

## ¿Cuán hábiles son los profesores para juzgar la complejidad de las preguntas de la prueba PISA y predecir el desempeño de los estudiantes chilenos en lectura?

HOW SKILLED ARE TEACHERS TO ASSESS THE COMPLEXITY OF PISA TEST QUESTIONS AND PREDICTING THE PERFORMANCE OF CHILEAN STUDENTS IN READING?

Verónica Villarroel Henríquez\*, María Angélica San Martín Toro\*\*, Carmen Nuñez Ramos\*\*\*, Carolina Hernández Opazo\*\*\*\* y Isidora Castillo Rabanal\*\*\*\*\*

Recibido: octubre 16, 2023. Aceptado: mayo 29, 2024.

**Resumen:** Se analiza el juicio evaluativo de profesores en 10 ítems de la Prueba PISA, como también su capacidad de predecir el desempeño de los estudiantes en ellos. Participaron 16 profesores de lenguaje de establecimientos educacionales públicos, subvencionados y privados. Utilizando un enfoque cuantitativo de investigación, se les aplicó un cuestionario de respuestas cerradas y abiertas que medía 5 dimensiones en cada uno de los ítems. Los profesores analizaron los 5 ítems en que los estudiantes mostraron el más alto desempeño y los 5 en que exhibieron el más bajo rendimiento. Los resultados revelaron que los profesores no pudieron predecir con exactitud el desempeño de los estudiantes en estos ítems. Para los docentes la dificultad del ítem dependía de la habilidad cognitiva evaluada. En particular, los docentes perciben que las preguntas más sencillas eran aquellas en que la información podía ser identificada de manera explícita y presentaban contextos familiares para los estudiantes. Las más difíciles requerían de habilidades cognitivas superiores. Se concluye que la familiaridad con el tipo de ítem y la evaluación situada en contextos, son variables y relevantes de considerar para explicar el desempeño de los estudiantes en esta prueba internacional.

**Palabras claves:** Competencias, Lectura, Evaluación del docente, Prueba de lectura, Enseñanza secundaria, Rendimiento escolar.

---

\* Doctora en Psicología por la Pontificia Universidad Católica de Chile. Afiliación institucional: Universidad San Sebastián. Correo electrónico: veronica.villarroel@uss.cl. Orcid: <https://orcid.org/0000-0002-3000-2248>

\*\* Magister en Psicología Educacional por la Universidad del Desarrollo. Afiliación institucional: Facultad de Psicología, Universidad del Desarrollo. Correo electrónico: angelicasanmartin@gmail.com. Orcid: <https://orcid.org/0000-0003-4715-5782>

\*\*\* Magister en Psicología Educacional por la Universidad del Desarrollo. Afiliación institucional: Facultad de Psicología, Universidad del Desarrollo. Correo electrónico: carmennunezramos@gmail.com. Orcid: <https://orcid.org/0000-0001-6223-4660>

\*\*\*\* Psicóloga clínica. Afiliación institucional: Facultad de Psicología, Universidad del Desarrollo. Correo electrónico: carolinahernandezopazo@gmail.com. Orcid: <https://orcid.org/0000-0001-8640-3334>

\*\*\*\*\* PhD Learning and Leadership Education. Afiliación institucional: Institute of Education, University College London. Correo electrónico: isidora.rabanal.21@ucl.ac.uk. Orcid: <https://orcid.org/0000-0002-9941-1437>

**Abstract:** The evaluative judgment of teachers on 10 items of the PISA test, as well as their ability to predict the performance of Chilean students on them, is analyzed. Sixteen language teachers from public, subsidized, and private educational institutions participated. Using a quantitative research approach, a questionnaire with closed and open-ended responses was administered to measure five dimensions for each item. Teachers analyzed the five items in which students exhibited the highest performance and the five in which they showed the lowest performance. The results revealed that teachers could not accurately predict student performance on these items. For teachers, item difficulty depended on the cognitive skill being assessed. In particular, teachers perceived that the easiest questions were those in which information could be explicitly identified and presented familiar contexts to students. The most difficult questions required higher cognitive skills. It is concluded that item type familiarity and the evaluation situated in contexts are relevant variables to consider in explaining student performance on this international test.

**Key words:** Skills, Reading ability, Teacher appraisal, Reading tests, High school, Student performance.

## Introducción

La competencia lectora es una habilidad fundamental para participar como ciudadano en el mundo (Cabrera-Pommiez et al., 2021). La información y el conocimiento se transmite, principalmente, a través de textos escritos (en distintos medios) que impactan en el aprendizaje y el desarrollo de la ciudadanía (Silva-Quiroz y Maturana-Castillo, 2017). Por estos motivos, la lectura facilita que todos los habitantes, de manera igualitaria, puedan acceder, comprender y reflexionar sobre todo tipo de información, de manera activa y crítica, de acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE) (Organization for Economic Co-Operation and Development [OECD], 2016).

Leer involucra interpretar, crear sentido y construir un espacio de intersubjetividad (Delgado-Cerrillo, 2007). Contribuye a comprender el mundo, como también al sí mismo, despertando la imaginación y la curiosidad (Ortiz-Salazar y Peña Castaño, 2019). Considerando su relevancia, disponer de un sistema educativo que proporcione una enseñanza de calidad en el área de la lectura constituye una aspiración universal y con ese espíritu se inició el Programa para la Evaluación Internacional de los Alumnos (PISA, según siglas en inglés), impulsado por la OCDE (Sanz Ponce et al., 2020).

