

Análisis de validez basado en procesos de respuesta de un instrumento para evaluar pensamiento funcional^{1,2}

Paulina Araya³

Orcid: 0000-0001-6629-6906

Beltrán Pantoja⁴

Orcid: 0009-0004-6407-8859

Andrea Valenzuela⁵

Orcid: 0000-0003-4948-8564

Resumen

Diversos estudios desde la perspectiva del early algebra han empleado pruebas escritas para evaluar el pensamiento funcional de los estudiantes. Sin embargo, la validez de las interpretaciones derivadas de estos instrumentos ha sido escasamente investigada. Este estudio tuvo como propósito validar las puntuaciones de un instrumento diseñado para evaluar el pensamiento funcional en estudiantes de tercer grado, aportando evidencia basada en los procesos de respuesta. Se administró una prueba y una entrevista cognitiva retrospectiva a 24 estudiantes de tercer grado de una escuela de Santiago de Chile. Las respuestas obtenidas en la prueba y la entrevista fueron codificadas y comparadas para examinar la concordancia entre las interpretaciones derivadas de ambos instrumentos. Los resultados mostraron una correlación alta y significativa en el desempeño general, y revelaron que el ítem con mayor discrepancia entre la prueba y la entrevista implicaba describir la relación entre variables utilizando lenguaje natural. En conclusión, el instrumento permite capturar de manera fiable el pensamiento funcional de los estudiantes, aunque podría subestimar su desempeño en la expresión de reglas funcionales mediante lenguaje natural. Futuros estudios podrían emplear este instrumento, considerando la inclusión de entrevistas breves para los estudiantes en la categoría funcional numérico, con el fin de captar de manera más precisa su capacidad de generalización en lenguaje natural sin depender de sus habilidades de redacción.

1- Disponibilidad de datos: El conjunto de datos que respalda los resultados de este estudio están disponibles en Harvard Dataverse y se puede acceder a ellos en <https://doi.org/10.7910/DVN/4JALQ7>

2- Agradecimientos: Esta investigación se realizó en el marco del proyecto FONDECYT de postdoctorado No. 3220465, financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) de Chile y el Núcleo Milenio para el Estudio del Desarrollo Temprano de Habilidades Matemáticas (MEMAT).

3- Universidad Diego Portales, Santiago, Chile. Contacto: paulina.araya7@mail.udp.cl

4- Universidad de O'Higgins, Rancagua, Chile. Contacto: beltran.pantoja@uoh.cl

5- Pontificia Universidad Católica de Chile, Santiago, Chile. Contacto: apvalenzuela@uc.cl



<https://doi.org/10.1590/S1678-4634202551287163es>
This content is licensed under a Creative Commons attribution-type BY 4.0.



Palabras clave

Evaluación – Evidencia de validez – Pensamiento algebraico temprano – Pensamiento funcional – Educación primaria.

*Validity evidence based on response processes for an instrument assessing functional thinking**

Abstract

Various studies from the perspective of early algebra have employed written tests to assess students' functional thinking. However, there has been little research on the validity of the interpretations derived from these instruments. The purpose of this study was to validate the scores of an instrument designed to assess functional thinking in third-grade students by providing evidence based on response processes. A test and a retrospective cognitive interview were administered to 24 third-grade students from a school in Santiago, Chile. Responses from both the test and the interview were coded and compared to assess the agreement between the interpretations derived from each instrument. The results showed a strong and statistically significant correlation in overall performance and revealed that the item with the greatest divergence between the test and the interview involved describing the relationship between variables using natural language. In conclusion, the instrument offers a reliable means of capturing students' functional thinking, although it may underestimate their performance when expressing functional rules using natural language. Future research could use this instrument while incorporating brief interviews for students in the numerical-functional category, in order to more accurately capture their ability to generalize in natural language without relying on their writing skills.

Keywords

Assessment – Validity evidence – Early algebraic thinking – Functional thinking – Primary education

Introducción

En la actualidad, diversos planes curriculares han incorporado la enseñanza del álgebra como un componente transversal desde el inicio de la escolaridad (Kieran, 2022; Blanton *et al.*, 2019; NCTM, 2000). Estas propuestas recomiendan que los estudiantes comiencen identificando relaciones y estructuras matemáticas generales a través de situaciones adecuadas para su edad, lo que establece las bases para la comprensión progresiva de conceptos algebraicos avanzados, como la notación variable (Blanton *et al.*, 2019; Mason *et*

al., 2009). De esta forma, se espera que un desarrollo temprano del pensamiento algebraico contribuya a mitigar los bajos desempeños matemáticos asociados con la enseñanza tardía del álgebra (Kaput, 2008; Moses y Cobb, 2001; Museus *et al.*, 2013).

Una de las áreas del pensamiento algebraico es el pensamiento funcional, que implica la generalización de relaciones de covariación, así como la representación y justificación de estas relaciones a partir de distintos registros (Brizuela *et al.*, 2015). Diversos estudios, desde esta perspectiva, han mostrado consistentemente que, desde los primeros grados escolares, los estudiantes tienen la capacidad de analizar cantidades que covarían y gradualmente desarrollan la habilidad de generalizar dichas relaciones utilizando registros más o menos complejos, hasta alcanzar la expresión de sus generalizaciones a través del lenguaje simbólico convencional (Kieran, 2022; Radford, 2018).

En esta línea, distintas investigaciones han examinado los cambios en el pensamiento funcional de los estudiantes antes y después de la implementación de intervenciones educativas, (Blanton *et al.*, 2019; Chimoni *et al.*, 2021; Ureña *et al.*, 2019). Un aspecto metodológico central de estos estudios radica en la forma en que se evalúa el aprendizaje de los estudiantes. Para ello, se han utilizado principalmente dos formas de evaluación: instrumentos escritos (pruebas) y entrevistas. Los estudios con un enfoque cualitativo han optado mayormente por el uso de entrevistas (Blanton *et al.*, 2017; Brizuela *et al.*, 2015; Goñi-Cervera *et al.*, 2022; Ureña *et al.*, 2019), debido a que ofrecen la ventaja de permitir que los estudiantes expresen sus ideas sin depender de sus habilidades de lectura y escritura (Larsson; Granhag, 2005), aunque requieren un tiempo considerable tanto para su aplicación como para su análisis.

Por otro lado, las investigaciones que evalúan el pensamiento funcional con un enfoque cuantitativo (Blanton *et al.*, 2015, 2019; Chimoni *et al.*, 2021; Stylianou *et al.*, 2019) han empleado predominantemente pruebas de lápiz y papel. Este tipo de instrumento ofrece la ventaja de evaluar a una mayor cantidad de estudiantes en un tiempo reducido, lo que resulta especialmente relevante en estudios con muestras numerosas. Es importante destacar que, en general, los instrumentos de evaluación utilizados han sido desarrollados por los propios investigadores, sin un análisis exhaustivo de su validez para medir el constructo deseado.

