



## Construcción y validación de un cuestionario para evaluar la comprensión del gráfico de caja

Construction and validation of a questionnaire to assess boxplot comprehension

*Construção e validação de um questionário para avaliar a compreensão do gráfico de caixa*

ARTÍCULO ORIGINAL



Escanea en tu dispositivo móvil o revisa este artículo en:

<https://doi.org/10.33996/revistahorizontes.v8i33.750>

**Rodrigo Gutiérrez-Martínez**   
rogutierrezmartinez@gmail.com

**Audy Salcedo**   
audy.salcedo@gmx.com

**Danilo Díaz-Levicoy**   
dddiaz01@hotmail.com

Universidad Católica del Maule. Talca, Chile

Artículo recibido 9 de enero 2023 | Aceptado 30 de enero 2023 | Publicado 2 de abril 2024

### RESUMEN

El gráfico de caja se ha incorporado a los currículos oficiales de distintos países, pasando a ser parte de la formación de profesores. Por ello, que este trabajo expone el proceso de diseño y validación de un instrumento para evaluar la comprensión del gráfico de caja para futuros profesores de Educación Básica. Para construir el cuestionario se revisaron libros de texto e investigaciones previas, tipos de tareas y demanda cognitiva. Se trata de una investigación mixta. Para estimar la validez de contenido se siguió la técnica del juicio de expertos, quienes puntuaron los ítems según claridad, pertinencia y relevancia, y comentarios para su mejora. Además, se calculó el coeficiente de validez de contenido. El cuestionario final reúne una muestra de ítems que se consideran apropiados para medir la comprensión del gráfico de caja en el profesorado.

**Palabras clave:** Construcción de cuestionario; Validación de cuestionario; Cuestionario; Gráfico de caja; Comprensión

### ABSTRACT

The box plot has been incorporated into the official curricula of different countries, becoming part of teacher training. Therefore, this paper presents the design and validation process of an instrument to evaluate the comprehension of the box plot for future elementary school teachers. To construct the questionnaire, textbooks and previous research, types of tasks and cognitive demand were reviewed. This is a mixed research. To estimate the content validity, the technique of expert judgment was followed, who scored the items according to clarity, pertinence and relevance, and comments for improvement. In addition, the content validity coefficient was calculated. The final questionnaire gathers a sample of items considered appropriate to measure teachers' understanding of the box plot.

**Key words:** Questionnaire construction; Questionnaire validation; Questionnaire; Box plot; Comprehension

### RESUMO

O gráfico de caixa foi incorporado aos currículos oficiais de diferentes países, tornando-se parte da formação de professores. Portanto, este artigo apresenta o projeto e o processo de validação de um instrumento para avaliar a compreensão do gráfico de caixa para futuros professores do ensino fundamental. Para elaborar o questionário, foram revisados livros didáticos e pesquisas anteriores, tipos de tarefas e demanda cognitiva. Trata-se de uma pesquisa mista. Para estimar a validade do conteúdo, foi usado o julgamento de especialistas para pontuar os itens quanto à clareza, relevância e pertinência, além de comentários para aprimoramento. Além disso, foi calculado o coeficiente de validade do conteúdo. O questionário final reúne uma amostra de itens considerados adequados para medir a compreensão dos professores sobre o gráfico de caixa.

**Palavras-chave:** Construção de questionário; Validação de questionário; Questionário; Questionário; Box plot; Compreensão

## INTRODUCCIÓN

En la actualidad, las grandes cantidades de información estadística presente en medios de comunicación y redes sociales, hacen necesario que los ciudadanos tengan la capacidad de analizar de forma crítica y comprensiva. Esto se conoce con el nombre de cultura estadística (del Pino y Estrella, 2012). Por su parte, Batanero (2004) indica que este término surgió de forma espontánea entre la comunidad de estadísticos y educadores en estadística, haciendo alusión a la importancia del conocimiento estadístico para el ciudadano educado. Para Gal (2002) la cultura estadística está relacionada con la interrelación de dos componentes: la capacidad de las personas de evaluar críticamente la información estadística, argumentos y fenómenos estocásticos en diversos contextos, y la capacidad de comunicar sus reacciones a dicha información estadística.

En este marco, Arteaga et al., (2011) destacan el rol de las representaciones gráficas como un elemento de la cultura estadística, las que aparecen ampliamente en los medios de comunicación, prensa e internet en general. En este sentido, la comprensión de información a partir de gráficos estadísticos se vuelve fundamental, ya que una persona en la sociedad actual, debe leer constantemente y de forma crítica información gráfica proveniente de diferentes medios (Arteaga et al., 2016). El lenguaje gráfico, además, tiene un rol primordial en el análisis de datos y está presente en ámbitos tan diversos como la educación, redes sociales, la publicidad y movimientos que

promueven la apertura y transparencia de datos (Alcaraz et al., 2020).

Específicamente, la necesidad de mejorar la cultura estadística de la población en general ha motivado la inclusión de temas de estadística en el currículo de Educación Básica en diversas partes del mundo, incluyendo Estados Unidos, España, Francia y gran parte de los países de Iberoamérica (Cuétara et al., 2016). En Chile, siguiendo tendencias internacionales, el currículo introduce el eje temático de Datos y azar en el año 2009, el cual aborda el tratamiento de la estadística y su parte gráfica desde los primeros años de escolaridad (Ministerio de Educación de Chile [MINEDUC], 2009). En este sentido, el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, perteneciente al MINEDUC, indica que los profesores de Educación Básica en Chile deben manejar adecuadamente representaciones estadísticas como pictogramas, gráficos de barras y circulares, histogramas y gráfico de caja (MINEDUC, 2022).