PISA posee relevancia mundial en materia de evaluación educativa comparada (Díez-Mediavilla y Clemente-Egío, 2017). Nace con el objetivo de medir el nivel de preparación que tienen los estudiantes de 15 años para enfrentar desafíos de su vida futura. Es muy valorado porque entrega información sobre las competencias alcanzadas por los estudiantes. Es un programa de medición innovador y que resulta desafiante para estudiantes de 15 años. Utiliza problemas en la vida diaria, contextualizados a la realidad de los jóvenes, que deben ser resueltos aplicando los conocimientos y aprendizajes adquiridos en el aula (Caño y Luna, 2011).

Respecto a la complejidad de la evaluación PISA de lectura, se describen tres niveles. El más simple o básico implica identificar información explícita y realizar tareas de lectura simples, como localizar información en un texto. El nivel intermedio evalúa la capacidad de los estudiantes para comprender e interpretar información más compleja, incluyendo la identificación de ideas principales, detalles, inferencias sencillas y la capacidad para relacionar la información en un texto. Finalmente, en el nivel alto se evalúa la capacidad de los estudiantes para realizar tareas de lectura más sofisticadas, como la identificación y evaluación de argumentos, la interpretación de textos complejos y la realización de inferencias más elaboradas (2015).

Chile ha participado desde el año 2000 en esta prueba y, si bien, se observa un aumento sostenido en las áreas de lectura, matemáticas y ciencias, los resultados siguen evidenciando que un número importante de estudiantes no alcanza el nivel de desempeño mínimo en competencias necesarias para la vida. En todas las áreas, los estudiantes chilenos están bajo el promedio de los países de la OCDE y significativamente sobre el promedio de los países latinoamericanos (OCDE, 2010); tendencia que se repite en el informe correspondiente a PISA 2018 (OECD, 2019).

La debilidad de los estudiantes chilenos en la competencia de lectura se relaciona con los resultados presentados por Cabrera-Pommiez et al. (2021). Tras aplicar una prueba diagnóstica desde el 2017 al 2020 a 21.465 estudiantes que ingresan a la educación superior, encontraron que un grupo importante de los jóvenes no tiene un nivel satisfactorio de lectura académica y presenta dificultades para responder ítems que evaluaban lectura implícita, relaciones entre párrafos, inferencias e interpretación de propósitos comunicativos.

En un intento por comprender el desempeño de los estudiantes chilenos, (2015) estudiaron los aciertos y errores más frecuentes en los ítems de la Prueba PISA 2009, detectando que el tipo de ítem, la familiaridad de los estudiantes con ellos y la complejidad de las habilidades cognitivas involucradas en las preguntas influía en el desempeño logrado por ellos. Encontraron que los estudiantes tienden a cometer más errores en ítems de lectura que miden habilidades cognitivas de alta complejidad, como reflexionar y evaluar (2015).

En la misma línea, Valenzuela et al. (2015) analizaron las variables sistémicas e individuales que inciden en el incremento de puntajes de los estudiantes frente a la prueba PISA en el período 2000-2009, encontrando que las actitudes de los estudiantes explicaban un 25% de la mejora de los resultados, sin que las estrategias de aprendizaje se configuraran como significativas en la explicación. En otro estudio, Agasisti

et al. (2021) analizó una serie de factores escolares y su asociación con el buen rendimiento de estudiantes en desventaja socioeconómica. Utilizando datos de pruebas PISA de los años 2006, 2009, 2012 y 2015, provenientes de 56 países participantes, encontraron que los estudiantes con mejor desempeño PISA pertenecían a establecimientos con un buen clima escolar y que dedicaban más tiempo en asignaturas clave. Entonces, es posible concluir que, para comprender los resultados de los estudiantes, también es necesario explorar qué hacen los establecimientos educacionales y los docentes en el aula.

¿Por qué a los estudiantes chilenos les resulta más difícil un tipo de ítem que otro? ¿Qué aspectos del ítem son los que más influyen en su correcta o errónea resolución? Son preguntas que los propios docentes necesitan poder responder para comprender en profundidad las causas de este desempeño y poder incidir directamente en la mejora e innovación en prácticas de enseñanza y evaluación. Por ejemplo, Roberts-Sánchez et al. (2019) describen las experiencias evaluativas de docentes que se desempeñan en escuelas vulnerables de la comuna de Arica. Estos pasaron de aplicar prácticas evaluativas homogéneas, obtenidas en su formación inicial, a seleccionar sus prácticas considerando el contexto y las necesidades de sus estudiantes, valorando aquellas que les permiten tomar decisiones pedagógicas.

Las escuelas que mejoran sus resultados ponen un énfasis en cómo potenciar buenas prácticas docentes en el aula a partir de los resultados obtenidos en pruebas estandarizadas como una retroalimentación útil para las prácticas docentes (Agencia de Calidad de la Educación, 2017; Bellei et al., 2015). Esto se relaciona con lo expuesto por Dann (2017), quien destaca la necesidad de utilizar la retroalimentación para mejorar la docencia y el aprendizaje de los estudiantes, como también con lo propuesto por Castillo y Cabrerizo (2010), cuando afirman que la evaluación no termina con la calificación, sino en cómo ésta se usa para tomar decisiones sobre intervenciones futuras.

Los estudios antes expuestos se han enfocado principalmente en analizar las variables escolares e individuales que explican el rendimiento de los estudiantes en las pruebas PISA. En este estudio, en cambio, se busca profundizar en la perspectiva del profesor, específicamente en sus competencias para juzgar la complejidad de los ítems de la prueba PISA, predecir cómo sería el desempeño de los estudiantes chilenos en ellos y explorar posibles hipótesis explicativas de ese desempeño.

A partir de las conclusiones de (2015), se busca integrar la opinión de los docentes porque ellos son quienes construyen, aplican e interpretan los resultados de las evaluaciones de aula. Mientras más distancia y diferencias existan entre estas pruebas y las mediciones estandarizadas como PISA, menos probable es que mejore el desempeño de los estudiantes en estas últimas. De esta forma, la investigación tuvo por objetivos: a) determinar la habilidad de los docentes para emitir un juicio evaluativo en relación a la complejidad de ítems PISA y b) analizar la competencia de los profesores para estimar o predecir el desempeño de los alumnos en estos ítems PISA.