Un aspecto clave en la construcción de instrumentos de evaluación es llevar a cabo procesos de recolección de evidencia de validez, es decir, evidencia que respalde las interpretaciones derivadas del instrumento (AERA; APA; NCME, 2014). Una de las fuentes de evidencia es denominada evidencia basada en los procesos de respuesta, la que consiste en analizar el grado de concordancia entre las suposiciones de los desarrolladores del instrumento sobre cómo los evaluados responderán y los razonamientos que efectivamente experimentan estos últimos (AERA; APA; NCME, 2014; Willson; Miller, 2014). A pesar de su relevancia, esta fuente ha sido explorada con menor frecuencia en los estudios de validación (Padilla; Leighton, 2017). Sin embargo, su comprensión es esencial para garantizar la validez de las evaluaciones cognitivas.

Con el propósito de aportar evidencia de validez basada en los procesos de respuesta, este estudio tuvo como objetivo analizar en qué medida una prueba que evalúa el pensamiento funcional logra capturar este tipo de razonamiento en estudiantes de 3°



grado. De este modo, las preguntas de investigación abordadas en este trabajo son: 1) ¿en qué medida el pensamiento funcional de estudiantes de 3° grado capturado por una prueba es concordante con el identificado por medio de una entrevista cognitiva?, y 2) ¿en qué ítemes se observan mayores discrepancias entre las conclusiones derivadas a partir de la prueba y de la entrevista?

Pensamiento algebraico temprano

El pensamiento algebraico temprano es el razonamiento en el que participan niños de 5 a 12 años a medida que van otorgando significados a los objetos y las formas de pensar que encontrarán más tarde en el álgebra de la escuela secundaria (Kieran, 2022). Distintos estudios han mostrado que los estudiantes de estas edades son capaces de realizar exitosamente actividades centrales del álgebra, tales como identificar variables, ordenar datos en tablas, generalizar estructuras y expresar dicha generalidad utilizando diferentes medios de representación, como números, lenguaje natural, o símbolos convencionales y no convencionales (Brizuela *et al.*, 2015; Cañadas *et al.*, 2019). Radford (2018) señala que el pensamiento algebraico se caracteriza por la capacidad de tratar cantidades indeterminadas (variables o incógnitas) de forma analítica. En este sentido, el uso del simbolismo alfanumérico no es condición necesaria ni suficiente para pensar algebraicamente, ya que otros sistemas semióticos podrían dar cuenta de que un estudiante comprende la estructura de una secuencia y la trata de manera general y analítica (Cañadas *et al.*, 2019; Kieran, 2022; Radford, 2018).

Dentro de la investigación sobre el álgebra temprana, se distinguen tres grandes focos según los ámbitos involucrados: (i) igualdades y ecuaciones, (ii) aritmética generalizada y (iii) pensamiento funcional (Blanton *et al.*, 2015). El presente estudio se centra en el pensamiento funcional, el que se caracteriza por analizar y generalizar relaciones entre cantidades covariantes, y expresar estas relaciones a partir de distintas representaciones, como tablas, palabras, símbolos, entre otras (Blanton; Kaput, 2011). No se trata de un abordaje formal del concepto de función, sino del trabajo con tareas contextualizadas que involucran cantidades que covarían y que pueden modelarse mediante funciones. Es decir, tareas en las que los estudiantes deban generalizar un patrón y representar esta generalidad de formas distintas (Kieran, 2022; Pinto; Cañadas, 2021).

Niveles de sofisticación del pensamiento algebraico

El dominio del lenguaje algebraico es un proceso gradual que implica la incorporación progresiva de medios de representación más sofisticados para expresar generalidades matemáticas (Blanton *et al.*, 2015; Radford, 2018; Stephens *et al.*, 2017). Diversos autores han caracterizado los tipos de razonamiento que los estudiantes emplean al analizar patrones, organizándolos según su nivel de complejidad y sofisticación. Radford (2018) identifica tres tipos de pensamiento algebraico que surgen en diferentes etapas de la trayectoria escolar: el pensamiento factual, contextual y simbólico. El pensamiento factual se caracteriza por la comprensión de las estructuras operacionales de una secuencia y su aplicación a cualquier término, aunque sin explicitar cantidades indeterminadas. El

pensamiento contextual permite a los estudiantes verbalizar estructuras operacionales de manera generalizada, explicitando lo indeterminado mediante palabras. Finalmente, el pensamiento simbólico supone la capacidad de expresar estructuras operacionales utilizando símbolos alfanuméricos convencionales.

Por otro lado, Blanton y Kaput (2011) identificaron tres tipos de razonamiento en el análisis de patrones, que van desde el más simple al más complejo: (i) el pensamiento recursivo consiste en identificar la variación dentro de una sola secuencia; (ii) el pensamiento covariacional consiste en analizar cómo dos cantidades varían simultáneamente, pero sin identificar una relación directa y (iii) el pensamiento de correspondencia, que implica identificar una correlación directa entre las variables. Posteriormente, Stephens *et al.* (2017) ampliaron el esquema de Blanton y Kaput (2011) a partir de una clasificación de las respuestas de los estudiantes según el tipo de representación utilizada y la completitud de la generalización. Estos autores organizan las respuestas por nivel de dificultad en función del tipo de representación empleada –numérica, simbólica y mediante lenguaje natural– y distinguen, además, si la respuesta explicita solo una de las variables o ambas.

Es importante notar que el esquema de Stephens *et al.* (2017), basado en datos de pruebas escritas, considera que representar relaciones funcionales mediante lenguaje natural resulta más complejo que hacerlo mediante representación simbólica. No obstante, la cuestión de si el lenguaje natural o el simbólico es más accesible para los estudiantes ha mostrado resultados contradictorios (Kieran, 2022). A pesar de estas discrepancias, ambos modelos coinciden en que los estudiantes, al trabajar con secuencias que covarian, suelen comenzar con enfoques recursivos y de covariación. Luego, un cambio significativo hacia la generalización algebraica ocurre cuando emplean estrategias de correspondencia, representadas inicialmente a través de expresiones numéricas y progresivamente mediante formas de representación más complejas como el lenguaje natural y el simbólico. Por ende, una cuestión central en la evaluación del pensamiento funcional es determinar si los estudiantes son capaces, en primera instancia, de identificar las variables, organizar los datos en tablas e identificar patrones de recursividad y covariación, y, luego, observar si reconocen la correspondencia generalizando la situación a través de registros numéricos, verbales y simbólicos.

Evidencia de validez basada en los procesos de respuesta

La validez es un argumento basado en evidencia empírica y teórica que respalda la interpretación de los puntajes de una prueba diseñada para un uso en particular (AERA; APA; NCME, 2014). Las evidencias que podrían utilizarse para evaluar la validez de las puntuaciones de una prueba provienen de diferentes fuentes, por ejemplo, de su contenido, de la relación con otras variables o del proceso de respuesta, entre otras. Cada una de estas evidencias se unen para formar un argumento respecto a si es válida la interpretación del puntaje para su uso declarado. De ese modo, contar con evidencias de validez es fundamental para decidir respecto del uso de una prueba en particular.