Diversas investigaciones como la de Arteaga et al., 2016; Espinel, 2007; Monteiro y Ainley, 2007; Rodríguez-Alveal y Díaz-Levicoy, 2019; Salcedo et al., 2021 buscan descifrar qué tan bien comprenden los futuros profesores los gráficos estadísticos, pero es usual que se centran; principalmente, en un pequeño grupo de esas representaciones, como el gráfico de barras o el histograma, con ello dejan de lado otras, como el gráfico de caja, el cual es de uso común en reportes estadísticos y de investigación. Además de que

su estudio está considerado desde la Educación Primaria dentro del currículo de matemática de diversos países (e.g., National Council of Teachers of Mathematics [NCTM], 2000). En este sentido, son pocas las investigaciones que abordan la comprensión del gráfico de caja de futuros profesores o profesores en activo, ninguna encontrada en Chile. Esas investigaciones lo hacen con instrumentos que, al parecer, no han pasado por un proceso de validación, o al menos no se describen. Por ejemplo, Carvalho et al., (2019) presentan un instrumento con cuatro preguntas, pero no menciona una validación de este. En la misma línea, Edwards et al., (2017) también diseñan un instrumento para su evaluación, pero tampoco hablan de una validación.

Por otra parte, Retnawati et al., (2017) señalan que el desarrollo de instrumentos de calidad, válidos y confiables, juega un papel decisivo en la identificación de las competencias y de las dificultades de los estudiantes. De allí que este trabajo tiene por objetivo construir y validar un instrumento para evaluar la comprensión del gráfico de caja que tienen futuros profesores de Educación Básica. Con esto se espera (a) contribuir con un cuestionario que pueda ser utilizados en distintos contextos para evaluar de comprensión del gráfico de caja por futuros profesores, (b) ejemplificar el procedimiento a seguir para el diseño y validación, mediante juicio de experto, de un cuestionario de investigación.

## MÉTODO

Se trata de un estudio descriptivo, para establecer la validez de contenido de un cuestionario para evaluar la comprensión del gráfico de caja que tienen futuros profesores de Educación Básica, por medio de juicio de expertos. De acuerdo con Haynes et al., (1995) el estudio de validez de contenido de un instrumento de recolección de datos es un proceso multimétodo, que contempla un nivel cualitativo y otro cuantitativo. En lo cualitativo está el análisis de los aportes y observaciones proporcionados por cada evaluador y lo cuantitativos se encuentra en el uso de medidas estadísticas para el análisis de las puntuaciones asignadas a cada ítem, además, del uso de un coeficiente específico para evaluar la validez de contenido.

Para la elaboración del cuestionario, se realizó una revisión de los libros de texto de matemática de segundo ciclo de Enseñanza Básica de los últimos años en Chile, así como investigaciones realizadas sobre los diagramas de caja. Para caracterizar la complejidad que la resolución de un problema matemático plantea a los estudiantes (o profesores, como en este caso), es útil el disponer de un modelo teórico que brinde las herramientas para su análisis (Benedicto et al., 2015). Es por ello que se consideró apropiado tomar como referencia un modelo de niveles de demanda cognitiva y los tipos de tareas sobre gráficos estadísticos descritos en la literatura.

Se utilizó el modelo de niveles de demanda cognitiva, originalmente propuesto por Stein et al., (2000) y adaptado a la estadística por Salcedo (2015) y que considera los siguientes niveles: 1) tareas de memorización: actividad directa, sin ambigüedad, en el caso del gráfico, lectura literal, sin implicar interpretación; 2) tareas de procedimiento sin conexión: actividades que no demandan justificaciones del proceso realizado. Por ejemplo, estudiante concluye de forma simple y directa a partir de gráficos; 3) tareas de procedimiento con conexión: actividades que requieren pensamiento complejo para su resolución. Demandan procedimientos, aunque de forma general. Por ejemplo, estudiante extrae información no explícita a partir de los datos y su contexto; 4) tareas para hacer estadística: actividades que requieren pensamiento complejo y no algorítmico, exigiendo el entendimiento de los conceptos y procedimientos. Por ejemplo, el estudiante hace inferencias, siendo crítico con la información, de acuerdo con los datos y su contexto.

Los niveles van desde el más básico (tareas de memorización) al más complejo y que implica una mayor demanda cognitiva (tareas para hacer estadística). El modelo de demanda cognitiva se centra en el nivel de pensamiento exigido cuando este se enfrenta a una actividad estadística, para resolverla con éxito. Con respecto a los tipos de tareas sobre gráficos estadísticos propuestas

a los estudiantes, distintas investigaciones han caracterizado su presencia en libros de texto (Díaz-Levicoy et al., 2016, 2018). Algunas de ellas, y que se han considerado y adaptado para esta investigación, son: 1) leer: se pide una lectura literal del gráfico estadístico; 2) construir: se pide elaborar un gráfico de caja con la información proporcionada; 3) calcular: se debe leer la información que se presenta en el gráfico y realizar operaciones aritméticas a partir de ella; 4) interpretar: se pide extraer información no literal presente en el gráfico, dar puntos de vista y justificar sus afirmaciones.

Con esos referentes, se elaboró una tabla de especificaciones con la estructura tentativa del cuestionario y se hizo una primera producción de ítems. La discusión entre investigadores llevó a una depuración preliminar y se generó una primera versión del cuestionario. Esa propuesta inicial consistió en un cuestionario de 10 ítems, cada uno con dos versiones que se diferenciaban en aspectos como los datos manejados y el lenguaje utilizado. Luego de una depuración inicial, se seleccionaron 8 ítems adecuados para abordar los distintos aspectos importantes a evaluar, considerando las tareas de lectura, construcción, cálculo e interpretación, descritas previamente. En la Tabla 1 se especifica el tipo de tarea, nivel de demanda cognitiva y la forma en que son presentados los datos en cada uno de los ocho problemas propuestos.

