

COMPRENSIÓN DEL MUESTREO POR ESTUDIANTES CHILENOS DE EDUCACIÓN SECUNDARIA

KAREN RUIZ REYES
Universidad de Granada
karenruizreyes@gmail.com

JOSÉ MIGUEL CONTRERAS GARCÍA
Universidad de Granada
jmcontreras@ugr.es

RESUMEN

En inferencia estadística se reconoce la importancia del muestreo como uno de sus conceptos clave, lo que ha permitido su incorporación a nivel internacional en diferentes lineamientos curriculares y específicamente en el currículo chileno, dado que las primeras nociones de muestreo se introducen en el séptimo año de educación primaria. En este trabajo se presenta el análisis de las respuestas a un cuestionario de respuesta abierta, diseñado para evaluar la comprensión del muestreo, que fue aplicado a una muestra de 1241 estudiantes de secundaria chilenos, de octavo año de primaria, segundo año de secundaria y cuarto año de secundaria, en seis centros educativos. Se empleó una metodología mixta, con una descripción cualitativa de las respuestas y un análisis cuantitativo de sus frecuencias. Los resultados reflejan dificultades destacables en la utilización de los elementos relacionados al muestreo y sus propiedades en las diferentes situaciones problema planteadas. Por ejemplo, los estudiantes distinguen el concepto de muestra en contextos cercanos a sus experiencias; pero cuando se enfrentan a diferentes métodos de muestreo, no son capaces de identificar los sesgos asociados a la selección de muestras. Así, al momento de decidir si una muestra es representativa, identifican sobre todo los casos en que la muestra dada no lo es.

Palabras Clave: Investigación en educación estadística; muestra; muestreo; comprensión; inferencia estadística; educación secundaria

1. INTRODUCCIÓN

Al analizar diversos lineamientos curriculares de países como España, Estados Unidos y Chile (MEC, 2015; CCSSI, 2010; MINEDUC, 2009; 2012; 2015), se puede apreciar que la enseñanza de la estadística se centra, principalmente, conceptos descriptivos durante los primeros niveles de escolaridad, mientras que los temas de inferencia son enseñados en los últimos cursos de secundaria o en el nivel universitario. En la actualidad, algunos investigadores de educación matemática (Batanero, 2013; Makar & Ben-Zvi, 2011; Ben-Zvi, Bakker, & Makar, 2015), destacan la importancia de los fundamentos del razonamiento estadístico, sugiriendo así un rol más amplio y profundo para la estadística dentro de la matemática escolar.

Como señalan Burrill y Biehler (2011), la teoría del muestreo estudia cómo se seleccionan las muestras y sus propiedades, para obtener, a partir de ellas, conclusiones incuestionables. Llegar a comprender el muestreo como un concepto fundamental en la inferencia es indispensable para estudiar el contraste de hipótesis y los intervalos de confianza.

El Marco Curricular Chileno (MINEDUC, 2009; 2012; 2015) establece que la educación primaria dura ocho cursos (1° a 8° Básico); mientras que la educación secundaria dura cuatro cursos (1° a 4° Medio). En los primeros cursos de educación primaria, se estudian temas de estadística descriptiva, por lo que términos como *colección* o *grupo de datos*, se utilizan como sinónimos de muestreo, para describir al conjunto que se desea estudiar, pero sin vinculados a una población. El concepto de muestra se aborda formalmente como subconjunto de la población a partir del séptimo año de enseñanza primaria (12-13 años). Además, se estudia la idea intuitiva de la representatividad de la muestra, su aleatoriedad y la posibilidad de estimar resultados a partir de ella.