Una de las fuentes de evidencias de validez se relaciona con el proceso de respuesta. Esta fuente de evidencia busca respaldar el adecuado cruce entre el constructo que se



evalúa y la naturaleza de la respuesta entregada por quien responde la prueba. Por ejemplo, si una prueba tiene como finalidad evaluar el razonamiento matemático, es importante asegurar que quien responde efectivamente está aplicando dicho razonamiento, y no algún procedimiento que haya conducido al resultado de manera fortuita (AERA; APA; NCME, 2014). Por lo general, esta evidencia proviene del análisis de respuestas individuales que son recolectadas mediante entrevistas cognitivas.

La entrevista cognitiva es una técnica que permite acceder a los procesos mentales de los evaluados cuando responden preguntas de una prueba. Este tipo de entrevistas recogen información verbal adicional sobre la respuesta de un evaluado a una pregunta. De esta manera, la información levantada ayudará a juzgar si la pregunta está generando el proceso mental que el diseñador del ítem pretende (Padilla; Benítez, 2014).

Entre las técnicas para implementar una entrevista cognitiva se encuentran el pensamiento en voz alta y el diseño retrospectivo (Caicedo; Zalazar-Jaime, 2018; Willis, 2019). En la técnica de pensamiento en voz alta, los evaluados verbalizan sus pensamientos a medida que se enfrentan a los ítems. En cambio, en el diseño retrospectivo, los evaluados responden a los ítems de la prueba en condiciones similares a las de su futura aplicación y, luego, comienzan la entrevista cognitiva señalando los razonamientos que emplearon para responder. Las ventajas y desventajas de cada técnica han sido estudiadas en la literatura (Conrad; Blair, 2009). Una ventaja de la entrevista retrospectiva es que los evaluados responden a la prueba sin verse expuestos por el entrevistador ni desconcentrarse al verbalizar sus pensamientos, lo que puede ser un elemento importante en preguntas complejas, en las que el proceso puede involucrar varios pasos (Meadows, 2021).

Metodología

Como una manera de aportar evidencia de validez basada en los procesos de respuesta, este artículo tuvo como propósito analizar en qué medida una prueba que evalúa el pensamiento funcional logra capturar este tipo de razonamiento en estudiantes de 3° grado. Para ello, se analizó el punto en que las interpretaciones realizadas a partir de la prueba coincidían con el razonamiento descrito por los estudiantes en una entrevista cognitiva retrospectiva (Caicedo; Zalazar-Jaime, 2018). De esta manera, se codificaron las respuestas escritas y las de la entrevista y se aplicaron técnicas cuantitativas para identificar el grado de concordancia entre ambas medidas.

Respecto a los aspectos éticos de la investigación, los estudiantes, sus apoderados y el director del establecimiento consintieron voluntariamente su participación en este estudio, y tanto los protocolos seguidos como los consentimientos informados fueron revisados y aprobados por un comité de ética.

Participantes

Participaron 24 alumnos de 3° grado de una escuela pública de Santiago de Chile. Estos alumnos habían participado de una intervención de 6 semanas de duración, cuyo

propósito fue desarrollar el pensamiento funcional. Una vez a la semana, los estudiantes trabajaron en una situación funcional, empleando distintas formas de generalización incluida la notación variable. En las sesiones, los estudiantes no estudiaron ni el contexto ni la función abordada en la prueba. La elección de los estudiantes se realizó mediante un muestreo no probabilístico intencional (Castillo; Padilla, 2012) con el propósito de incluir estudiantes tanto de género femenino como masculino y que tuvieran distintos desempeños. Este número de participantes (N=24) es adecuado para levantar evidencias de validez basada en los procesos de respuesta (Padilla; Benítez, 2014; Willis, 2019).

Instrumentos

La prueba se elaboró en torno a uno de los contextos empleados por Blanton *et al.* (2015) (Figura 1). Todos los ítemes de la prueba se desprenden de este contexto. Se elaboraron 4 grupos de ítemes. El primer grupo evaluó habilidades aritméticas, específicamente, responder casos cercanos y ordenarlos en una tabla. El segundo grupo incluyó preguntas que requerían generalizar la relación funcional de forma numérica, es decir, hallar un resultado numérico para un término lejano de la secuencia. El tercer grupo tuvo solo una pregunta que incluía describir la relación entre las variables usando lenguaje natural. Por último, el cuarto grupo solicitaba expresar la relación usando notación simbólica. Los ítemes, los grupos y las habilidades evaluadas se describen en la Tabla 1. La prueba completa se encuentra disponible en el anexo 1 en dataverse⁶.

Figura 1- Situación funcional empleada en la prueba

María está organizando las mesas y sillas donde se sentarán los invitados a su fiesta. Tiene varias mesas pequeñas cuadradas que irán ordenadas en fila para que se sienten todos juntos. Si ella pone solo una mesa caben 4 invitados. Si pone 2 mesas caben 6 invitados y si pone 3 mesas caben 8 invitados.

1 mesa **2 mesas** **3 mesas**

Fuente: Elaboración propia.

⁶ Anexos disponibles en dataverse: <https://doi.org/10.7910/DVN/4JALQ7>

Tabla 1- Ítemes prueba por grupos, habilidad evaluada y enunciados

Grupo (ítemes de...)	Habilidad evaluada	N° ítem	Enunciado
Pensamiento aritmético	Aplicar la estructura de la secuencia a casos cercanos	1	¿Cuántos invitados caben si se juntan 4 mesas?
		2	¿Cuántos invitados caben si se juntan 7 mesas?
	Identificar variables	3A	Organiza en la siguiente tabla el número de mesas y el número de invitados (Se incluye tabla de 2x9).
	Ordenar datos en una tabla identificando valores	3B	
	Identificar patrones en la tabla	4	¿Observas algún patrón en la tabla? Descríbelo.
Generalización numérica	Generalizar la estructura de la secuencia aplicándola a casos lejanos	5	¿Cuántos invitados caben si se juntan 100 mesas?
		6	¿Cuántos invitados caben si se juntan 204 mesas?
Generalización verbal	Generalizar la estructura de la secuencia empleando lenguaje natural	7	Describe cómo calcular el número de invitados que caben en cualquier número de mesas.
Generalización en lenguaje simbólico	Generalizar la estructura de la secuencia empleando lenguaje simbólico	8	¿Cómo describirías el total de invitados que caben en un número cualquiera de mesas usando variables (letras)?

Fuente: Elaboración propia.

Se construyó un guion con preguntas para cada ítem de la prueba. El guion estuvo orientado por lo que Willson y Miller (2014) definen como rol narrador, en él se le pide a los entrevistados que describan por qué respondieron cada ítem de la manera en que lo hicieron. Por ejemplo, “En el ítem 4 escribiste que en 100 mesas había 200 invitados, ¿por qué pusiste 200? ¿cómo llegaste a ese número?”. Si la respuesta dada por el estudiante explicitaba un razonamiento que fuera claro para el entrevistador se pasaba al siguiente ítem, de lo contrario, se realizaban preguntas adicionales del tipo “creo que no te entendí bien. ¿Me lo podrías explicar nuevamente?”