**Tabla 1.** Tabla de especificación del cuestionario.

Ítem	Tarea	Soporte de datos	Demanda cognitiva
1a	Leer	Gráfico de caja	Nivel 1
1b	Leer	Gráfico de caja	Nivel 1
2a	Leer	Gráfico de caja	Nivel 1
2b	Leer	Gráfico de caja	Nivel 1
3	Construir	Información del diagrama	Nivel 2
4	Construir	Conjunto por extensión	Nivel 2
5a	Construir	Gráfico de barras	Nivel 2
5b	Leer	Gráfico de barras	Nivel 1
6a	Leer	Gráfico de caja	Nivel 1
6b	Calcular	Gráfico de caja	Nivel 2
6c	Calcular	Gráfico de caja	Nivel 2
6d	Leer	Gráfico de caja	Nivel 1
6e	Interpretar	Gráfico de caja	Nivel 3
7a	Interpretar	Gráfico de caja	Nivel 3
7b	Interpretar	Gráfico de caja	Nivel 3
7c	Interpretar	Gráfico de caja	Nivel 3
7d	Interpretar	Gráfico de caja	Nivel 3
8	Interpretar	Histograma	Nivel 3

Varios de los ítems se consideran subítems, abordando los primeros 3 niveles del modelo de demanda cognitiva. Hay 7 preguntas que se clasifican como tareas de memorización (primer nivel del modelo), cinco de procedimiento sin conexión (segundo nivel) y las últimas seis preguntas del cuestionario son tareas de procedimiento con conexión (tercer nivel). Debido a que el cuestionario fue confeccionado para ser aplicado en futuros profesores de Educación Básica, se decidió no abordar el cuarto nivel del modelo, tareas para hacer estadística.

Para la validación se recurrió al juicio de expertos, técnica comúnmente utilizada en la evaluación de diseños de cuestionarios (Hernández et al., 2014). Por su parte, Juárez-Hernández

y Tobón (2018) señalan que es un proceso fundamental en la creación de instrumento, pues esta proporciona evidencia sobre la calidad y precisión de este. El juicio de expertos consiste en solicitar a un grupo de especialistas la revisión del instrumento para que lo evalúen (Hernández y Mendoza, 2018), de acuerdo con unos criterios previamente establecidos. Luego es necesario establecer el grado de correspondencia entre las apreciaciones de los jueces al momento de valorar los ítems.

Se consultó a investigadores reconocidos de Chile, Brasil, Perú, Argentina y España y que son expertos en áreas relevantes para la construcción de este cuestionario: Estadística, Didáctica de la Matemática, Didáctica de la Estadística o

Evaluación de aprendizajes. Todos los expertos contactados tienen publicaciones en sus áreas de especialización y estudios de postgrado, la mayoría con doctorado. A los 9 expertos contactados vía correo electrónico se les envió una carta indicando detalles del instrumento, el objetivo e instrucciones de cómo evaluar, y se recogieron 8 valoraciones.

La valoración solicitada consistía en asignar un puntaje de 1 a 5, donde 1 es la calificación mínima y 5 la máxima, en tres apartados: claridad (los enunciados de cada ítem se entienden sin dificultad alguna, siendo conciso, exacto y directo), pertinencia (si el ítem pertenece a la categoría) y relevancia (el ítem es apropiado para representar a la categoría específica del constructo). Además, se pidió indicar sugerencias para mejorar los aspectos que el experto considerara necesario.

Una vez recibidas las evaluaciones de cada uno de los expertos, se realizó una valoración de cada ítem, considerando los aspectos evaluados (claridad, pertinencia y relevancia), calculando medidas de estadística descriptiva. Este análisis se complementó con el Coeficiente de Validez de Contenido (CVC), propuesto por Hernández-Nieto (2011). Este coeficiente establece una relación proporcional entre la evaluación promedio observada entre los jueces evaluadores y la evaluación máxima teórica, según la escala utilizada, ajustada, además, por la concordancia aleatoria entre los jueces.

El CVC ha sido utilizado en diferentes contextos. Por ejemplo, Balbinotti et al. (2007) lo utilizó para validar el contenido de la traducción de un cuestionario financiero con fin de utilizarlo

en Brasil. En general, con algunas excepciones, concluyeron que el instrumento es apto para ser aplicado en dicho país. Esto, basándose en el cálculo del CVC, considerando aceptable un valor superior al 0,70. En el ámbito de la salud, Goncalves Bos et al., (2017) validaron un cuestionario sobre la percepción de riesgos de caídas en adultos mayores. El cuestionario diseñado, que contó inicialmente con 36 preguntas, fue sometido a una evaluación por juicio de expertos en tres criterios: claridad del lenguaje, importancia y relevancia teórica. También usando un criterio de 0,70, el cuestionario final se redujo a 14 preguntas. Los autores indican que este coeficiente es el más utilizado en salud, especialmente en enfermería, para validar este tipo de instrumentos.

Además, de los análisis anteriores se consideraron los comentarios dejados por los expertos al momento de emitir los juicios sobre los ítems. Con toda esa información se procedió a modificar algunos ítems. Para efectos del análisis, se codificaron los expertos de la forma EX, donde E significa experto y X corresponde a un número del 1 al 8.