En el segundo año de enseñanza secundaria (2° Medio, 15-16 años) se estudia el tamaño muestral y las diferentes técnicas de conteo (permutaciones, variaciones y combinaciones) en el sorteo al azar, con o sin reposición. Finalmente, en el cuarto año de enseñanza secundaria (4° Medio, 17-18 años), los estudiantes realizan conjeturas sobre el tipo de distribución a la que tienden las medias muestrales \bar{x}_n (de tamaño n) cuando n aumenta, con lo que se introduce la noción de distribución muestral y su relación con la distribución normal (MINEDUC, 2015). Ruiz-Reyes, Begué, Batanero y Contreras (2017) exponen una descripción más amplia y detallada de los conceptos relativos al muestreo presentes en el currículo chileno, además de realizar una comparación con otras directrices curriculares a nivel internacional.

A pesar de la importancia otorgada al muestreo como uno de los conceptos elementales de la estadística inferencial, Ben-Zvi, Bakker y Makar (2015) consideran que se ha prestado menos de atención a la investigación sobre este tema en comparación con otros conceptos estadísticos. Por lo tanto, teniendo en cuenta esa información y la temprana presencia del concepto de muestra en el currículo chileno, surgió la necesidad de realizar un estudio exploratorio para evaluar la comprensión del muestreo con estudiantes de secundaria, y proporcionar información en el contexto chileno, donde no se han realizado estudios previos sobre este tema.

Este artículo se centra en la pregunta de investigación: ¿Qué comprenden los estudiantes chilenos de secundaria sobre muestra y muestreo? Para dar respuesta a esta interrogante, se elaboró y aplicó un cuestionario de seis ítems de respuesta abierta (ver Anexo). Los resultados proporcionados por una muestra de 1241 estudiantes chilenos de educación secundaria, se detallan en la sección Resultados de este documento.

2. ANTECEDENTES

Watson y Moritz (2000) demuestran que los estudiantes de 3° grado poseen nociones muy elementales y particulares de las muestras, concretamente de ideas derivadas de experiencias de la vida cotidiana con productos de muestra (por ejemplo, en el supermercado). Además, los estudiantes obtienen conclusiones sobre la población basadas en los datos de muestras muy pequeñas, sin preocuparse por el sesgo. Los alumnos de 6° grado formulan variadas creencias sobre el tamaño de la muestra y el método de muestreo. Por el contrario, la mayoría de los estudiantes de 9° grado consideraban la variación en la población, es decir, manifestaban la necesidad de contar con una muestra suficientemente grande y representativa. Sin embargo, a menudo no fueron capaces de identificar el sesgo asociado a la muestra.

Un estudio del razonamiento de los alumnos de secundaria sobre el muestreo desarrollado por Watson (2004), en donde se realizaron entrevistas longitudinales a 38 alumnos durante 3 o 4 años, de los grados 6° al 12° (11-18 años), evidencia que los estudiantes prefieren elegir muestras tomadas con algún procedimiento sesgado, es decir, optan por escoger las muestras que son entregadas voluntariamente, en lugar de elegir muestras aleatorias. Las respuestas de los estudiantes, en general, mejoran gradualmente desde la primera a la última entrevista.

Watson (2004) y Harradine, Batanero y Rossman (2011) sostienen que los estudiantes tienen dificultades con la idea de la variabilidad en las poblaciones, tienen demasiada confianza en muestras pequeñas y no consideran la importancia del tamaño de la muestra en las muestras aleatorias.

El estudio realizado por Meletioui-Mavrotheris y Papanastasiou (2015) analiza el razonamiento inferencial informal de 69 alumnos de 6° grado (11 años) sobre los siguientes conceptos: muestra, tamaño de la muestra, método de muestreo y sesgo. El 85% de los estudiantes habían escuchado la palabra muestra en contextos no relacionados a la escuela (tienda, farmacia, etc.), sin embargo, casi la mitad de ellos (46%), no explicó el concepto o dieron alguna respuesta incompleta. El 40% proporcionó definiciones que reflejaban la idea de una parte, sin el todo asociado que caracteriza la relación de muestra - población, o como una prueba de algo; sólo el 12% dieron respuestas más elaboradas y reconocieron la relación parte-todo entre la muestra y la población.