Para codificar las respuestas de los estudiantes en la prueba y en la entrevista se construyó una rúbrica basada en categorías teóricas. La ventaja de este tipo de rúbrica es que cada nivel de desempeño está descrito de manera cualitativa, lo que facilita la asignación de las respuestas a los diferentes niveles de desempeño (Förster *et al.*, 2017).

Para este estudio, el uso de la rúbrica incluye dos pasos. En el primero se codifican las respuestas a cada ítem. A modo de ejemplo, la tabla 2 muestra una parte de la rúbrica con los códigos para el ítem 7.

En el segundo paso se categorizó el desempeño de los estudiantes en cada grupo de ítemes (ver Tabla 1) considerando la codificación del primer paso. De esta manera, se les asignó un nivel (alto, medio o bajo). Por ejemplo, si un estudiante obtuvo un buen desempeño en los ítemes 5 y 6, que componen el grupo de generalización numérica, se codificó como nivel alto. La regla de asignación de niveles para cada grupo se describe de manera detallada en el anexo 2 disponible en dataverse.

Tabla 2- Códigos rúbrica ítem 7: “Describe cómo calcular el número de invitados que caben en cualquier número de mesas”

Código	Descriptor
A	Incluye operación correcta y nombra ambas variables correctamente: “se multiplica por dos el número de mesas y se le suma dos para obtener el número de personas”.
B	Incluye operación correcta y nombra una variable: “el número de mesas se multiplica por dos y se le suma dos”.
C	Describe operación correcta “se multiplica por dos y se le suma dos” o “se suma ambas veces el mismo número y se le suma dos”.
D1	Identifica solo la suma e incluye al menos una variable. Ejemplo, “al número de mesas se le suma dos”.
D2	Identifica solo la multiplicación e incluye al menos una variable. Ejemplo, “el número de mesas se multiplica por dos”.
E	Describe operación incompleta sin incluir variables. Ejemplo, “se multiplica por dos”.
F	La respuesta es incorrecta pero coherente con lo que hizo en los casos grandes específicos, por ejemplo, antes multiplicó por 4 y anota “se multiplica por 4”.
G	Incluye estrategia irrelevante. Ejemplo, “se dibujan las mesas y las personas y luego se cuentan las personas”.
H	No se entiende la idea descrita por el estudiante.
J	Escribe un ejemplo correcto, con o sin palabras que lo acompañen. Ejemplo, “por ejemplo, si tienes 70 mesas debes hacer $70+70+2=142$ ”.
NR	No responde.

Fuente: Elaboración propia.

Aplicación de instrumentos

La prueba fue aplicada por un investigador del proyecto a la semana siguiente de culminar la intervención. Al comienzo de la aplicación, el investigador leyó las preguntas detenidamente y dio la instrucción de escribir en la prueba todos los cálculos realizados. La aplicación tardó 60 minutos, tiempo en el que todos los estudiantes alcanzaron a responder la prueba completa.

Por otra parte, la entrevista retrospectiva se realizó al día siguiente de la prueba en una sala de la escuela en la que se instaló una cámara de video. El entrevistador fue llamando uno a uno a los estudiantes y les entregó su prueba para que recordaran sus respuestas al momento de responder la entrevista.

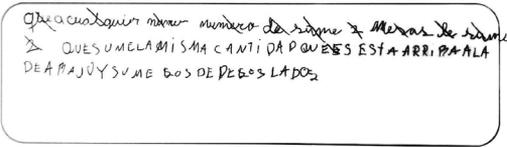
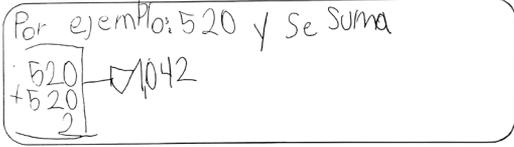
Proceso de codificación

La rúbrica fue aplicada por dos codificadores de manera independiente y los resultados fueron posteriormente contrastados. Para determinar el grado de concordancia entre los codificadores, se usó el índice Kappa de Cohen para cada ítem. Se obtuvieron valores entre 0,742 y 1, lo que indica un grado de concordancia adecuado (Cerdeira; Villarroel, 2008). Los desacuerdos fueron discutidos y consensuados para llegar a una única codificación.



Para ejemplificar el proceso de codificación, la figura 2 muestra la respuesta de dos estudiantes al ítem 7. La respuesta de E02 fue codificada con el código “C” (descrito en la tabla 2) y la respuesta de E29 fue codificada como “J”.

Figura 2- Respuestas de dos estudiantes al ítem 7

(E02)	(E29)
<p>7. Describe cómo calcular el número de personas que caben en cualquier número de mesas</p> 	<p>7. Describe cómo calcular el número de personas que caben en cualquier número de mesas</p> 

Fuente: Datos de la investigación.

Las entrevistas fueron transcritas y dos codificadores, de manera independiente, asignaron a cada respuesta el código de la rúbrica que mejor reflejaba el razonamiento descrito por el estudiante. Se estimó el índice de Kappa de Cohen para describir el grado de concordancia entre los codificadores para cada ítem. Los valores oscilaron entre 0,646 y 1, lo que se considera un acuerdo adecuado. Las discrepancias fueron discutidas y consensuadas.

Para ejemplificar este proceso de codificación, se presentan fragmentos de las entrevistas realizadas a dos estudiantes, E08 y E16, cuando respondieron a la pregunta ¿cómo llegaste a tu respuesta en el ítem 5? Pese a que ambas estudiantes habían escrito en su prueba la misma respuesta incorrecta: 102 invitados, las entrevistas mostraron razonamientos distintos. Los fragmentos y el análisis se muestran a continuación:

E08: Aquí sí me equivoque.

I: A ver, cuéntame.

E08: Lo que pasa es que tendría que sumar 100 más 100 más 2, y me equivoqué y puse 102.

I: ¿Y por qué tendrías que sumar 100 más 100 más 2?

E08: Porque en la parte de arriba, siempre, como en esta, siempre sale la misma cantidad de invitados y siempre hay dos, por eso 100 más 100 más 2, 202.

E16: Me di cuenta que yo no podía hacer 100 mesas porque me duraría hasta mañana. Entonces, pues lo imaginé en mi mente y supe que eran 102, porque iba de dos en dos... por eso 100+2.

Al analizar la entrevista se determinó que E08, pese a haber cometido un error de cálculo, identificaba correctamente la estructura de la situación funcional (código A). En cambio, E16, identificó solo una de las operaciones involucradas (código B1).

Análisis

A partir de las codificaciones realizadas, se definieron para cada estudiante las siguientes variables: el desempeño en cada ítem de la prueba y en cada ítem de la entrevista; el desempeño en cada grupo de ítems de la prueba y, por último, en cada grupo de ítems de la entrevista. Para analizar el grado de concordancia entre los códigos asignados en la prueba y la entrevista, se calcularon los porcentajes de coincidencia correspondientes para cada ítem y grupo de ítems.

Luego, para estudiar la relación entre las respuestas obtenidas en la entrevista y en la prueba, se utilizó el coeficiente de correlación tau de Kendall (Kendall, 1938). Este coeficiente permite establecer asociaciones entre variables ordinales. En consecuencia, dado que las variables en este estudio corresponden a categorías ordinales de desempeño derivadas de la evaluación, este coeficiente resulta especialmente adecuado para establecer asociaciones entre ellas.