## RESULTADOS Y DISCUSIÓN

En este apartado se detallan los principales resultados obtenidos a partir de la validación solicitada al grupo de expertos. La Tabla 2 resume las puntuaciones otorgadas por los expertos en cada ítem del cuestionario, mostrando el criterio evaluado, la frecuencia de cada puntuación, junto a la media, moda, desviación estándar y coeficiente de simetría.

**Tabla 2.** Puntuación asignada por los expertos en cada ítem.

Ítem	Criterio	Puntuación					Media	Moda	Desv. est.	Coef. de asimetría
		1	2	3	4	5				
1a	Claridad	0	0	0	0	8	5	5	0	###
	Pertinencia	0	0	0	2	6	4,8	5	0,5	-1
	Relevancia	0	0	0	1	7	4,9	5	0,4	-3
1b	Claridad	0	0	1	0	7	4,8	5	0,7	-3
	Pertinencia	0	0	0	3	5	4,6	5	0,5	-1
	Relevancia	0	0	0	1	7	4,9	5	0,4	-3
2a	Claridad	0	0	1	0	7	4,8	5	0,7	-3
	Pertinencia	0	0	0	1	7	4,9	5	0,4	-3
	Relevancia	0	0	0	0	8	5	5	0	###
2b	Claridad	0	1	1	1	5	4,3	5	1,2	-1
	Pertinencia	0	0	1	2	5	4,5	5	0,8	-1
	Relevancia	0	0	2	0	6	4,5	5	0,9	-1
3	Claridad	0	1	2	0	5	4,1	5	1,2	-1
	Pertinencia	0	0	0	2	6	4,8	5	0,5	-1
	Relevancia	0	0	0	1	7	4,9	5	0,4	-3
4	Claridad	0	0	1	1	6	4,6	5	0,7	-2
	Pertinencia	0	0	0	2	6	4,8	5	0,5	-1
	Relevancia	0	0	0	2	6	4,8	5	0,5	-1
5a	Claridad	0	1	2	0	5	4,1	5	1,2	-1
	Pertinencia	0	0	0	2	6	4,8	5	0,5	-1
	Relevancia	0	0	0	1	7	4,9	5	0,4	-3
5b	Claridad	0	1	1	1	3	4	5	1,3	-1
	Pertinencia	0	0	0	3	3	4,5	4	0,5	0
	Relevancia	0	0	1	1	4	4,5	5	0,8	-2
6a	Claridad	0	0	1	0	7	4,8	5	0,7	-3
	Pertinencia	0	0	2	1	5	4,4	5	0,9	-1
	Relevancia	0	1	1	0	6	4,4	5	1,2	-2
6b	Claridad	0	0	1	1	6	4,6	5	0,7	-2
	Pertinencia	0	0	2	1	5	4,4	5	0,9	-1
	Relevancia	0	1	1	0	6	4,4	5	1,2	-2
6c	Claridad	0	0	1	1	6	4,6	5	0,7	-2
	Pertinencia	0	0	2	1	5	4,4	5	0,9	-1
	Relevancia	0	1	1	0	6	4,4	5	1,2	-2
6d	Claridad	1	0	3	0	4	3,8	5	1,5	-1
	Pertinencia	0	1	0	1	6	4,5	5	1,1	-2
	Relevancia	1	0	0	0	7	4,5	5	1,4	-3
6e	Claridad	0	0	1	1	6	4,6	5	0,7	-2
	Pertinencia	0	0	0	1	7	4,9	5	0,4	-3
	Relevancia	0	0	0	0	8	5	5	0	###

Ítem	Criterio	Puntuación					Media	Moda	Desv. est.	Coef. de asimetría
		1	2	3	4	5				
7a	Claridad	0	0	1	3	3	4,3	4	0,8	-1
	Pertinencia	0	0	1	1	4	4,5	5	0,8	-2
	Relevancia	0	0	1	0	5	4,7	5	0,8	-2
7b	Claridad	0	0	0	3	4	4,6	5	0,5	-0
	Pertinencia	0	0	0	2	4	4,7	5	0,5	-1
	Relevancia	0	0	0	1	5	4,8	5	0,4	-2
7c	Claridad	0	0	1	2	4	4,4	5	0,8	-1
	Pertinencia	0	0	1	1	4	4,5	5	0,8	-2
	Relevancia	0	0	0	0	6	5	5	0	###
7d	Claridad	0	0	1	1	5	4,6	5	0,8	-2
	Pertinencia	0	0	0	1	6	4,9	5	0,4	-3
	Relevancia	0	0	0	0	7	5	5	0	###
8	Claridad	0	0	2	1	4	4,3	5	1	-1
	Pertinencia	0	0	0	2	5	4,7	5	0,5	-1
	Relevancia	0	0	0	1	6	4,9	5	0,4	-3

Nota: “###” representa datos faltantes debido a que el experto por alguna razón no evaluó dicho ítem.

En esta Tabla 2 se observa que los promedios de las puntuaciones son en su mayoría superiores 4, de igual manera la moda es de 5 (máximo posible) salvo la pertinencia del ítem 5b y la claridad del 7a. Hay una evaluación en general alta, reflejada en los coeficientes de asimetría negativos en cada uno de los ítems. Este primer análisis sugiere que los ítems son apropiados para evaluar la comprensión del gráfico de caja. Además, de forma complementaria, se recurrió al CVC, el

cual ayuda a determinar la validez de cada ítem y del instrumento en su totalidad (Hernández-Nieto, 2002).

