba CODICE-Derecho 2011						
	Grupo total	Género		Dependencia		
		Masc.	Fem.	Mu	PS	PP
Nº	242	145	97	56	81	101
Confiabilidad	0,84	0,86	0,81	0,79	0,82	0,88
Error de medida	4,22	4,21	4,20	4,43	4,36	3,91

Nota: Masc=grupo masculino; Fem=grupo femenino; Mu=Municipal; PS=Particular Subvencionado; PP=Particular Pagado.

## Propiedades psicométricas de los ítems: dificultad y discriminación

Así como la confiabilidad es una propiedad psicométrica del instrumento, en el caso de los ítems se consideran como fundamentales la dificultad<sup>10</sup> y la discriminación<sup>11</sup>. En el instrumento aplicado el año 2009, 38 preguntas cerradas (76%) presentaron índices de dificultad y discriminación aceptables, como se observa en la tabla 8 y figura 1.

<sup>10</sup> Índice de dificultad de un ítem (ID), “la proporción de sujetos que lo aciertan de aquellos que han intentado resolverlo” (MUÑIZ, J. *Ibid.* P. 218). Este índice se considera que debe estar comprendido entre 10 y 90% (10%>p<90%). Excepcionalmente, es posible variar un poco este rango, atendiendo a los fines y propósitos de la prueba.

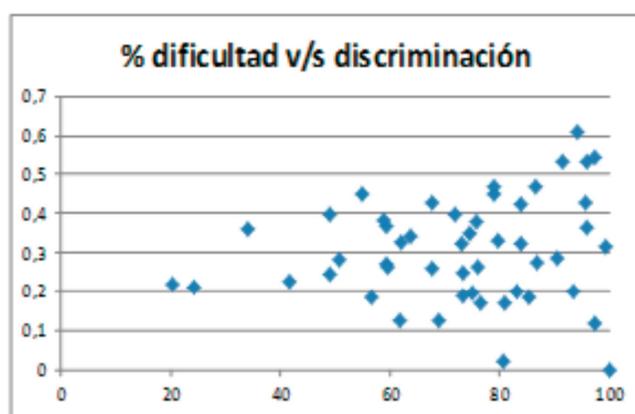
<sup>11</sup> Índice de discriminación, “se define como la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones en el test” (MUÑIZ, J. *Ibid.* P. 219). Correlación Biserial (Rb). Entre un ítem de tipo dicotómico –y por extensión, perteneciente a un nivel de medición de tipo discreto—con su escala de referencia, la cual adopta un nivel de medición de tipo continuo. La literatura especializada sugiere, para pruebas de rendimiento óptimo, contar con índices Rb mayores que 0,200 o 0,250 (Rb ≥ 0,200 o Rb ≥ 0,250).

La mayor concentración de preguntas (32) se ubicó entre el 70% y 99% de dificultad. De estas preguntas, 8 (16%) superan el 90% de dificultad, con un índice de discriminación superior a 0,200. Por otro lado, el 20% (10 ítemes) presenta índices de discriminación superiores a 0,400, concentradas en una dificultad de 70 a 90%.

**Tabla 8. Índice de Dificultad e Índice de Discriminación de los ítemes de selección múltiple prueba CODICE-Derecho año 2009.**

Dificultad y Discriminación CÓDICE-Derecho 2009							
Rb % dificultad	<0,100	0,100 – 0,199	0,200 – 0,299	0,300 – 0,399	0,400 – 0,499	0,500 – 0,599	TOTAL
0,0 - 9,9	0	0	0	0	0	0	0
10,0 – 19,9	0	0	0	0	0	0	0
20,0 – 29,9	0	0	2	0	0	0	2
30,0 – 39,9	0	0	0	1	0	0	1
40,0 – 49,9	0	0	2	1	0	0	3
50,0 – 59,9	0	1	3	2	1	0	7
60,0 – 69,9	0	2	1	2	1	0	6
70,0 – 79,9	0	3	2	5	2	0	12
80,0 – 89,9	1	2	2	1	2	0	8
90,0 – 99,9	1	2	2	2	1	3	12
<b>TOTAL</b>	<b>2</b>	<b>10</b>	<b>14</b>	<b>14</b>	<b>7</b>	<b>3</b>	<b>50</b>