## Método

La investigación se desarrolló a través de un diseño no experimental de corte transversal, correspondiente a un modelo de expansión de datos cuantitativos (Onwuegbuzie y Collins, 2007). Se empleó tanto un análisis cuantitativo de datos a partir de la aplicación de un cuestionario con preguntas de respuesta cerrada y abierta. Considerando estas últimas, se realizó también un análisis de contenido de las opiniones de los docentes sobre los ítems.

## Muestra

La muestra fue de tipo no probabilística, el muestreo fue intencional o de conveniencia, y los participantes fueron elegidos siguiendo determinados criterios. Los criterios de inclusión considerados fueron: a) docentes de enseñanza media de lenguaje, y b) que durante los últimos dos años hayan realizado clases en segundo año de enseñanza media en la provincia de Concepción, Chile, y c) los establecimientos participantes debían poseer un puntaje SIMCE (Sistema de Medición de la Calidad de la Educación) dentro del promedio nacional para su dependencia. De esta forma, los colegios públicos se encontraban en un rango de puntaje entre los 240 a 250, los particulares subvencionados entre 260 a 290, y los de dependencia privada pagada sobre los 290 puntos.

Participaron 16 docentes (9 mujeres y 11 hombres) de lenguaje (6 pertenecían a la educación pública, 5 a la subvencionada y 5 a la privada). El promedio de edad de los profesores fue 44 años (D.E.= 2.4).

## Instrumento

El instrumento corresponde a un protocolo construido para medir el Juicio Evaluativo y Predictivo de los profesores que contaba de dos partes:

**a) ¿Qué se evalúa?** Se seleccionaron 10 ítems de lectura de la prueba PISA según el alto o bajo desempeño de los estudiantes chilenos en ellos. Se consideró alto desempeño cuando los estudiantes acertaron en un 70% o más en la respuesta correcta, y se consideró bajo desempeño cuando se acertaba en la respuesta correcta en 30% o menos, siguiendo la línea de investigación de (2015). Estos ítems se distribuyeron de manera variada en un documento, informando el número del ítem y la opción de respuesta correcta.

Del conjunto de 10 preguntas: 5 de ellas en que los estudiantes chilenos mostraron más bajo desempeño (ítems: 2, 4, 6, 8, 10) y las 5 en que exhibieron más alto desempeño (ítems: 1, 3, 5, 7 y 9) en PISA de lectura, como se ve en Tabla 1.

Tabla 1

Identificación de las preguntas evaluadas por los profesores

Ítem	Tipo	Lectura	
		N° del ítem del cuestionario	% de acierto en PISA por estudiantes chilenos
Alto Desempeño	SM	1	96.8
	SMC	3	41.4
	RCC	5	92.7
	RC	7	96.0
	RCA	9	48.3
Bajo Desempeño	SM	2	25.4
	SMC	4	6.70
	RCC	6	19.0
	RC	8	21.1
	RCA	10	17.2

De las 10 preguntas, 2 era de selección múltiple (SM, multiple choice), 2 eran de selección múltiple compleja (SMC, complex multiple choice), 2 eran de respuesta construida de tipo cerrada (RCC, closed constructed response), 2 eran de respuesta corta (RC, short response), 2 eran de respuesta construida de tipo abierta (RCA, open response).

Los profesores evaluaban de manera ciega, es decir, desconocían el desempeño de los estudiantes chilenos en estas preguntas de PISA que ellos estaban evaluando.

**b) ¿Cómo se evalúa?** A través de un Cuestionario de Juicio Evaluativo de cada ítem, que consideraba 5 dimensiones. Cada dimensión contó con 16 indicadores, y cinco opciones de respuesta: (1) muy en desacuerdo, (2) desacuerdo, (3) ni de acuerdo ni en desacuerdo, (4) de acuerdo, (5) muy de acuerdo. El instrumento fue validado por 5 jueces, mostrando un índice de correlación intraclase (ICC) de 0.86.

Las dimensiones evaluadas fueron:

**-Formulación del ítem.** El ítem presentaba una redacción clara, vocabulario pertinente, familiaridad de los estudiantes con el tipo de pregunta y si la información entregada era útil para resolver la situación planteada.

**-Contenido del ítem.** El contenido del ítem está presente o no en el currículum escolar, si se trata de manera principal o tangencial y si forma parte de la enseñanza formal o informal.

-**Contextualización del ítem.** Calidad de la contextualización y autenticidad del problema. Si el ítem era realista, si presentaba un problema relevante y de posible aplicación para la vida del estudiante.

-**Habilidad cognitiva medida por el ítem.** Si el ítem requería de un razonamiento simple o más elaborado, como acceder y extraer información, interpretar e integrar, reflexionar y evaluar.

-**Complejidad del ítem.** Cuán complejo resulta para los estudiantes su resolución. Se evalúa si el ítem requería de la integración de distintas variables, si necesita de planificación y división de la tarea, y si la resolución del problema es compleja.

El cuestionario presentaba, además, dos preguntas de respuesta abierta, en cada una de las 5 dimensiones para profundizar en el análisis. Asimismo, para cada ítem PISA, se pedía predecir el desempeño de los estudiantes en él, a través de preguntas cerradas con 5 opciones de respuesta.

## Procedimiento

Se entrevistó a los directores de los distintos establecimientos, solicitándoles autorización de sus profesores. El director del colegio y los profesores firmaron un consentimiento informado donde aceptaban participar y autorizaban el uso de los datos recogidos para la investigación. A cada docente, se le dio a conocer el objetivo del estudio y se le entregó el set de ítems de lectura y el Cuestionario de Juicio Evaluativo.