Ahora bien, para dar cuenta del desempeño general en la prueba, se establecieron cinco categorías basadas en el modelo teórico empleado (ver Tabla 3). Todos los estudiantes fueron categorizados en una de ellas en función del desempeño obtenido en los grupos de ítems, a excepción de 2 estudiantes cuyos desempeños no se ajustaron al modelo. Por esa razón, se estableció una sexta categoría denominada trayectoria irregular.

Tabla 3- Criterio de asignación de categoría de desempeño en función de los resultados por grupo de ítem

Categorías de desempeño	Grupos de ítems			
	Pensamiento aritmético	Generalización numérica	Generalización verbal	Generalización simbólica
Pensamiento funcional simbólico	Alto o medio	Alto	Alto	Alto
Pensamiento funcional verbal	Alto o medio	Alto	Alto	Medio o bajo
Pensamiento funcional numérico	Alto o medio	Alto	Medio o bajo	Medio o bajo
Pensamiento aritmético	Alto o medio	Medio o bajo	Medio o bajo	Medio o bajo
Bajo desempeño	Bajo	Medio o bajo	Medio o bajo	Medio o bajo

Fuente: Elaboración propia.

Las categorías de desempeños, incluida trayectoria irregular, fueron transformadas a variables ordinales para poder estudiar su correlación. El orden de las categorías definidas de mayor a menor sofisticación fue el siguiente: funcional simbólico, funcional verbal, funcional numérico, trayectoria irregular, pensamiento aritmético y bajo desempeño. Se



ubicó la categoría trayectoria irregular por sobre pensamiento aritmético, ya que todos los casos correspondían a estudiantes que habían logrado desempeño alto en los ítems de generalización numérica, pero habían obtenido desempeño medio o bajo en todos los otros grupos de ítems.

Resultados

A continuación, se presentan los porcentajes de concordancia entre los códigos asignados a partir de la entrevista y de la prueba para cada ítem (ver Tabla 4). Se aprecia una coincidencia superior al 90% en los ítems p1, p2, p3a, p3b, p5 y p8, mientras que el ítem que alcanzó la menor coincidencia (58.5%) fue el ítem 7.

Tabla 4- Porcentaje de coincidencia entre el código asignado a partir de la prueba y la entrevista para cada ítem

N° pregunta	1	2	3A	3B	4	5	6	7	8
Coincidencia	95,8%	95,8%	95,8%	95,8%	87,5%	91,7%	83,3%	58,3%	91,7%

Fuente: Datos de la investigación

En los ítems que evaluaron habilidades aritméticas (ítems p1 a p4) se observó que aquello que los estudiantes respondieron en la prueba coincidió en mayor medida con lo que expresaron en la entrevista.

En los ítems P5 y P6 (relacionados con la generalización de la situación funcional a casos lejanos), hubo un 91,7% y un 83,3% de coincidencia entre la prueba y la entrevista respectivamente. Para estos casos se observó que quienes respondieron los ítems de manera correcta, efectivamente, realizaron los razonamientos anticipados, es decir, generalizaron la estructura de la situación funcional (el doble más dos) para números específicos. Las diferencias entre el código asignado a partir de la prueba y de la entrevista fue mayor en el caso del ítem 6 y corresponden a estudiantes que llegaron a resultados equivocados mediante procedimientos correctos sin incluir los cálculos en su respuesta, por ejemplo, E14 expresó “sumé 204 más 204 más 2 y me dio 500”. Debido a que varios estudiantes cometieron un error de cálculo al sumar $408+2$, una versión futura de la prueba debería considerar la modificación del ítem p6, reemplazando el valor de 204, por ejemplo, por 203, para evitar este tipo de error.

El ítem P8 también mostró un porcentaje alto de coincidencia (91%) entre la prueba y la entrevista, lo que sugiere que los estudiantes en su mayoría escribieron las expresiones algebraicas que consideraron adecuadas para representar la relación observada.

El ítem que mostró la menor coincidencia (58,5%) entre la prueba escrita y la entrevista fue el ítem P7. Este requería que los estudiantes escribieran en sus palabras una descripción de la relación entre las variables. Las discrepancias en los códigos se debieron a que, en general, las respuestas de los estudiantes fueron más completas en la entrevista. Por ejemplo, E06, quien en la prueba escribió “que se le sume 2”, explicó en la



entrevista: “quería explicar, pero no me salía la explicación, así que puse que se le sume 2” y, luego, proporcionó una explicación que demostraba su comprensión de la estructura de la situación funcional. Otro ejemplo fue E01, quien en su prueba escribió la relación sin nombrar variables (código C en Tabla 2), pero en la entrevista las señaló claramente, por lo que obtuvo el código A.

La Tabla 5 muestra el porcentaje de coincidencia entre los grupos de ítems evaluados en la prueba y en la entrevista, en ella se observa que en todos los grupos la coincidencia es sobre 83%. Nuevamente, el grupo con una menor coincidencia fue el que contenía el ítem P7. En general, se observa que la agrupación de ítems y la definición de los niveles alto, medio y bajo permitió aunar distintos desempeños cuyo código daba cuenta de un rendimiento similar, lo que se traduce en una mejora de los porcentajes de coincidencia. En el caso del ejemplo anterior, los desempeños de E01 en la prueba (código C) como en la entrevista (código A) fueron codificados como altos.

Tabla 5- Porcentaje de coincidencia entre el código asignado a partir de la prueba y la entrevista para cada grupo de ítems

Grupos de ítems	Ítems aritméticos	Ítems de generalización numérica	Ítems de generalización en lenguaje natural	Ítems de generalización simbólica
Coincidencia	87,5%	91,7%	83,3%	91,7%

Fuente: Datos de la investigación.

La tabla 6 presenta los resultados del análisis de correlación entre la prueba y la entrevista, tanto para cada grupo de ítems como para la categoría de desempeño general en la prueba. Se identificó una correlación alta y estadísticamente significativa entre los grupos de ítems, lo que muestra que, a pesar de discrepancias observadas y descritas previamente, las medidas asignadas para cada grupo de ítems van en la misma dirección. Además, la categoría de desempeño final mostró una correlación alta y significativa ($\tau_b=0,913$, $p < 0,001$; $N=24$), lo que sugiere que las conclusiones relacionadas con el pensamiento funcional de los estudiantes, derivadas de la prueba, se alinean en gran medida con las deducciones extraídas de las entrevistas.

Tabla 6- Correlación entre la prueba y la entrevista por grupos de ítems y por categoría de desempeño

	Grupo de ítems (ítems de...)				Categoría de desempeño
	Pensamiento aritmético	Generalización numérica	Generalización en lenguaje natural	Generalización en lenguaje simbólico	
Correlación	0,891***	0,871***	0,868***	0,920***	0,913***

Nota: Para estimar la correlación se usó la tau de Kendall.