3. MARCO TEÓRICO

Para analizar las respuestas de los estudiantes al cuestionario se utilizaron los lineamientos propuestos por Zieffler, Garfield, Delmas, y Reading (2008). Definen el Razonamiento Inferencial Informal (RII) como:

El modo en que los estudiantes usan su conocimiento estadístico informal para hacer argumentos para apoyar inferencias sobre poblaciones desconocidas basadas en muestras observadas. (p. 44)

Además, estos autores delimitan tres componentes del RII a considerar:

- (1) Hacer juicios, afirmaciones o predicciones sobre poblaciones basadas en muestras, pero sin utilizar procedimientos y métodos estadísticos formales (por ejemplo, valor p, pruebas t); (p. 45)

Por ejemplo, si el estudiante hizo inferencias razonables sobre una o más poblaciones basadas en una o más muestras. (p. 53)

- (2) Aprovechar, utilizar e integrar conocimiento previo [...], en la medida en que este conocimiento esté disponible; (p. 45)

Por ejemplo, cómo utilizó e integró el estudiante los conocimientos informales (es decir, el conocimiento cotidiano del contexto del problema, el conocimiento previo de los conceptos estadísticos, el conocimiento y la experiencia del mundo real y el lenguaje estadístico) al hacer las inferencias. (p. 53)

- (3) Articular argumentos basados en evidencia para juicios, afirmaciones o predicciones sobre poblaciones basadas en muestras. (p. 45)

Por ejemplo, cómo ha utilizado el estudiante la evidencia para apoyar sus argumentos en la realización de inferencias, y también, qué tan bien la evidencia utilizada apoyó las inferencias realizadas. (p. 53)

4. METODOLOGÍA

En este estudio se utilizó una metodología mixta (Hernández, Fernández y Baptista, 2014), que incluye un componente cualitativo, ya que se describen las respuestas entregadas por los estudiantes; y un análisis cuantitativo descriptivo, específicamente mediante tablas que resumen las principales variables consideradas en este estudio.

Se aplicaron seis ítems de respuesta abierta (ver Anexo) - validados por expertos - que se basaron en algunas de las investigaciones previas descritas anteriormente en la sección de Antecedentes. En la Tabla 1, asignamos la referencia a la fuente de la cual se adaptó cada pregunta, incluyendo las preguntas de elaboración propia.

Se establecieron cinco grupos de preguntas (ver Tabla 1), que tenían por objetivo medir alguno de los aspectos relacionados al muestreo. En total, el cuestionario tenía 20 preguntas distribuidas entre los cinco grupos de objetivos propuestos.

Tabla 1. Investigación de referencia

Grupo	Objetivo	Ítem	Fuente
1	Comprender la definición del concepto de muestra	1	Watson (2004);
		2	Meletiou-Mavrotheris y Papanastasiou (2015)
2	Sugerir un método de muestreo	3	Elaboración propia
3	Decidir si una muestra es representativa	4	Del Valle, Muñoz y Santís (2013), p. 247
4	Identificar la muestra y la población	5	Muñoz, Rupin y Jimenez (2013), p. 275
5	Obtener todas las muestras posibles de una población finita	6	Elaboración propia

A partir de la revisión de expertos realizada, para cada pregunta se han establecido las categorías de análisis a priori, en función a la respuesta correcta esperada a cada pregunta planteada en los seis

ítems que conforman el cuestionario. Éstas se describen a continuación, y sus características principales se detallarán en tablas de frecuencia de la sección Resultados.

En el Ítem 1 (ver Anexo), se espera que el estudiante sea capaz de explicar en sus palabras lo que se entiende por “muestra” y mencionar cualquier contexto en que se haya escuchado el concepto. La respuesta se considera correcta si el estudiante señala alguna de las siguientes tres situaciones: (1) *Sí. Es una parte de un conjunto más grande*; (2) *Sí. Da ejemplos en contextos, pero no explica significado*; (3) *Sí. Cuando se da a probar algo*.