Para cada ítem se le pidió completar la pauta diseñada especialmente para el estudio.

La primera parte de la evaluación requería marcar en un casillero su nivel de acuerdo o desacuerdo con el indicador planteado en cada dimensión. La segunda parte consistió en realizar una valoración de los ítems anteriormente evaluados, y finalmente respondieron preguntas relacionadas con el posible desempeño de los alumnos en el ítem.

Los profesores contaban con un plazo de 30 días para entregar el cuestionario completo.

## Análisis de datos

Se realizó un análisis descriptivo de los datos según sexo, establecimiento, dependencia, tipo de prueba, evaluación de cada indicador y predicción sobre el nivel de desempeño de los estudiantes en cada ítem.

Adicionalmente, se analizó el grado de concordancia entre los profesores en relación a la evaluación de cada ítem y sus cinco dimensiones, utilizando el estadístico descriptivo Alpha de Krippendorff (Hayes y Krippendorff, 2007). Se analizó la evaluación de las cinco dimensiones a través del promedio de las evaluaciones, para conocer la dimensión mejor evaluada por ellos.

Posteriormente, se analizó la concordancia entre la predicción de los profesores sobre el desempeño de los estudiantes en los ítems analizados. Debido a que las variables no se distribuían en forma normal, se optó por usar la prueba no paramétrica de Kruskal- Wallis, para observar la existencia de diferencias.

Finalmente, se realizó un análisis de contenido sobre las atribuciones al desempeño de los estudiantes, acompañado de algunos extractos prototípicos para cada categoría, codificando con un número arábigo a cada profesor, una letra de identificación del sexo (m para mujer y h para hombre) y una letra para la dependencia administrativa (m=público, ps=particular subvencionado y pp=particular pagado). A modo de ejemplo, el docente número 1 corresponde a un hombre de un colegio particular pagado (docente 1hpp).

## Resultados

A continuación, se presentan los resultados derivados del análisis cuantitativo, seguido del análisis de contenido cualitativo.

### Descripción de los profesores sobre la complejidad de los ítems

En las Tablas 2, 3 y 4 se muestra el promedio (M) de la evaluación que realizaron los profesores a los ítems de lectura en la escala desde 1= muy en desacuerdo a 5= muy de acuerdo. Los indicadores presentados en las tablas están designados con la palabra que mejor representa la afirmación propuesta en la pauta de evaluación de ítems.

La Tabla 2 muestra que se observan diferencias significativas entre la evaluación promedio de la dimensión: formulación del ítem ( $M=4.45$ ) y las otras dimensiones ( $\text{Chi cuadrado}= 7.03, p< 0.05$ ). Es decir, en las preguntas en que los estudiantes presentan mejor desempeño, los docentes las perciben altamente bien formuladas. Por otra parte, hay diferencias significativas entre la dimensión: complejidad ( $M=2.26$ ) y el resto de las dimensiones ( $\text{Chi cuadrado}= 6.92, p< 0.05$ ), que evidencia que los docentes perciben que estos ítems son de más baja complejidad y no requieren de la integración de distintas variables o la división de la tarea para llegar a su resolución., lo que explicaría el mejor desempeño de los estudiantes en ellos. El indicador que obtuvo una mayor evaluación de los profesores fue redacción con un promedio de 4.58, de un máximo de 5.



Tabla 2

Promedio de evaluación en ítems donde los estudiantes mostraron alto desempeño

Dimensión	Indicador	Ítems Alto Desempeño (Tipo)					M
		1 (SM)	3 (SMC)	5 (RCC)	7 (RC)	9 (RCA)	
Formulación del ítem	Redacción	4.63	4.56	4.56	4.69	4.50	4.58
	Vocabulario	4.19	4.25	4.63	4.69	4.56	4.46
	Conocido	4.56	3.81	4.38	4.50	3.94	4.23
	Información	4.63	4.44	4.63	4.75	4.38	4.56
	M	4.50	4.26	4.55	4.65	4.34	4.45
Contenido	Currículum	4.63	3.50	4.25	3.94	3.94	4.05
	Frecuencia	4.56	3.44	3.88	3.81	3.38	3.81
	Aplicación	1.69	2.75	2.13	2.00	2.50	2.21
	M	3.62	3.23	3.42	3.25	3.27	3.35
Contextualización	Realista	4.31	3.50	4.13	4.38	4.13	4.09
	Importante	3.50	2.56	2.94	3.75	3.25	3.20
	Familiar	4.00	2.63	3.88	4.50	3.63	3.72
	M	3.93	2.89	3.65	4.21	3.67	3.67
Habilidad cognitiva	Reproducir	4.44	3.50	4.19	4.00	3.19	3.86
	Analizar	2.38	3.75	2.00	2.69	3.63	2.89
	Reflexionar	2.25	2.81	1.81	1.88	3.44	2.43
	M	3.02	3.35	2.66	7.31	3.42	3.06
Complejidad	Integrar	2.19	3.44	1.75	2.31	3.25	2.58
	Planificar	2.00	2.56	1.75	1.88	2.81	2.20
	Resolver	1.63	2.75	1.56	1.56	2.56	2.01
	M	1.94	2.91	1.68	1.91	2.87	2.26
	M total	3.47	3.39	3.28	3.46	3.57	

La Tabla 3, al igual que en la evaluación donde los estudiantes obtuvieron alto desempeño, muestra que se observan diferencias significativas entre la evaluación

promedio de la dimensión: formulación del ítem (M=4.30) y las otras dimensiones (Chi cuadrado= 7.27,  $p < 0.05$ ). Es decir, en las preguntas en que los estudiantes presentan más bajo desempeño, los docentes también las perciben altamente bien formuladas. Por otra parte, hay diferencias significativas entre la dimensión: complejidad (M=3.00) y el resto de las dimensiones (Chi cuadrado= 8.21,  $p < 0.05$ ), que evidencia que los docentes perciben que estos ítems son de baja complejidad y no requieren de la integración de distintas variables o la división de la tarea para llegar a su resolución.