En la Tabla 3 se muestra el cálculo del CVC para cada ítem del cuestionario. Como se aprecia, no todos los jueces evaluaron todas las preguntas. Por ejemplo, el experto E2 olvidó evaluar los ítems 7 y 8. Sin embargo, como el CVC se calcula por ítem, la ausencia de algunas evaluaciones no es impedimento para su cálculo.

Tabla 3. Estimación del CVC a partir de la evaluación de los expertos.

Ítem	Jueces								$\Sigma x_{ij}$	J	CVC <sub>i</sub>	Pei	CVC
	1	2	3	4	5	6	7	8					
1a	14	15	15	15	15	13	15	15	117	8	0,975	0,0000000596	0,975
1b	12	14	15	15	15	13	15	15	114	8	0,950	0,0000000596	0,950
2a	14	15	15	15	15	15	13	15	117	8	0,975	0,0000000596	0,975
2b	14	14	15	13	15	12	8	15	106	8	0,883	0,0000000596	0,883
3	14	12	15	13	15	11	15	15	110	8	0,917	0,0000000596	0,917
4	14	15	15	15	14	12	13	15	113	8	0,942	0,0000000596	0,942



Ítem	Jueces								$\Sigma x_{ij}$	J	CVC <sub>i</sub>	Pei	CVC
	1	2	3	4	5	6	7	8					
5a	14	12	15	13	15	11	15	15	110	8	0,917	0,0000000596	0,917
5b	14	11	15	#	14	9	#	15	78	6	0,867	0,0000214335	0,867
6a	14	15	15	13	15	10	15	11	108	8	0,900	0,0000000596	0,900
6b	14	14	15	13	15	10	15	11	107	8	0,892	0,0000000596	0,892
6c	14	14	15	13	15	10	15	11	107	8	0,892	0,0000000596	0,892
6d	14	11	15	13	13	6	15	15	102	8	0,850	0,0000000596	0,850
6e	14	15	15	13	15	14	15	15	116	8	0,967	0,0000000596	0,967
7a	13	#	14	14	15	9	#	15	80	6	0,889	0,0000214335	0,889
7b	14	#	14	14	15	12	#	15	84	6	0,933	0,0000214335	0,933
7c	13	#	15	14	15	11	#	15	83	6	0,922	0,0000214335	0,922
7d	14	#	15	14	15	15	13	15	101	7	0,962	0,0000012143	0,962
8	14	#	15	13	15	12	13	15	97	7	0,924	0,0000012143	0,924

**Nota:** “#” representa datos faltantes debido a que el experto por alguna razón no evaluó dicho ítem.

Los valores de los CVC oscilan entre 0,850 y 0,975, por lo que el nivel de validación y concordancia interna de los ítems es bueno o excelente (Hernández-Nieto, 2002, 2011). El ítem que presentó el resultado más bajo fue el 6d, esto se debe principalmente a que el E6 lo evaluó excepcionalmente mal, considerándolo poco relevante (evaluado con un 1/5) y poco pertinente (2/5). Esta visión no es compartida por los otros expertos. Por ejemplo, el E5 indicó:

*En el caso del ítem d, he valorado con menor puntaje a la claridad porque me ha quedado la duda, pues si hablamos de lo que queda por debajo o encima de la mediana, no podríamos hablar de la mayoría. Pero como no conozco el objetivo, no podría [sic] decir si es adecuado o no. Es decir, si lo que ustedes esperan es ver si los estudiantes se dan cuenta de que no se puede hablar de mayoría, entonces me parece bien (E5).*

La pregunta precisamente busca evaluar qué tan bien conocen el concepto de mediana los futuros profesores de matemática, por lo que se consideró adecuada su inclusión en la versión final del cuestionario. En particular, Hernández-Nieto (2002) recomienda mantener únicamente los ítems con un CVC mayor a 0,80, como se puede apreciar en la Tabla 3, todos los ítems cumplen con esa condición, razón por la cual se consideró pertinente aceptarlos todos.

El CVC total del instrumento fue de 0,920. Según Hernández-Nieto (2002, 2011), el CVC mide el grado de acuerdo entre los jueces evaluadores sobre cada ítem y el cuestionario en general, el resultado encontrado sugiere una validez y concordancia excelente. Ello significa, de acuerdo con Cohen y Swerdlik (2006), que el cuestionario está conformado por una muestra representativa de la población de ítems que se podrían utilizar para estimar la comprensión del

gráfico de caja en futuros profesores de Educación Básica. Un tercer momento de análisis de los ítems se hizo a partir de los comentarios y sugerencias formulados por los evaluadores. Se consideraron todos los aportes de los jueces, pero a continuación se exponen los más relevantes y que colaboraron con la mejora de los ítems del cuestionario.

**Ítem 1.** El primer ítem del cuestionario tuvo calificaciones de los expertos que fluctúan entre 4,8 y 5,0 en claridad, 4,6 y 4,8 en pertinencia y de 4,9 en relevancia, además de un CVC en promedio de 0,963. Pese a tener una alta calificación en los tres conceptos evaluados, se modifica según los comentarios de los siguientes expertos:

*En el ítem a, “estudiante” debe ir en plural. (E2)*

*Sugiero arreglar la redacción entrego algunas ideas: En base a la representación gráfica dada responda las siguientes preguntas. Eso de registra... no sería más adecuado hacer notar que REPRESENTA o RESUME en un gráfico el tiempo en minutos que demoran sus estudiantes. (E6)*

*Sugiero cambiar por “información” por “datos” en el enunciado. Debe decir “Qué porcentaje de estudiantes tardó”. (E7)*

Se acogieron los comentarios de los expertos E6 y E7 y se modificó el ítem. En la parte a) se cambió “estudiante” por “estudiantes”, de acuerdo con lo sugerido por E5 y E7.