En el Ítem 2 (ver Anexo), se les solicita a los estudiantes que apliquen el concepto de muestra y su representatividad en una situación contextualizada, con el propósito de evaluar si comprenden el concepto de muestra. Las respuestas correctas para la pregunta 1 incluyen la caracterización de una muestra como “*un subconjunto de la población*” o que corresponde a “*un grupo de estudiantes de secundaria chilenos*”. Para la pregunta 2, algunas razones que se consideran correctas pueden incluir: “*porque es difícil encuestar a todos*”; “*porque es un proceso costoso*”; “*porque lleva mucho tiempo*”. Finalmente, para la pregunta 3, la respuesta correcta para la primera parte es “*No, no estoy de acuerdo con que seleccionen sólo 10 estudiantes*”; y para la segunda parte deben indicar el tamaño de la muestra que elegirían.

El Ítem 3 (ver Anexo) pide a los estudiantes que sugieran un método de muestreo, conocido el tamaño de la muestra, pero con el tamaño de la población desconocida. La respuesta correcta es *mencionar algún método de muestreo y justificar dicha elección*.

El Ítem 4 (ver Anexo) se solicita determinar si la muestra señalada es o no representativa; en la segunda parte, los estudiantes deben justificar su elección. En la Tabla 2 se presentan las respuestas correctas del Ítem 4.

Tabla 2. Respuestas correctas Ítem 4

Muestra	¿Es representativa?	Explica por qué
1	Sí	Porque se eligieron al azar.
2	No	Porque se pregunta solo a un sector de usuarios.
3	No	Porque no se considera la variabilidad.

En el Ítem 5 (ver Anexo) los estudiantes deben ser capaces de identificar la población y la muestra respectiva para cada uno de los conjuntos de datos proporcionados. Las respuestas correctas de este ítem se presentan en la Tabla 3.

Tabla 3. Respuestas correctas Ítem 5

Población	Muestra
1. Todos los yogures producidos.	1. Un número determinado de cada sabor.
2. Los precios de los tipos de carne de la carnicería.	2. Los precios de una variedad de cada tipo de carne (vacuno, pescado, ave)
3. Todas las hormigas en el insectario.	3. Cierta cantidad de hormigas del insectario.
4. Todas las ciudades del país.	4. Una ciudad por región.

El Ítem 6 (ver Anexo) solicita que los estudiantes determinen, todas las diferentes muestras (con y sin) reemplazo, que se obtienen de una población finita de cuatro elementos. En la Tabla 4 se especifican las respuestas correctas del Ítem 6.

Tabla 4. Respuestas correctas del Ítem 6

Muestras con reemplazo	Muestras sin reemplazo
1-2; 2-1; 3-1; 4-1.	1-2; 2-3.
1-3; 2-3; 3-2; 4-2.	1-3; 2-4.
1-4; 2-4; 3-4; 4-3.	1-4; 3-4.

Para todos los ítems, todas las respuestas que no se ajusten a los criterios mencionados, se considerarán como respuestas incorrectas.

4.1. PARTICIPANTES Y CONTEXTO

En este estudio, la muestra estuvo constituida por 1241 estudiantes chilenos; 274 (22.1%) de octavo año de primaria (8° Básico, 13-14 años), 529 (42.6%) de segundo año de secundaria (2° Medio, 15-16 años) y 438 (35.3%) de cuarto año de secundaria (4° medio, 17-18 años). Todos ellos asisten a seis diferentes centros educativos de secundaria de la ciudad de Osorno, Chile, durante el primer semestre del año escolar 2018. De ellos 132 (10.6%) asisten a centros de dependencia particular pagada (privados), 411 (33.1%) a centros particulares subvencionados (concertados) y 698 (56.2%) a centros municipales (públicos). El grupo está conformado por 690 (55.6%) mujeres y 551 (44.4%) hombres.

Los estudiantes respondieron los cuestionarios por escrito. Se les consedieron 45 minutos para responder al cuestionario, como una actividad que formaba parte de sus clases regulares de matemáticas, con la presencia de uno de los investigadores, quien explicó el objetivo del cuestionario y aclaró las posibles dudas que pudieran surgir sobre la manera de completar el cuestionario.