Respecto a los indicadores que podrían dificultar el desempeño de sus estudiantes son aquellos relacionados con la dimensión habilidad cognitiva. Se observó que los indicadores reproducir y analizar alcanzaron valores sobre 3, acercándose más bien al nivel “de acuerdo”. Los profesores percibieron que para dar solución al ítem se requiere de una habilidad cognitiva compleja, como interpretar textos a través del análisis de la información presente en ellos.

Tabla 3

Promedio de evaluación en ítems donde los estudiantes mostraron bajo desempeño

Dimensión	Indicador	Ítems Bajo Desempeño (Tipo)					M
		2 (SM)	4 (SMC)	6 (RCC)	8 (RC)	10 (RCA)	
Formulación del ítem	Redacción	4.31	4.44	4.38	4.25	4.75	4.42
	Vocabulario	4.44	4.56	4.63	4.31	4.69	4.52
	Conocido	4.50	3.50	4.00	3.56	4.19	3.95
	Información	4.19	4.31	4.63	4.00	4.56	4.33
	M	4.36	4.20	4.41	4.03	4.54	4.30
Contenido	Currículum	4.50	3.94	3.94	3.75	4.25	4.07
	Frecuencia	4.44	3.69	3.63	2.69	3.88	3.66
	Aplicación	1.94	2.00	2.44	2.19	2.06	2.12
	M	3.62	3.21	3.33	2.87	3.39	3.28
Contextualización	Realista	3.25	3.75	4.25	3.19	4.31	3.75
	Importante	3.31	2.81	3.38	2.69	2.81	3.00
	Familiar	3.50	3.25	3.38	2.75	3.13	3.20
	M	3.35	3.27	3.67	2.87	3.41	3.31

Tabla 3. Promedio de evaluación en ítems donde los estudiantes mostraron bajo desempeño (Continuación)

Habilidad cognitiva	Reproducir	2.81	3.31	3.56	3.31	3.19	3.23
	Analizar	3.63	4.06	3.25	4.00	3.56	3.70
	Reflexionar	3.00	3.19	2.69	3.31	2.63	2.96
	M	3.14	3.52	3.16	3.54	3.12	3.29
Complejidad	Integrar	2.88	3.38	3.19	3.94	2.88	3.25
	Planificar	2.44	3.19	2.81	3.69	2.50	2.92
	Resolver	2.81	2.81	2.75	3.38	2.44	2.83
	M	2.71	3.12	2.91	3.67	2.60	3.00
	M total	3.50	3.51	3.56	3.44	3.49	

En la tabla 4 se comparan los promedios de la evaluación que hicieron los profesores de los ítems en que los estudiantes chilenos mostraron alto y bajo desempeño. Todas de las evaluaciones de los ítems con bajo desempeño tienden a ser más bajas que los de alto desempeño, pero no necesariamente esta diferencia es significativa.

Los indicadores en que se observan diferencias significativas son: a) contextualización familiar (M alto desempeño= 3.72; M bajo desempeño= 3.20), donde lo profesores interpretan el contexto de los ítems de más bajo desempeño como menos familiar del de los donde los estudiantes muestran mejor desempeño (Chi cuadrado= 8.11,  $p < 0.05$ ); b) habilidad cognitiva reproducir (M alto desempeño= 3.86; M bajo desempeño= 3.23), donde lo profesores interpretan el que los ítems de más bajo desempeño exigen menos habilidad de reproducción que el de los ítems con mejor desempeño (Chi cuadrado= 9.03,  $p < 0.05$ ); c) habilidad cognitiva analizar (M alto desempeño= 2.89; M bajo desempeño= 3.70), donde lo profesores interpretan el que los ítems de más bajo desempeño exigen mayor habilidad de análisis que el de los ítems con mejor desempeño (Chi cuadrado= 9.58,  $p < 0.05$ ); d) habilidad cognitiva reflexionar (M alto desempeño= 3.43; M bajo desempeño= 2.96), donde lo profesores interpretan el que los ítems de más bajo desempeño exigen mayor habilidad de reflexión que el de los ítems con mejor desempeño (Chi cuadrado= 7.89,  $p < 0.05$ ); y d) todos los indicadores de complejidad (M alto desempeño= 2.58/ 2.20/ 2.01; M bajo desempeño= 3.25/ 2.92/ 2.83), donde lo profesores interpretan el que los ítems de más bajo desempeño exigen mayor habilidad de integrar, planificar y resolver que el de los ítems con mejor desempeño (Chi cuadrado= 8.15/ 8.55/ 8.03,  $p < 0.05$ ).

**Tabla 4**

**Comparación entre la evaluación de los ítems en que los estudiantes mostraron alto y bajo desempeño**

Dimensión	Indicador	M (ítems alto desempeño)	M (ítems bajo desempeño)
Formulación del ítem	Redacción	4.58	4.42
	Vocabulario	4.46	4.52
	Conocido	4.23	3.95
	Información	4.56	4.33
	M	4.45	4.30
Contenido	Currículum	4.05	4.07
	Frecuencia	3.81	3.66
	Aplicación	2.21	2.12
	M	3.35	3.28
Contextualización	Realista	4.09	3.75
	Importante	3.20	3.00
	Familiar	3.72	3.20
	M	3.67	3.31
Habilidad cognitiva	Reproducir	3.86	3.23
	Analizar	2.89	3.70
	Reflexionar	2.43	2.96
	M	3.06	3.29
Complejidad	Integrar	2.58	3.25
	Planificar	2.20	2.92
	Resolver	2.01	2.83
	M	2.26	3.00

### **Análisis de contenido**

El análisis de contenido se realizó considerando la predicción de los profesores sobre el desempeño de los estudiantes en el área de lectura en cada uno de los ítems, su concordancia con los resultados obtenidos en PISA y las atribuciones que expresaron los profesores respecto del desempeño de los estudiantes.