**Ítem 2.** Este ítem obtuvo calificaciones de media mayor a 4,2 y moda 5, siendo la claridad la dimensión peor evaluada. El CVC del ítem en promedio fue de 0,929. Algunos comentarios de los expertos fueron:

*En lugar de “círculos negros”, escribiría “puntos”. (E1)*

*Me parece interesante hacer la mirada vertical de este tipo de gráfico claro que si se trata de alumnos de enseñanza básica les puede complicar atendiendo que leemos de izquierda a derecha luego su lectura horizontal es más atendible. (E6)*  
*Sería más adecuado hablar de gráfico o de diagrama. A lo mejor puede generar algo de confusión utilizar ambos términos en el mismo ejercicio. ... mide la altura de cada uno, en centímetros, y las representa en el siguiente gráfico. (E8)*

Se hicieron cambios en la redacción, corrigiendo algunos errores identificados por los expertos. Además, se modificaron algunos vocablos por otros considerados más apropiados. Por ejemplo, se cambió la palabra “diagrama” por “gráfico”, siguiendo las indicaciones de E8 y “círculos” por “puntos”, según lo sugerido por E1.

**Ítem 3.** Este ítem fue evaluado con una media de 4,1 en claridad, 4,8 en pertinencia y 4,9 en relevancia. La moda fue 5 en todos los apartados. El CVC fue de 0,917. Si bien fue un ítem con valoración alta por parte de los expertos, algunos comentarios fueron considerados para su mejora.

*Propuesta de enunciado: Una empresa quiere conocer el nivel de satisfacción de sus clientes con relación al servicio entregado. Para tal fin, pide a 100 de los clientes que completen una encuesta de satisfacción. Tras el análisis de los datos se obtuvo que la puntuación mínima fue de 1, mientras que la máxima fue de 80. Sabiendo además que los valores de los cuartiles 1, 2 y 3 fueron 10, 30 y 60 puntos respectivamente, represente dicha información en un gráfico adecuado.* (E8)

El E8 entrega una sugerencia de enunciado considerada pertinente para los objetivos del ítem, por lo que se toma parte de sus sugerencias para la redacción del problema. Otro comentario destacado es el del E6, quien sugiere cambiar “notas” por “puntajes”, sugerencia considerada en los cambios.

**Ítem 4.** El ítem 4 fue evaluado por los expertos con una media en claridad de 4,6, en pertinencia de 4,8 y en relevancia de 4,8. La moda fue de 5 en los tres apartados y el CVC obtenido fue de 0,942. Los cambios a este ítem fueron motivados por los siguientes comentarios:

*Los resultados de la prueba de matemática aplicada a los 19 estudiantes de sexto básico se muestran a continuación.* (E4)  
*Construya un gráfico de caja que represente las notas obtenidas en la prueba por este grupo de estudiantes. No es necesario escribir lo último.* (E7)

Se incorpora “se muestran a continuación” como sugiere el E4, además se elimina “por este grupo de estudiantes”, como indica el E7.

**Ítem 5.** Las puntuaciones dadas por los expertos en claridad (4,1 de media y moda 5), pertinencia (4,7 de media y moda 5) y relevancia (4,7 de media y 5 de moda) son altas. El ítem tuvo en promedio un CVC de 0,892. Estas calificaciones fueron acompañadas de algunos comentarios que fueron utilizados para hacer mejoras al ítem.

Se cambiaron algunas palabras para hacer la pregunta más comprensible y evitar confusiones. Como sugirió el E4, se modificó la palabra habitación por dormitorio. Además, se cambió la frase “niños encuestados” por “estudiantes encuestados” siguiendo el comentario de E7. Por último, en respuesta al comentario de E8, se cambió el gráfico propuesto por uno con barras más separadas, de forma de dejar en claro que el gráfico no es un histograma.

**Ítem 6.** Este ítem está compuesto por cinco preguntas, cada una de ellas evaluada con una moda de 5. Los promedios fluctúan entre 4,3 y 4,8 y el CVC total del ítem fue de 0,9.

Se hicieron modificaciones leves al enunciado del problema en vista de los comentarios de E6 y E7. Las preguntas a), b), c) y d) no sufrieron modificaciones, mientras que e) se mejoró siguiendo lo comentado por E1, que sugirió agregar un

“contexto” más amplio a la pregunta. A E8 le llamó la atención la forma directa de preguntar en los ítems a), b) y c), pero se decidió mantenerlas como estaban, pues el preguntar de distintas formas podría ayudar a obtener información sobre posibles errores que puedan estar cometiendo los futuros profesores al momento de responder.

**Ítem 7.** Este ítem está compuesto de cuatro preguntas. Fue evaluado por los expertos con una media de 4,5 en relación a la claridad, de 4,7 en pertinencia y 4,9 en cuanto a la relevancia; todas con moda de 5. El CVC total del ítem fue de 0,926.

Algunas sugerencias de los evaluadores permitieron mejorar la redacción de los ítems, de manera de hacerlos más claros para los futuros encuestados. Siguiendo lo señalado por E3 y E5, se quitó “De acuerdo con los gráficos” de la pregunta 7a, pues es redundante. Se cambió la redacción de 7b y 7c siguiendo las sugerencias de algunos de los expertos. Además, se corrigió un error de transcripción señalado por E6, cambiando “los mostrado” por “lo mostrado”.

**Ítem 8.** Este ítem fue evaluado en claridad con una media de 4,3 y 5 de moda, en pertinencia con 4,7 de media y 5 de moda y relevancia con 4,9 de media y 5 de moda. El CVC fue de 0,924.