Cabe señalar que, al momento de aplicar el cuestionario, no se preguntó a los profesores a cargo de cada clase, ni a los estudiantes que participaron en este estudio, si éstos habían recibido instrucción previa sobre muestras o muestreo, ya que este aspecto no formaba parte del objetivo de nuestra investigación. Como se mencionó anteriormente, aunque el currículum chileno incluye el concepto de muestra en el séptimo grado de educación primaria, no podemos afirmar que los estudiantes participantes de este estudio han recibido instrucción previa en este tema.

5. RESULTADOS

Los resultados del estudio se describen mediante el análisis de las respuestas de los estudiantes a cada uno de los ítems del cuestionario (ver Anexo). Se presentan las ideas que expresan los estudiantes y también los porcentajes de respuesta correcta e incorrecta por pregunta.

La diversidad de respuestas proporcionadas para el *Ítem 1* (ver Anexo), se detallan en la Tabla 5. Un porcentaje muy bajo no responde (1.5%) o no ha escuchado hablar de ese concepto (2.1%), mientras que un alto porcentaje (96.4%) confirma haber oído la palabra muestra anteriormente. Por otra parte, cuando responden que *una muestra es una parte de un conjunto más grande*; se observa que el porcentaje de respuesta aumenta paulatinamente a medida que aumenta la edad de los estudiantes. Además, la mayoría de los estudiantes afirman haber oído la palabra muestra, pero sólo un 43.8% es capaz de dar ejemplos en varios contextos sin explicar su significado.

Tabla 5. Análisis respuestas *Ítem 1*

Curso	8° Básico	2° Medio	4° Medio	Total
Respuestas	n (%)	n (%)	n (%)	
No responde	5 (0.4)	8 (0.6)	5 (0.4)	18 (1.5)
No. Nunca	7 (0.6)	15 (1.2)	4 (0.3)	26 (2.1)
Sí. Es una parte de un conjunto más grande	63 (5.1)	134 (10.8)	166 (13.4)	363 (29.3)
Sí. Da ejemplos en contextos, pero no explica significado	133 (10.7)	270 (21.8)	141 (11.4)	544 (43.8)
Sí. Da una Explicación sin fundamento	66 (5.3)	81 (6.5)	106 (8.5)	253 (20.4)
Sí. Cuando se da a probar algo	0 (0)	21 (1.7)	16 (1.3)	37 (3)
Total	274	529	438	1241

Al comparar las respuestas por curso (ver Tabla 6), se observa que el mayor porcentaje de respuesta correcta corresponde a los estudiantes de 2° año de secundaria (34.2%), seguidos por los alumnos de 4° año de secundaria (26%) y por último los alumnos de 8° año de primaria (15.8%).

En resumen, 944 estudiantes (76.1%) dan una respuesta correcta al indicar que *han escuchado la palabra muestra en algún contexto* (centro de salud, supermercado, tienda, farmacia, etc.), pero no

necesariamente en la clase de matemática; lo que indica que sus respuestas están basadas en aspectos estadísticos informales. Por otra parte, sólo 297 (23.9%) dan una respuesta incorrecta, al no responder o dar ejemplos en contextos poco claros.

Tabla 6. Respuestas correctas e incorrectas ítem 1

Curso	8° Básico	2° Medio	4° Medio	Total (%)
Respuestas	n (%)	n (%)	n (%)	
Incorrectas	78 (6.3)	104 (8.4)	115 (9.3)	297 (23.9)
Correctas	196 (15.8)	425 (34.2)	323 (26)	944 (76.1)
Total	274	529	438	1241

En cuanto a la primera pregunta del *Ítem 2* (ver Anexo), los estudiantes manifiestan tener una idea bastante clara del concepto de muestra en un contexto cercano a su propia realidad, dado que sólo 27 (2.2%) no responden, como se ve en la Tabla 7.