El análisis se dividió considerando los ítems de alto y bajo acierto, estableciendo la presencia o ausencia de concordancia entre los profesores, junto con las razones que expusieron para ello.

Según el análisis del área de lectura, de los diez ítems evaluados, cinco correspondían a la categoría de alto desempeño y tres de ellos presentaron mayor concordancia entre los profesores, correspondientes al ítem 1, 5 y 7, donde los 16 profesores los evaluaron positivamente. Aquí, los docentes argumentaron que la información explícita en los textos presentados no requiere de una habilidad cognitiva mayor, ya que sólo debían reconocer la información; aspecto que es intencionado en el aula y el currículum. Describiéndolo así:

“La extracción de información explícita es una habilidad básica; por lo tanto, la respuesta correcta es fácil de determinar” (Docente 49mm).

En cuanto a los ítems 3 y 9 de alto acierto, no hubo concordancia. Quienes evaluaron estos ítems como de alto acierto, fundamentaron su decisión en la información explícita y la familiaridad del ítem por su utilización en las evaluaciones, como se refleja en el siguiente comentario:

“Es un ítem muy utilizado dentro de las evaluaciones, por lo cual es fácil para ellos desarrollarla” (Docente 47mm).

Por su parte, los profesores que consideraron los ítems de bajo desempeño, argumentaron que se requiere de habilidades complejas para responder, lo cual condiría al alumno, ya que este tipo de ítem no es trabajado en aula frecuentemente:

“La habilidad de interpretación de texto multimodales es débil en los estudiantes” (Docente 50mm).

Cabe destacar que en ítems de bajo desempeño hay menor concordancia entre la opinión de los docentes y los resultados de los estudiantes.

“La principal razón puede basarse en que, para poder responder el ítem, sólo los alumnos debían localizar la información que se presentaba de manera explícita en el texto” (Docente 6mmp).

Los profesores que valoraron los ítems como de bajo desempeño, plantearon que la dificultad estaría dada por la exigencia de una habilidad cognitiva superior, señalando que:

“Falta de la habilidad para deducir o inferir información, el alumno está acostumbrado a lo explícito o textual” (Docente 49mm).

Un grupo de docentes, planteó que en el ítem 8 los estudiantes obtendrían buen desempeño, aludiendo que en el texto se presentó información explícita. Los demás profesores, evaluaron el ítem como de alta dificultad.

“Es más complejo lo que se le pide, porque es una habilidad que requiere de práctica, pues se debe leer cada texto con atención y selecciona lo que está en todos, en algunos o sólo en uno” (Juez 3mmp).

“La pregunta requiere el conocimiento de una leyenda original que supone el conocimiento por parte de los estudiantes” (Docente 48mm).

## Discusión

El estudio permitió conocer el juicio evaluativo de profesores de lenguaje sobre la dificultad de los ítems en la prueba PISA de lectura, y la concordancia de sus opiniones con el desempeño mostrado por los estudiantes.

Respecto a los ítems, se concluye que, en general, los docentes no pudieron predecir con exactitud, si los ítems de la prueba que estaban analizando, habían mostrado un bajo o alto acierto en el desempeño de los estudiantes chilenos. Considerando esta dificultad, es necesario destacar que los profesores alcanzan mayor capacidad de predicción en los ítems donde los estudiantes presentan mejor desempeño.

Esta dificultad para reconocer los ítems, donde los estudiantes muestran más bajo desempeño, podría incidir dificultades para la implementación de innovaciones educativas orientadas a fortalecer las competencias evaluadas en las distintas áreas. En otras palabras, se les haría más difícil detectar en qué enfocarse para planificar prácticas remediales futuras (Castillo y Cabrerizo, 2010), en especial en aquellas habilidades que a los estudiantes les cuesta más lograr.

Cuando se indaga respecto a lo que podría facilitar un mejor o peor desempeño en los estudiantes, los docentes concluyen que, en general, la formulación de los ítems favorece el desempeño de los estudiantes, ya que éstos contaban con una redacción clara, un vocablo familiar y comprensible para los estudiantes. En cambio, plantean que, los factores que dificultaría la correcta resolución del ítem sería la complejidad, contextualización y la habilidad cognitiva medida.

Las evaluaciones que entregaron los profesores sobre el desempeño de los estudiantes fueron variadas frente a un mismo ítem. Sin embargo, concordaron en que los ítems de menor dificultad fueron aquellos con información explícita y familiar a lo tratado por el estudiante, involucrando predominantemente habilidades cognitivas básicas; mientras que los de mayor dificultad fueron aquellos donde la información no se entrega de forma textual, suponiendo el empleo de habilidades cognitivas superiores para interpretarla. Estas atribuciones serían consistentes con los hallazgos derivados del estudio de 2015.

Si bien los docentes detectaron que la dificultad del ítem radica en la habilidad cognitiva evaluada, no lograron reconocer a cabalidad el nivel de habilidad medida. Es posible hipotetizar que la dificultad para reconocer las habilidades cognitivas medidas por los ítems responda a un enfoque funcional del aprendizaje, visualizando el desafío de diseñar y proponer situaciones problemáticas realistas, donde se incorporen transversalmente los diferentes niveles de complejidad cognitiva, en línea con los resultados de la investigación de Radišić<sup>1</sup> y Baucal (2018).