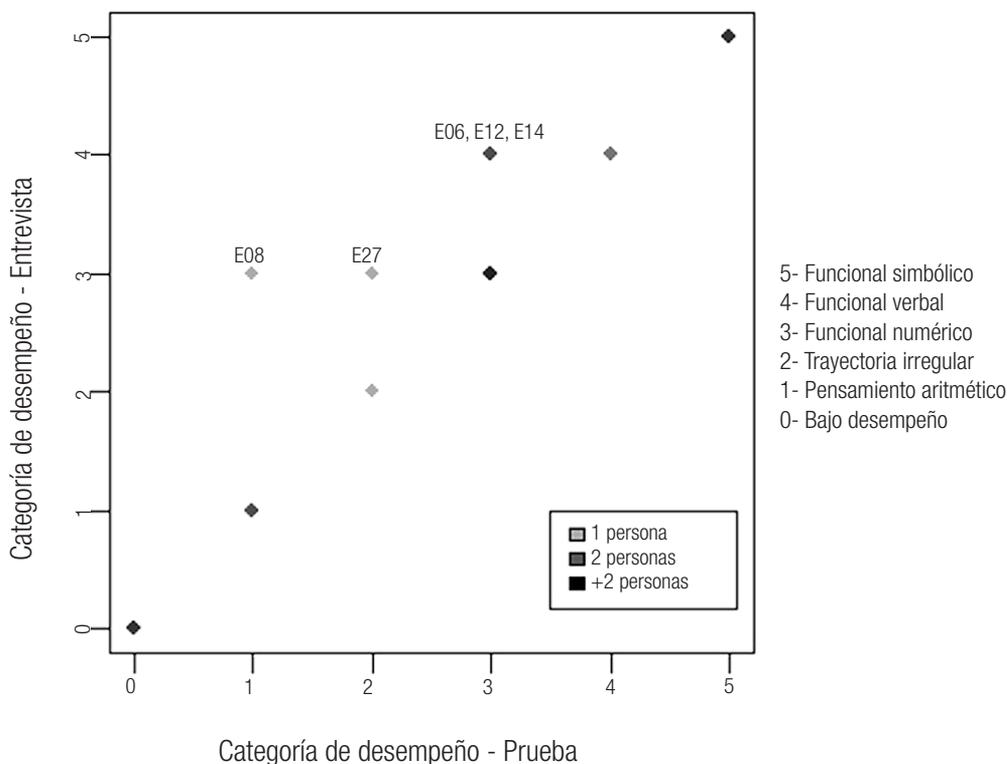
*** $p < 0,001$.

Fuente: Datos de la investigación.



Por último, el gráfico de la figura 3 presenta la correlación entre las categorías de desempeño general obtenidas a partir de la prueba y la entrevista. La mayoría de los casos (19 de 24) se encuentran en la diagonal del gráfico, lo que muestra coincidencia entre ambos instrumentos. En los casos en los que no hubo coincidencia, la entrevista situó a los estudiantes en una categoría de desempeño de un nivel mayor al de la prueba, es decir, la prueba tendió a subestimar el desempeño de estos 5 estudiantes.

Figura 3- Gráfico de dispersión entre las categorías de desempeño asignadas a partir de la prueba y de la entrevista



Fuente: Datos de la investigación.

Respecto a los casos que no están en la diagonal, tres estudiantes (E06, E12 y E14) se concentran en el mismo punto. En los tres casos se trata de estudiantes que en la prueba quedaron caracterizados con la categoría funcional numérico, pero en la entrevista se determinó que estaban en funcional verbal. Estos estudiantes no pudieron responder el ítem 7 en la prueba escrita, pero lo hicieron correctamente en la entrevista. Los otros dos puntos se debieron a aspectos menos frecuentes. E08 pasó de la categoría pensamiento aritmético a funcional numérico, ya que respondió de manera incorrecta los ítemes de generalización numérica y al entrevistarlo se pudo constatar que solo se debía a errores de cálculo. Mientras

que E27 había quedado categorizada en trayectoria irregular en la prueba por no responder los ítems P1 y P2 y solo realizar dibujos. Sin embargo, en la entrevista evidenció conocer los resultados, por lo que se le asignó la categoría funcional numérico.

Discusión

Las pruebas diseñadas para evaluar el pensamiento funcional a menudo parten de la premisa de que los estudiantes, al responder los distintos ítems, aplicarán razonamientos tales como identificar patrones y generalizar las relaciones entre variables. Analizar hasta qué punto los estudiantes efectivamente emplean estos razonamientos constituye un tipo de evidencia crucial. Este estudio se propuso aportar evidencia de validez basada en los procesos de respuesta (AERA; APA; NCME, 2014), determinando en qué medida las interpretaciones derivadas de una prueba diseñada para evaluar el pensamiento funcional de estudiantes de 3º grado coincidían con los razonamientos expresados por los estudiantes en una entrevista cognitiva. El análisis de concordancia entre la prueba y la entrevista permitió identificar aquellos ítems en los que se presentaron las mayores discrepancias. En términos generales, se encontró una correlación alta y significativa ($N=24$) entre el nivel de pensamiento funcional atribuido a través de la prueba y el determinado mediante la entrevista, lo que permite sustentar que la prueba desencadena el proceso mental que busca medir.

En concordancia con otros autores como Radford (2018), Stephens *et al.* (2017) y Blanton *et al.* (2015), el instrumento desarrollado en este estudio permitió identificar una trayectoria discernible en la evolución del pensamiento funcional de los estudiantes, en ella se distinguen categorías de desempeño que muestran distintos niveles de sofisticación. La categoría pensamiento aritmético identificó a los estudiantes que solo pudieron trabajar con casos cercanos y registrar algunos datos en una tabla. La categoría funcional numérico agrupó a quienes lograron establecer una relación de correspondencia y aplicarla a casos remotos, pero no articularon reglas generales. En la categoría funcional verbal, los estudiantes no solo generalizaron mediante registros numéricos, sino que también pudieron hacerlo a través del lenguaje natural escrito o hablado de forma indeterminada. La categoría funcional simbólico identificó estudiantes que generalizaron utilizando todos los registros de representación, incluyendo la notación variable.

A diferencia de Stephens *et al.* (2017), en el presente estudio no se observa que los estudiantes utilicen registros simbólicos sin la habilidad de expresar la relación en palabras, ya sean escritas o habladas. Es decir, aquellos estudiantes que logran generalizar correctamente mediante notación variable, también son capaces de expresar la relación en lenguaje natural, aunque algunos puedan tener dificultades para expresar sus ideas por escrito. En concordancia con autores como Radford (2018) y Pinto y Cañadas (2021), nuestros hallazgos sugieren que los estudiantes tienden a expresar la relación entre variables a través del lenguaje natural antes de adquirir la representación simbólica.

Al llevar a cabo un análisis detallado de las correspondencias entre la prueba y la entrevista en relación con cada ítem, se encontró que las discrepancias más importantes se presentaron en el ítem 7, en el que se solicitaba a los estudiantes que escribieran en



sus propias palabras la relación existente entre las variables. Se observó que solo el 58% de los estudiantes contestó en la prueba aquello que sabía, mientras que el 42% restante, elaboró respuestas más completas durante la entrevista. Esto sugiere que redactar una regla funcional generalizada puede ser complejo para varios estudiantes. Lo anterior, es consistente con estudios que señalan las dificultades que enfrentan los estudiantes de esta edad para expresar ideas matemáticas por escrito (Blackstock-Bernstein *et al.*, 2022; Hughes *et al.*, 2020).