**Figura 1. Gráfico % de dificultad v/s discriminación**



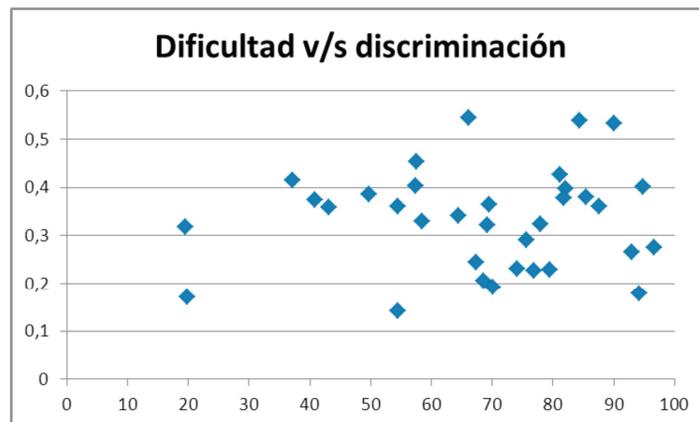
En la aplicación del año 2011, 30 preguntas (75%) presentaron índices dentro de rango, con la mayor concentración de preguntas (25), según grado de dificultad, entre 50 y 90%.

Del total de preguntas, 4 (10%) superan el 90% de dificultad, con un índice de discriminación superior a 0,200. Por otro lado, el 20% (8 ítems) presenta índices de discriminación superiores a 0,400, concentradas en una dificultad de 80 a 99%, como se observa en la Tabla 9 y Figura 2.

**Tabla 9. Índice de Dificultad e Índice de Discriminación de los ítems de selección múltiple prueba CODICE-Derecho año 2011.**

Dificultad y Discriminación CÓDICE-Derecho 2009							
Rb % dificultad	<0,100	0,100 – 0,199	0,200 – 0,299	0,300 – 0,399	0,400 – 0,499	0,500 – 0,599	TOTAL
0,0 – 9,9	0	0	0	0	0	0	0
10,0 – 19,9	0	1	0	1	0	0	2
20,0 – 29,9	0	1	0	0	0	0	1
30,0 – 39,9	0	0	0	0	1	0	1
40,0 – 49,9	0	0	0	3	0	0	3
50,0 – 59,9	0	2	0	2	2	0	6
60,0 – 69,9	0	0	2	3	0	1	5
70,0 – 79,9	0	2	4	1	0	0	7
80,0 – 89,9	0	1	0	4	1	1	6
90,0 – 99,9	2	1	2	0	1	1	6
<b>TOTAL</b>	<b>2</b>	<b>8</b>	<b>8</b>	<b>14</b>	<b>5</b>	<b>3</b>	<b>40</b>

**Figura 2. Gráfico % de dificultad v/s discriminación**



## Conclusiones

Las pruebas CODICE–Derecho, aplicadas los años 2009 y 2011, están en proceso de experimentación y se enmarcan dentro de un interés por parte de la Facultad de Derecho de la Universidad de Chile de contar con información respecto de sus estudiantes, en los ámbitos de la Lectura y la Escritura, más allá de los datos que proporciona la prueba de ingreso a la educación superior (PSU®).

Las pruebas aplicadas, desde la perspectiva de las propiedades psicométricas a nivel de instrumento se observa que las estimaciones de confiabilidad, tanto para las preguntas cerradas como para las preguntas abiertas, las que varían entre 0,86 y 0,92 y 0,79 y 0,88, respectivamente, son adecuadas para los fines del instrumento.