Todo lo anterior evidencia la necesidad de familiarizar a los estudiantes con diferentes tipos de ítems, realistas y contextualizados, y con diferentes niveles de comple-

alidad, en las evaluaciones de aula. Por ejemplo, comenzar a integrar en las pruebas, ítems típicos de PISA, que son más extensos, con contextos que incluyen textos en distinto formato, como ocurre en una revista o periódico (por ejemplo, con entrevista, gráficos, tablas, reportes), que apelen a situaciones de la vida cotidiana, y que los estudiantes deban comprender e interpretar.

Respecto a las limitaciones del estudio, se puede señalar el análisis sólo de ítems de mayor y menor acierto, quedando fuera una porción de preguntas frente a las que los estudiantes reflejan un desempeño promedio, y que posiblemente se vinculen a aquellas más típicamente puestas en práctica por los docentes, tanto en términos de contenido, de estructura de la pregunta y de la habilidad cognitiva exigida. Por otra parte, la escasez de estudios a nivel nacional sobre esta temática, perfilan estos resultados como fundamentalmente exploratorios, aun cuando en términos analíticos revisten un alcance descriptivo-comparativo.

Adicionalmente, no se tomó en cuenta la creciente necesidad de abordar las nuevas necesidades lectoras que derivan del desarrollo y masificación de las nuevas tecnologías de la información (Sánchez-Claros, 2016). Considerando que esta prueba busca presentar problemas contextualizados para los estudiantes (Caño y Luna, 2011), los establecimientos educacionales debiesen considerar este nuevo contexto para así fomentar el desarrollo de competencias digitales (Consejo de la Unión Europea, 2016, 30 de mayo). Trabajos futuros deberían considerar este nuevo contexto al investigar las competencias lectoras del estudiantado.

Las fortalezas del estudio están relacionadas con emplear un diseño mixto de investigación, que ha permitido profundizar en las valoraciones y atribuciones de los docentes respecto de los ítems y el desempeño de los estudiantes, entregando información valiosa para la mejora de prácticas educativas relacionadas a la competencia lectora dentro del aula. Si bien la muestra puede considerarse pequeña, participaron docentes de establecimientos educativos chilenos de todo tipo de dependencia, buscando representar los distintos contextos que pueden ser encontrados en el país.

La literatura reconoce que las prácticas educativas son inseparables de las prácticas pedagógicas. Adhiriendo al consenso sobre la importancia de la evaluación en los procesos de enseñanza y aprendizaje, este estudio orienta a los docentes a reflexionar sobre cómo sus acciones han resultado eficaces para favorecer el aprendizaje de los estudiantes en el área de lectura. Gutiérrez-Fresneda y del Olmo Ibáñez (2019) sugieren que la comprensión lectora de los estudiantes puede ser mejorada en el aula a través del desarrollo de estrategias en base a preguntas tipo test, como puede ser considerada la prueba PISA. Sin embargo, para que estas actividades tengan éxito, es necesario que los docentes puedan mejorar su juicio evaluativo ante ítems de comprensión lectora, para determinar qué hace que algunos de ellos se vuelvan más complejos de resolver para sus estudiantes.

Este estudio puede aportar al fomento de una cultura evaluativa, que considere el error como una oportunidad más de aprendizaje para estudiantes y docentes, sumando la importancia de la participación de los docentes como protagonistas en la implementación de mejoras educativas.

Se concluye que es necesario avanzar en la investigación sobre evaluación educativa de modo de reforzar las competencias de los docentes en ejercicio en el área de la evaluación y el juicio evaluativo. Asimismo, fortalecer la formación inicial docente en evaluación para el aprendizaje.

Finalmente, se sugiere replicar esta metodología en el análisis del desempeño frente a otras evaluaciones estandarizadas, de carácter nacional e internacional, con otras poblaciones de estudio.

## Referencias

- Agencia de Calidad de la Educación. (2017). Informe de Resultados PISA 2015. Competencia científica, lectora y matemática en estudiantes de quince años en Chile. Agencia de Calidad de la Educación. <http://archivos.agenciaeducacion.cl/>
- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2021). What school factors are associated with the success of socio-economically disadvantaged students? An empirical investigation using PISA data. *Social Indicators Research*, 157, 749–781. <https://doi.org/10.1007/s11205-021-02668-w>
- Bellei, C., Vanni, X., Valenzuela, J. P., y Contreras, D. (2015). Una mirada a los procesos de mejoramiento escolar en Chile. Apuntes sobre Mejoramiento Escolar N°2, Enero 2015. Anillo de Ciencias Sociales sobre Mejoramiento de la Efectividad Escolar en Chile. [https://www.researchgate.net/profile/Xavier-Vanni/publication/319243286\\_Una\\_mirada\\_a\\_los\\_procesos\\_de\\_mejoramiento\\_escolar\\_en\\_Chile\\_Apuntes\\_sobre\\_Mejoramiento\\_Escolar\\_N2\\_Enero\\_2015\\_Anillo\\_de\\_Ciencias\\_Sociales\\_sobre\\_Mejoramiento\\_de\\_la\\_Efectividad\\_Escolar\\_en\\_Chile/links/599d78d4a6fdcc5003507685/Una-mirada-a-los-procesos-de-mejoramiento-escolar-en-Chile-Apuntes-sobre-Mejoramiento-Escolar-N2-Enero-2015-Anillo-de-Ciencias-Sociales-sobre-Mejoramiento-de-la-Efectividad-Escolar-en-Chile.pdf](https://www.researchgate.net/profile/Xavier-Vanni/publication/319243286_Una_mirada_a_los_procesos_de_mejoramiento_escolar_en_Chile_Apuntes_sobre_Mejoramiento_Escolar_N2_Enero_2015_Anillo_de_Ciencias_Sociales_sobre_Mejoramiento_de_la_Efectividad_Escolar_en_Chile/links/599d78d4a6fdcc5003507685/Una-mirada-a-los-procesos-de-mejoramiento-escolar-en-Chile-Apuntes-sobre-Mejoramiento-Escolar-N2-Enero-2015-Anillo-de-Ciencias-Sociales-sobre-Mejoramiento-de-la-Efectividad-Escolar-en-Chile.pdf)
- Cabrera-Pommiez, M., Lara-Inostroza, F., & Puga-Larraín, J. (2021). Evaluación de la lectura académica en estudiantes que ingresan a la Educación Superior. *Ocnos. Revista de Estudios Sobre Lectura*, 20(3), 1-18. [https://doi.org/10.18239/ocnos\\_2021.20.3.2614](https://doi.org/10.18239/ocnos_2021.20.3.2614)
- Caño, A. y Luna, F. (2011). PISA - Comprensión lectora: I Marco y análisis de los ítems. Instituto Vasco de Evaluación e Investigación Educativa. <https://docplayer.es/6689474-Pisa-comprension-lectora-i-marco-y-analisis-de-los-tems-proyecto-de-evaluacion-internacional-del-alumnado-de-15-anos.html>
- Castillo, S. y Cabrerizo, J. (2010). Evaluación educativa de aprendizajes y competencias. Pearson Prentice Hall.
- Consejo de la Unión Europea (2016, 30 de mayo). Conclusiones del Consejo de la Unión Europea sobre el desarrollo de la alfabetización mediática y el pensamiento crítico a través de la educación y la formación. Consejo de la Unión Europea. <https://www.consilium.europa.eu/es/press/press-releases/2016/05/30/eyps-conclusions-developing-media-literacy/>
- Dann, R. (2017). Developing feedback for pupil learning. *Teaching, learning, and assessment in schools*. Routledge. <https://doi-org.libproxy.ucl.ac.uk/10.4324/9781315564210>