Considerar este aspecto puede ser relevante al establecer conclusiones basadas en este tipo de ítems. Por ejemplo, en el estudio de Stephens *et al.* (2017), que empleó instrumentos escritos, se reportó que los estudiantes hallaban más sencillo expresar la generalización mediante notación variable que a través de palabras. Los resultados del presente estudio sugieren que la dificultad para generalizar usando lenguaje natural podría estar más asociada con el registro escrito que con el lenguaje natural en sí.

Por último, en el análisis de la categoría de desempeño general asignada, se observó que en el 79% de los estudiantes evaluados (19 de 24), la prueba y la entrevista coincidieron en su clasificación. En los casos restantes (21%), la prueba subestimó el nivel que se les atribuyó según la entrevista. Estas discrepancias se encontraron principalmente en estudiantes clasificados en la categoría funcional numérico. Para estos estudiantes, la entrevista permitió precisar su capacidad de generalizar reglas funcionales mediante lenguaje natural. Estos resultados sugieren que la aplicación del instrumento podría beneficiarse de la incorporación de entrevistas breves que complementen la prueba, focalizadas en el ítem 7 y en estudiantes de la categoría funcional numérico. Esta adaptación metodológica podría enriquecer la evaluación del pensamiento funcional sin depender excesivamente de las habilidades de redacción de los estudiantes.

Conclusiones

Este estudio aporta evidencia de validez basada en los procesos de respuesta que respalda las interpretaciones derivadas de una prueba escrita, diseñada para evaluar el pensamiento funcional en estudiantes de 3° grado (9-10 años). En términos generales, el instrumento analizado –compuesto por la prueba y una rúbrica– capturó de manera fiable el pensamiento funcional de los estudiantes. Un hallazgo relevante fue que el ítem con mayor discrepancia entre la prueba y la entrevista involucraba la redacción de la relación funcional usando lenguaje natural. Esto sugiere que futuras investigaciones podrían considerar la inclusión de herramientas complementarias, como entrevistas breves, para examinar si estudiantes caracterizados en la categoría funcional numérico pueden expresar sus generalizaciones mediante lenguaje natural independientemente de sus habilidades de redacción.

Si bien estos hallazgos son significativos, es importante señalar que la evidencia fue obtenida en un curso de educación primaria de una escuela pública de nivel socioeconómico bajo. Para evaluar su aplicabilidad en otros contextos, sería necesario realizar análisis de validez adicionales con estudiantes de distintos niveles socioeconómicos. Esta ampliación permitiría no solo explorar la generalidad de las conclusiones, sino también determinar



si las dificultades con el lenguaje escrito se presentan en otros grupos sociodemográficos. Además, cabe señalar que los procesos de respuesta corresponden a una de varias posibles fuentes de evidencia de validez, por lo que futuros estudios podrían indagar en otras fuentes para robustecer la validez de las interpretaciones derivadas de la prueba.

Referencias

AERA; APA; NCME. **Standards for Educational & Psychological Testing**. Washington, DC: American Educational Research Association, 2014.

BLACKSTOCK-BERNSTEIN, Anne; WOODBRIDGE, Amy y BAILEY, Alison. Comparing elementary students' explanatory language across oral and written modes. **The Elementary School Journal**, Chicago, v.122, n. 3, p. 315-340, 2022. <https://doi.org/10.1086/718077>

BLANTON, Maria; BRIZUELA, Barbara; GARDINER, Angela Murphy; SAWREY, Katie; NEWMAN-OWENS, Ashley. Progression in first-grade children's thinking about variable and variable notation in functional relationships. **Educational Studies in Mathematics**, v. 95, p. 181-202, 2017. <https://doi.org/10.1007/s10649-016-9745-0>

BLANTON, Maria; KAPUT, James. Functional thinking as a route into algebra in the elementary grades. *In*: CAI, Jinfa; KNUTH, Eric (ed.). **Early algebraization**. Berlin: Springer, 2011. p. 5-23.

BLANTON, Maria; STEPHENS, Ana; KNUTH, Eric; GARDINER, Angela Murphy; ISLER, Isil; KIM, Jee-Seon. The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. **Journal for Research in Mathematics Education**, Reston, v. 46, n. 1, p. 39-87, 2015. <https://doi.org/10.5951/jresmetheduc.46.1.0039>

BLANTON, Maria; STROUD, Rena; STEPHENS, Ana; GARDINER, Angela Murphy; STYLIANOU, Despina; KNUTH, Eric; ISLER-BAYKAL, Isil; STRACHOTA, Susanne. Does early algebra matter? The effectiveness of an early algebra intervention in grades 3 to 5. **American Educational Research Journal**, Washington, DC., v. 56, n. 5, p. 1930-1972, 2019. <https://doi.org/10.3102/0002831219832301>

BRIZUELA, Barbara; BLANTON, Maria; SAWREY, Katharine; NEWMAN-OWENS, Ashley; GARDINER, Angela Murphy. Children's use of variables and variable notation to represent their algebraic ideas. **Mathematical Thinking and Learning**, Cham, v. 17, n. 1, p. 34-63, 2015. <https://doi.org/10.1080/10986065.2015.981939>

CAICEDO, Estefanía; ZALAZAR-JAIME, Mauricio Federico. Entrevistas cognitivas: revisión, directrices de uso y aplicación en investigaciones psicológicas. **Avaliação Psicológica**, São Paulo, v. 17, n. 3, p. 362-370, 2018. <https://doi.org/10.15689/ap.2018.1703.14883.09>

CAÑADAS, María; BLANTON, Maria; BRIZUELA, Barbara. Special issue on early algebraic thinking. **Journal for the Study of Education and Development**, Madrid, v. 42, n. 3, p. 469-478, 2019. <https://doi.org/10.1080/02103702.2019.1638569>



CASTILLO, Miguel; PADILLA, José Luís. How cognitive interviewing can provide validity evidence of the response processes to scale items. **Social Indicators Research**, Dordrecht, v. 114, p. 963-975, 2012. <https://doi.org/10.1007/s11205-012-0184-8>.

CERDA L, Jaime; VILLARROEL DEL P, Luis. Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. **Revista Chilena de Pediatría**, Santiago de Chile, v. 79, n. 1, p. 54-58, 2008.