Tabla 7. Análisis respuestas *Ítem 2*, pregunta 1

Curso	8° Básico	2° Medio	4° Medio	Total (%)
Respuestas	n (%)	n (%)	n (%)	
No responde	6 (0.5)	15 (1.2)	6 (0.5)	27 (2.2)
Incorrecto	57 (4.6)	70 (5.6)	32 (2.6)	159 (12.8)
Una parte, una pequeña parte	153 (12.3)	289 (23.3)	262 (21.1)	704 (56.7)
Un porcentaje	16 (1.3)	42 (3.4)	30 (2.4)	88 (7.1)
Una cantidad	39 (3.1)	107 (8.6)	102 (8.2)	248 (20)
Un subconjunto de la población	3 (0.2)	6 (0.5)	6 (0.5)	15 (1.2)
Total	274	529	438	1241

Se observa que un 12.8% (159) respondió incorrectamente (ver Tabla 7); mientras que un 83.8% (1040) mencionó ideas como *una pequeña parte*, *un porcentaje*, *una cantidad*, y sólo 15 estudiantes (1.2%) dan una definición más precisa, es decir, un *subconjunto de la población*. En resumen, el 85% de los estudiantes responde correctamente a este ítem.

El análisis de las respuestas por curso muestra que el 16.9% de los estudiantes de 8° año de primaria respondió correctamente esta pregunta; este porcentaje aumenta si se consideran los estudiantes de 2° año de secundaria (35.8%) y, de manera similar, los de 4° año de secundaria (32.2%).

En el caso de la pregunta 2 (ver Tabla 8), el 65.5% (812) de los estudiantes tienen una idea más clara de los factores que son importantes al seleccionar una muestra de una población; por ejemplo, mencionan ideas sobre el *excesivo tiempo* que puede llevar el estudio, el *costo de implementación* de la entrevista o *el acceso para entrevistar a la población total*. Esto significa que un alto porcentaje de los estudiantes toma en consideración los factores que afectan a la selección de muestras representativas; mientras que sólo un 34.5% (429) respondió incorrectamente o no dio ninguna respuesta.

Tabla 8. Análisis respuestas *Ítem 2*, pregunta 2

Curso	8° Básico	2° Medio	4° Medio	Total (%)
Respuestas	n (%)	n (%)	n (%)	
No responde	14 (1.1)	22 (1.8)	14 (1.1)	50 (4)
Incorrecto	104 (8.4)	169 (13.6)	106 (8.5)	379 (30.5)
Porque es difícil preguntar a todos	79 (6.4)	165 (13.3)	145 (11.6)	389 (31.3)
Porque es costoso	2 (0.2)	4 (0.3)	1 (0.08)	7 (0.6)
Porque lleva mucho tiempo	60 (4.8)	118 (9.5)	85 (6.8)	263 (21.2)
Es más rápido o fácil	3 (0.2)	18 (1.5)	41 (3.3)	62 (5)
Porque obtienen los mismos resultados	12 (1)	33 (2.7)	46 (3.7)	91 (7.3)
Total	274	529	438	1241

La principal razón mencionada para seleccionar una muestra es *que es difícil preguntar a todos*

(31.3%); como segunda opción, señalan *que lleva mucho tiempo* (21.2%). Los estudiantes de 2° año de secundaria son los más propensos a indicar estas respuestas, seguidos por los estudiantes de 4° año de secundaria y, por último los alumnos de 8° año de primaria.

En las respuestas a la pregunta 3 del Ítem 2 (ver Tabla 9), se observa que el 19.8% no responde; un 35.6% no está de acuerdo con la elección de una muestra de “*sólo 10 estudiantes*”; y el 44.6% está de acuerdo con la afirmación propuesta.