Por otro lado, a nivel de ítems, en ambas pruebas, una cantidad importante de sus preguntas, 76% y 75%, respectivamente, cumplen con los estándares de dificultad y discriminación exigidos a instrumentos de medición.

Dado que la prueba CODICE–Derecho está en un proceso de experimentación se siguen evaluando algunas de sus propiedades.

Por una parte, se continuará realizando una evaluación de los puntajes de corte por niveles, en función de alguna variable criterio, con el objetivo de lograr mayor precisión en la clasificación de las competencias de lectura y escritura manifestadas por los y las estudiantes.

Por otra parte, se contempla realizar un estudio de sesgo, particularmente, a través de los procedimientos estadísticos Funcionamiento Diferencial del Ítem (DIF, en inglés) en todos los subgrupos de interés como género, dependencia u otros.

Además, se evaluará la pertinencia de aplicar algunos modelos de análisis de la Teoría de Respuesta al ítem para complementar el modelo de Teoría Clásica de Test.

Por último, se procederá a efectuar las validaciones pendientes de constructo y de validez predictiva del instrumento.

## Referencias bibliográficas

- AMERICAN PSYCHOLOGICAL ASSOCIATION, AMERICAN EDUCATIONAL RESEARCH ASSOCIATION Y NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. *Standards for Educational and Psychological Testing*. Washinton DC: American Psychological Association, 1999.
- ARNOUX, E. *La lectura y la escritura en la universidad*. Buenos Aires: EUDEBA, 2002.
- ATORRESI, A. *Competencias para la vida en las evaluaciones de lectura y escritura*. (SERCE – LLECE, LABORATORIO LATINOAMERICANO DE EVALUACIÓN DE LA CALIDAD DE

- LA EDUCACIÓN), UNESCO - OREALC, 2005. Disponible en web: [http://www.oei.es/evaluacioneducativa/competencias\\_vida\\_evaluaciones\\_lectura\\_escritura\\_atorrési.pdf](http://www.oei.es/evaluacioneducativa/competencias_vida_evaluaciones_lectura_escritura_atorrési.pdf) [consultado el 15 de noviembre de 2011].
- BRAVO, D. y CONTRERAS, D. *Competencias y destrezas básicas de la población adulta, Chile*. Santiago: Departamento de Economía, Universidad de Chile, 2001.
- CARLINO, P. *Escribir, leer y aprender en la universidad. Una introducción a la alfabetización académica*. México: Fondo de Cultura Económica, 2005.
- CARLINO, Paula y MARTÍNEZ, Silvia (coords.). *La lectura y la escritura, un problema asunto de todos*. Neuquén: Universidad de Coimahuén, 2009.
- GOODMAN, Y. “Las raíces de la Alfabetización”. Revista *Infancia y aprendizaje* N° 58, 1992. Pp. 29-42.
- MUÑIZ, J. *Teoría Clásica de los Tests*. Madrid: Pirámide, 2003.
- KALMAN, J. “El acceso a la cultura escrita: la participación social y la apropiación de conocimientos en eventos cotidianos de lectura y escritura”. En: *Revista Mexicana de Investigación Educativa*. 2003, vol. VIII, núm. 17. P. 37 - 66.
- PERONARD, M., GÓMEZ MACKER, L., PARODI, G. y NÚÑEZ, P. *Comprensión de textos escritos: De la teoría a la sala de clases*. Santiago: Editorial Andrés Bello, 1998.
- PARODI, G., IBÁÑEZ, R. y PERONARD, M. *Saber leer*. Madrid: Aguilar, 2010.
- PARODI, G. (Ed.). *Alfabetización académica y profesional en el siglo XXI. Leer y escribir desde las disciplinas*. Santiago: Editorial Planeta, 2010.
- SOLÉ, I. *Estrategias de Lectura*. Barcelona: Editorial Graó, 2006.
- SPEARMAN, C. *The Abilities of Man*. Londres: McMillan, 1927.