En vista de lo bien evaluado y los comentarios de los expertos, el ítem este sólo se cambió en términos de redacción, manteniendo la idea de la

pregunta y los gráficos propuestos. Por otro lado, algunos comentarios de los expertos no fueron considerados para la modificación de los ítems, principalmente porque a estos se da respuesta en otros apartados del instrumento. Por ejemplo, E3 sugiere una pregunta de nivel superior al de tareas de procedimiento sin conexión, pero ellas están consideradas en los ítems 6e en adelante. Lo mismo aplica para comentarios de los expertos E6 y E8, ya que en otros ítems hay preguntas que abordan dicha sugerencia. El E5 sugiere la inclusión de preguntas que aborden los datos atípicos en el gráfico de caja, sin duda un aspecto valioso, pero fue desdeñado ya que la identificación y análisis de datos atípicos es un contenido no abordado en enseñanza básica, en consecuencia, no fue incorporado al cuestionario.

*Sugiero hacer otra pregunta en el que se aplique el nivel 3, para aprovechar conocer el entendimiento del diagrama de caja, por parte del estudiante. (E3)*

*El ítem cumple con los requisitos de claridad, pertinencia y relevancia. En este último requisito he puesto menor valoración, sólo porque creo que podría presentarse un lote de datos que tenga algún valor atípico, de forma que no quede la idea de que los bigotes inician y concluyen siempre en el mínimo y máximo respectivamente. Por favor, es sólo una idea que dependerá de si los estudiantes de esta edad conocen cómo es el procedimiento para determinar los valores alejados. (E5)*

*Sugiero incluir una pregunta acerca de la información que entrega el gráfico.*

*Dado que de lo contrario estaríamos potenciando lo procedimental algo que debemos erradicar del sistema escolar. (E6)*

*Sería interesante preguntar por alguno de los cuartiles, por ejemplo, ¿Cuál fue la calificación máxima que obtuvo el 50% de los estudiantes de clase? De esta forma se solicita también la interpretación del gráfico, y no solamente se centra en la representación. (E8)*

La versión final del cuestionario, luego de atender a las sugerencias dadas por los expertos, se incluye en el anexo.

## CONCLUSIONES

Un aspecto fundamental en la investigación es la recolección de datos, para ello se debe contar con instrumentos válidos y confiables. Es por esto que es importante continuar desarrollando investigaciones que avancen en la producción de instrumentos de medición consolidados. En este trabajo se expuso el proceso a seguir para el diseño y validación, mediante juicio de experto, de un cuestionario para evaluar la comprensión del gráfico de caja para futuros profesores de Educación Básica.

El cuestionario elaborado pasó por diversos filtros para lograr su validez de contenido. Además de las medidas estadística descriptivas que se usan en estos casos, se usó el CVC ya que es un coeficiente utilizado en otras disciplinas con éxito (Balbinotti et al., 2007; Goncalves Bos et al., 2017), pero no se ha utilizado constantemente en

educación. Si bien en este caso no fue necesario eliminar preguntas, posiblemente por la depuración previa a la que había sido sometido el cuestionario y los altos CVC obtenidos por las preguntas propuestas, es común que el cálculo de coeficientes conlleve a la reducción de la cantidad de preguntas en la versión final del instrumento.

El instrumento presentado es de relevancia pues no se encontraron en la literatura cuestionarios validados para medir la comprensión del gráfico de caja por parte de futuros profesores de Enseñanza Básica. Como su contenido fue validado, puede ser un insumo adecuado para aplicar en este grupo. Eso significa que los ítems que conforman el cuestionario son representativos y relevantes de un marco conceptual específico (Setiawan et al., 2020). En este caso, según la taxonomía de exigencias cognitivas y tareas seleccionada para evaluar la comprensión del gráfico de caja. No obstante, es necesario continuar con el proceso y hacer una prueba piloto para estimar la confiabilidad del cuestionario.

Finalmente, un instrumento de medición solo es válido para medir un propósito específico, pero lograrlo es importante ya que ayuda a conseguir datos con mayor precisión, lo cual se traduce en un menor error de varianza porque el error de medición es bajo. En el caso de la matemática en general, y de la estadística en particular, es relevante ya que mejorar la enseñanza y diseñar estrategias de aprendizaje plausibles, pasa por tener información válida y confiable del logro de aprendizaje de los estudiantes.

**CONFLICTO DE INTERESES.** Los autores declaran que no existe conflicto de intereses para la publicación del presente artículo científico.