CHIMONI, María; PITTA-PANTAZI, Demetra; CHRISTOU, Constantinos. The impact of two different types of instructional tasks on students' development of early algebraic thinking (El impacto de dos tipos diferentes de tareas instruccionales en el desarrollo del pensamiento algebraico temprano de los estudiantes). **Journal for the Study of Education and Development**, Madrid, v. 44, n. 3, p. 503-552, 2021. <https://doi.org/10.1080/02103702.2020.1778280>

CONRAD, Frederick G.; BLAIR, Johnny. Sources of error in cognitive interviews. **The Public Opinion Quarterly**, Oxford, v. 73, n. 1, p. 32-55, 2009. <https://doi.org/10.1093/poq/nfp013>

FÖRSTER, Carla; ZEPEDA, Sandra; NUÑEZ, Claudio. Instrumentos para la evaluación de aprendizajes, ¿con qué evaluar? *In*: FÖRSTER, Carla E. (ed.). **El poder de la evaluación en el aula**: mejores decisiones para promover aprendizaje. Santiago de Chile: Universidad Católica de Chile, 2017. p. 177-230.

GOÑI-CERVERA, Juncal; CAÑADAS, María; POLO-BLANCO, Irene. Generalizations in students with autism spectrum disorder: an exploratory study of strategies. **ZDM—Mathematics Education**, Berlin, v. 54, n. 6, p. 1333-1347, 2022. <https://doi.org/10.1007/s11858-022-01415-w>

HUGHES, Elizabeth; RICCOMINI, Paul; LEE, Joo-Young. Investigating written expressions of mathematical reasoning for students with learning disabilities. **The Journal of Mathematical Behavior**, v. 58, n. 2, 2020. <https://doi.org/10.1016/j.jmathb.2020.100775>

KAPUT, James. What is algebra? What is algebraic reasoning? *In*: KAPUT, James; CARRAHER, Davis; BLANTON, Maria (ed.). **Algebra in the early grades**. New York: Taylor and Francis, 2008. p. 5-17.

KENDALL, Maurice. A new measure of rank correlation. **Biometrika**, Oxford, v. 30, n 1/2, p. 81-93, 1938. <https://doi.org/10.2307/2332226>

KIERAN, Carolyn. The multi-dimensionality of early algebraic thinking: background, overarching dimensions, and new directions. **ZDM Mathematics Education**, Berlin, v. 54, n. 6, p. 1131-1150, 2022. <https://doi.org/10.1007/s11858-022-01435-6>

LARSSON, Anneli; GRANHAG, Pär Anders. Interviewing children with the cognitive interview: assessing the reliability of statements based on observed and imagined events. **Scandinavian journal of psychology**, Stockholm, v. 46, n. 1, p. 49-57. 2005. <https://doi.org/10.1111/j.1467-9450.2005.00434.x>

MASON, John; STEPHENS, Max; WATSON, Anne. Appreciating mathematical structure for all. **Mathematics Education Research Journal**, Sydney, v. 21, n. 2, p. 10-32, 2009. <https://doi.org/10.1007/BF03217543>



MEADOWS, Keith. Cognitive Interviewing Methodologies. **Clinical Nursing Research**, Thousand Oaks, v. 30, n. 4, p.375-379. 2021. <https://doi.org/10.1177/10547738211014099>

MOSES, Robert y COBB, Charles. **Radical equations: math literacy and civil rights**. Boston: Beacon Press. 2001.

MUSEUS, Samuel; PALMER, Robert; DAVIS, Ryan; MARAMBA, Dina. Special issue: racial and ethnic minority students' success in STEM education. **ASHE Higher Education Report**, Hoboken, v. 36, n. 6, p. 1-140. 2013. <https://doi.org/10.1353/rhe.2013.0046>

NCTM. National Council of Teachers of Mathematics. **Principles and standards for school mathematics**. Reston: NCTM. 2000.

PADILLA, José Luís; BENÍTEZ, Isabel. Validity evidence based on response processes. **Psicothema**, Oviedo, v. 26, n. 1, p. 136–144, 2014. <https://doi.org/10.7334/psicothema2013.259>

PADILLA, José Luís; LEIGHTON, Jacqueline. Cognitive interviewing and think aloud methods. *In*: ZUMBO, Bruno; HUBLEY, Anita (ed.). **Understanding and investigating response processes in validation research**. [S. l.]: Springer International/Springer Nature, 2017. p. 211-228. https://doi.org/10.1007/978-3-319-56129-5_12

PINTO, Eder; CAÑADAS, María. Generalizations of third and fifth graders within a functional approach to early algebra. **Mathematics Education Research Journal**, Sydney, v. 33, n. 1, p. 113-134. 2021. <https://doi.org/10.1007/s13394-019-00300-2>

RADFORD, Luis. The emergence of symbolic algebraic thinking in primary school. *In*: KIERAN, Carolyn (ed.). **Teaching and learning algebraic thinking with 5- to 12-year-olds**. Cham: Springer, 2018. p. 3-25. ICME-13 Monographs. https://doi.org/10.1007/978-3-319-68351-5_1

STEPHENS, Ana; FONGER, Nicole; STRACHOTA, Susanne; ISLER, Isil; BLANTON, Maria; KNUTH, Eric; GARDINER, Angela. A learning progression for elementary students' functional thinking. **Mathematical Thinking and Learning**, Cham, v. 19, n. 3, p. 143-166, 2017. <https://doi.org/10.1080/10986065.2017.1328636>

STYLIANOU, Despina; STROUD, Rena; CASSIDY, Michael; KNUTH, Eric; STEPHENS, Ana; GARDINER, Angela; DEMERS, Lindsay. Putting early algebra in the hands of elementary school teachers: examining fidelity of implementation and its relation to student performance. El álgebra temprana en manos del docente de primaria: un análisis de la fidelidad de ejecución y su relación con el rendimiento de los escolares. **Journal for the Study of Education and Development**, Madrid, v. 42, n. 3, p. 523-569, 2019. <https://doi.org/10.1080/02103702.2019.1604021>

UREÑA, Jason; RAMÍREZ, Rafael; MOLINA, Marta. Representaciones de la generalización de una relación funcional y el vínculo con la mediación del entrevistador. **Journal for the Study of Education and Development**, Madrid, v. 42, n. 3, p. 591-614, 2019. <https://doi.org/10.1080/02103702.2019.1604020>



WILLIS, Gordon. **Cognitive interviewing**: A tool for improving questionnaire design. Thousand Oaks: Sage, 2019.

WILLSON, Stephanie; MILLER, Kristen. Data collection. *In*: MILLER, Kristen; WILLSON, Stephanie; CHEPP, Valeria; PADILLA, José-Luís (ed.). **Cognitive interviewing methodology**. Hoboken: Wiley, 2014. p. 15-33.

Recibido en: 30.05.2024

Revisado en: 08.10.2024

Aprobado en: 25.11.2024

Editora responsable: Profa. Dra. Tatiana Hochgreb-Haegele

Paulina Araya es profesora asistente de la Facultad de Educación de la Universidad Diego Portales. Licenciada en ciencias exactas y profesora de matemática y física por la Universidad de Chile y doctora en educación por las Universidades Diego Portales y Alberto Hurtado.

Beltrán Pantoja es profesor de matemática de la Universidad Alberto Hurtado y Magíster en educación con mención en evaluación de aprendizajes por la Pontificia Universidad Católica de Chile.

Andrea Valenzuela es profesora adjunta de la Facultad de Educación de la Pontificia Universidad Católica de Chile. Educadora de párvulos, Magíster en educación con mención en evaluación de aprendizajes en la misma universidad.