- Delgado Cerrillo, B. (2007). Fundamentos del proceso lector. Motivar la lectura en la Educación Secundaria. *Ocnos. Revista de Estudios Sobre Lectura*, (3), 39-53. [https://doi.org/10.18239/ocnos\\_2007.03.03](https://doi.org/10.18239/ocnos_2007.03.03)
- Díez Mediavilla, A., y Clemente Egío, V. (2017). La competencia lectora. Una aproximación teórica y práctica para su evaluación. *Investigaciones sobre lectura*, 7, 22-35.
- Gutiérrez-Fresneda, R. y Del Olmo, M. T. (2019). Mejora de la comprensión lectora mediante la formulación de preguntas tipo test. *Revista Investigaciones sobre lectura*, 11, 93-104. <https://doi.org/10.37132/isl.v0i11.286>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1, 77-89. <http://www.afhayes.com/public/kalpha.pdf>
- Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2), 281-316. <https://files.eric.ed.gov/fulltext/EJ800183.pdf>
- Organization for Economic Co-Operation and Development. (2019). PISA 2018 Results (Volume I): What Students Know and Can Do. Organization for Economic Co-Operation and Development. <https://doi.org/10.1787/5f07c754-en>.
- Organization for Economic Co-Operation and Development. (2016). Skills matter: Further results from the Survey of Adult Skills. Organization for Economic Co-Operation and Development. <http://dx.doi.org/10.1787/9789264258051-en>
- Organización para la Cooperación y el Desarrollo Económico. (2010). Resultados del informe PISA 2009: Tendencias de aprendizaje. Cambios en el rendimiento de los estudiantes desde 2000 (volumen V). <https://www.oecd-ilibrary.org/docserver/9789264177543-es.pdf?expires=1654302327&id=id&accname=guest&checksum=6F5C14D12A0B8E8DAD80B18DD-7053FE0>
- Ortiz-Salazar, M. A., & Peña Castaño, J. M. (2019). La lectura en la infancia y niñez: Incidencia en la construcción del sujeto lector. *Sophía*, 15(2), 111-117.
- Radišić1, J. & Baucal, A. (2018). Teachers' reflection on PISA items and why they are so hard for students in Serbia. *European Journal of Psychology of Education*, 33, 445-466. <https://doi.org/10.1007/s10212-018-0366-0>
- Roberts-Sánchez, K., Rodríguez-Gómez, D., y Silva, P. (2019). Saber pedagógico en evaluación del profesorado de escuelas vulnerables de la comuna de Arica y sus demandas a la formación inicial docente. *Paideia*, (64), 65-95. <https://revistas.udec.cl/index.php/paideia/article/view/2135>
- Sánchez-Claros, J. P. (2016). Nuevas alfabetizaciones en un entorno multimodal: Nuevas necesidades lectoras para un entorno textual múltiple. *Investigaciones sobre lectura*, 6, 51-57. <https://www.redalyc.org/pdf/4462/446246761004.pdf>
- Sanz Ponce, R., Serrano Sarmiento, Á., & González Bertolín, A. (2020). PISA: El precio pedagógico de una evaluación internacional. *Revista electrónica de investigación educativa*, 22, e22, 1-13. <https://doi.org/10.24320/redie.2020.22.e22.2673>

Silva-Quiroz, Juan, & Maturana-Castillo, Daniela (2017). Una propuesta de modelo para introducir metodologías activas en educación superior. *Innovación educativa* (México, DF), 17(73), 117-131. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-26732017000100117&lng=es&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-26732017000100117&lng=es&tlng=es).

Valenzuela, J.P., Gómez, G. Sotomayor, C. (2015). The role of reading engagement in improving national achievement: An analysis of Chile's 2000–2009 PISA results. *International Journal of Educational Development*, 40, 28-39. <https://doi.org/10.1016/j.ijedudev.2014.11.011>