## REFERENCIAS

- Alcaraz, R., Ribera, M., y Granollers, T. (2020). La accesibilidad de los gráficos estadísticos para personas con baja visión y visión cromática deficiente: revisión del estado del arte y perspectivas. *Interacción. Revista digital de AIPO*, 1(1), 59-75. <https://repositori.udl.cat/handle/10459.1/68293>
- Arteaga, P., Batanero, C., Cañadas, G.R., y Contreras, J.M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números*, 76, 55-67.
- Arteaga, P., Batanero, C., Contreras, J.M., y Cañadas, G.R. (2016). Evaluación de errores en la construcción de gráficos estadísticos elementales por futuros profesores. *Revista Latinoamericana de Investigación en Matemática Educativa*, 19(1), 15-40. <https://dx.doi.org/10.12802/relime.13.1911>
- Balbinotti, M., Benetti, C., y Soares, P. (2007). Translation and validation of the Graham-Harvey survey for the Brazilian context. *International Journal of Managerial Finance*, 3(1), 26-48. <https://doi.org/10.1108/17439130710721644>
- Batanero, C. (2004). Los retos de la cultura estadística. *Yupana*, 1(1), 27-37. <https://doi.org/10.14409/yu.v1i1.238>
- Benedicto, C., Jaime, A., y Gutiérrez, A. (2015). Análisis de la demanda cognitiva de problemas de patrones geométricos. En C. Fernández, M. Molina y N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 153- 162). SEIEM.
- Carvalho, M. J., Fernandes, J. A., y Freitas, A. (2019). Construção e interpretação de diagramas de extremos e quartis por alunos portugueses do 9º ano de escolaridade. *Bolema: Boletim de Educação Matemática*, 33, 1508-1532. <https://doi.org/10.1590/1980-4415v33n65a25>
- Cohen, R., y Swerdlik, M. (2006). *Pruebas y evaluación psicológicas: Introducción a las pruebas y a la medición* (6ta ed.). Mc Graw Hill.
- Cuétara, Y., Salcedo, I., y Hernández, M. (2016). La enseñanza de la estadística: antecedentes y actualidad en el contexto internacional y nacional. *Atenas*, 3(35), 125-140.
- del Pino, G., y Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo*, 49(1), 53-64. <https://doi.org/10.7764/PEL.49.1.2012.5>
- Díaz-Levicoy, D., Batanero, C., Arteaga, P., y Gea, M.M. (2016). Gráficos estadísticos en libros de texto de Educación Primaria: un estudio comparativo entre España y Chile. *Boletim de Educação Matemática*, 30(55), 713-737. <https://doi.org/10.1590/1980-4415v30n55a20>
- Díaz-Levicoy, D., Osorio, M., Arteaga, P., y Rodríguez-Alveal, F. (2018). Gráficos estadísticos en libros de texto de matemática de Educación Primaria en Perú. *Bolema: Boletim de Educação Matemática*, 32(61), 503-525. <https://doi.org/10.1590/1980-4415v32n61a10>
- Edwards, T. G., Özgün-Koca, A., y Barr, J. (2017). Interpretations of boxplots: helping middle school students to think outside the box. *Journal of Statistics Education*, 25(1), 21-28. <https://doi.org/10.1080/10691898.2017.1288556>
- Espinel, M. C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. En M. Camacho, P. Flores, M.P. Bolea (Eds.), *Investigación en Educación Matemática* (pp. 99-120). SEIEM.
- Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25. <https://doi.org/10.2307/1403713>
- Goncalves Bos, A., Morsch, P., Myskiw, M., y Myskiw, J. (2017). Development and validation of a questionnaire to assess older adult's perception about fall risks. *Journal of Gerontology & Geriatric Research*, 6(2), 1000412. <https://doi.org/10.4172/2167-7182.1000412>
- Haynes, S. N., Richard, D. C. S., y Kubay, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment*, 7(3), 238-247.

- Hernández, R., Fernández, C., y Baptista, P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill.
- Hernández, R., y Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw Hill.
- Hernández-Nieto, R. (2002). *Contributions to statistical analysis*. Universidad de Los Andes.
- Hernández-Nieto, R. (2011). *Instrumentos de recolección de datos en Ciencias Sociales y Ciencias Biomédicas*. Universidad de los Andes.
- Juárez-Hernández, L. G., y Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Espacios*, 39(53), 23. <https://www.revistaespacios.com/cited2017/cited2017-23.html>
- MINEDUC. (2009). *Objetivos fundamentales y contenidos mínimos obligatorios de la Educación Básica y Media. Actualización 2009*. MINEDUC. <http://bibliotecadigital.mineduc.cl/handle/20.500.12365/270>
- MINEDUC (2022). *Estándares Disciplinarios Educación General Básica Matemática*. CPEIP.
- Monteiro, C., y Ainley, J. (2007). Investigating the Interpretation of Media Graphs among Student Teachers. *International Electronic Journal of Mathematics Education*, 2(3), 187-207. <https://doi.org/10.29333/iejme/183>
- NCTM (2000). *Principles and standards for school mathematics*. NCTM.
- Retnawati, H., Kartowagiran, B., Arlinwibowo, J. y Sulistyaningsih, E. (2017). Why are the Mathematics National Examination Items Difficult and What Is Teachers' Strategy to Overcome It? *International Journal of Instruction*, 10(3), 257-276. <https://doi.org/10.12973/iji.2017.10317a>
- Rodríguez-Alveal, F., y Díaz-Levicoy, D. (2019). Evaluación de la alfabetización gráfica del profesorado de Educación Básica en formación y en activo. *Educar em Revista*, 35, 85-103. <https://doi.org/10.1590/0104-4060.68977>
- Salcedo, A. (2015). Exigencia cognitiva de las actividades de estadística en textos escolares de Educación Primaria. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López-Martín (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2 (pp. 307-315). Universidad de Granada.
- Salcedo, A., González, J., Sarco, A., y González, J. (2021). Maestros en formación leen e interpretan gráficos estadísticos. *Jornal Internacional de Estudos em Educação Matemática*, 13, 374-384. <https://doi.org/10.17921/2176-5634.2020v13n4p374-384>
- Stein, M. K., Smith, M. S., Henningsen, M., y Silver, E. (2000). *Implementing standards-based mathematics instruction: a casebook for professional development*. Teachers College Press.
- Setiawan, A., Pusporini, W., y Dardjito, H. (2020). Observation instrument for student social attitude in primary schools: validity and reliability. *Jurnal Penelitian dan Evaluasi Pendidikan*, 24(1), 76-87. <https://doi.org/10.21831/pep.v24i1.31868>