Tabla 9. Análisis respuestas Ítem 2, pregunta 3 a

Curso	8° Básico	2° Medio	4° Medio	
Respuestas	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	Total (%)
No responde	69 (5.6)	91 (7.3)	86 (6.9)	246 (19.8)
No, no estoy de acuerdo	79 (6.4)	203 (16.3)	160 (12.9)	442 (35.6)
Sí, estoy de acuerdo	126 (10.2)	235 (19)	192 (15.4)	553 (44.6)
Total	274	529	438	1241

Para esta pregunta, sólo el 35.6% responde correctamente. A pesar que, en las preguntas anteriores, los estudiantes evidencian un nivel elemental comprensión de la muestra y las limitaciones de realizar un estudio *preguntándole a todos los individuos* de una determinada población, al momento de cuantificar los sujetos de la muestra sus concepciones no son tan acertadas. Además, podría inferirse que estas respuestas pueden estar influidas por el hecho que la población estudiada es desconocida, lo que dificulta la cuantificación y puede dar lugar a errores de interpretación de la pregunta. Una vez más, se observa que los estudiantes de 2° año de secundaria son los que mejor responden a esta pregunta, seguidos por lo de 4° año de secundaria; y, por último, el los estudiantes de 8° año de primaria.

Si consideramos la segunda parte de la pregunta 3: *¿Cuántos alumnos elegirías tú?* (ver Tabla 10), el 34.3% indica que sólo “*elegiría a 10 alumnos*”, tal como se indica en la primera parte de la pregunta; el 36.8% señala que consideraría “*más alumnos*” dependiendo del *total de alumnos entrevistados*, o del número *total de centros educativos que forman parte del estudio*; y sólo el 9.2% considera que *todos los estudiantes de secundaria chilenos deberían ser entrevistados*, lo que prueba que olvidan las razones entregadas en la pregunta anterior, sobre seleccionar una muestra representativa de la población.

Se observa que el 56.6% de los estudiantes afirma que se debe seleccionar una muestra *mayor a 10 estudiantes* (considerando todas las razones presentadas); aunque no se refieren expresamente a la *representatividad de la muestra*: algunos consideran que debe *ser seleccionada al azar*, o *según el número total de alumnos* entrevistados. De este porcentaje, se observa que el 22.8% de los estudiantes de 2° año de educación secundaria, el 21.3% de 4° año de secundaria y el 12.5% de los estudiantes de 8° año de primaria respondieron correctamente a la pregunta.

Tabla 10. Análisis respuestas Ítem 2, pregunta 3 b

Curso	8° Básico	2° Medio	4° Medio	
Respuestas	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	Total (%)
No responde	15 (1.2)	29 (2.3)	23 (1.9)	67 (5.4)
10 alumnos	91 (7.3)	198 (16)	137 (11)	426 (34.3)
Más alumnos	96 (7.7)	187 (15.1)	174 (14)	457 (36.8)
50%	13 (1.1)	27 (2.2)	21 (1.7)	61 (4.9)
Todos	35 (2.8)	45 (3.6)	34 (2.7)	114 (9.2)
Depende del total	11 (0.9)	22 (1.8)	35 (2.8)	68 (5.5)
Al azar	0 (0)	1 (0.1)	2 (0.1)	3 (0.2)
Menos de 10	13 (1.1)	20 (1.6)	12 (1)	45 (3.6)
Total	274	529	438	1241

Cuando se les solicita a los estudiantes que sugieran: *¿Cómo seleccionarían una muestra?* (con un tamaño de muestra dado), en un contexto cercano a su realidad (Ítem 3, ver Anexo), se pudo observar que el 38.9% de ellos *mencionan un método de muestreo pero no justifican su elección*; y el 5.7% se refieren a la *elección de una muestra sesgada*, por ejemplo, perteneciente a un nivel escolar, o a la consideración de otras características como el sexo o la edad de los participantes (ver Tabla 11).

Tabla 11. Análisis respuestas Ítem 3

Curso	8° Básico	2° Medio	4° Medio	Total (%)
Respuesta	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	
No contesta	32 (2.6)	101 (8.1)	49 (4)	182 (14.7)
Menciona un método de muestreo, pero no justifica	112 (9)	124 (10)	247 (19.9)	483 (38.9)
Menciona obtener una muestra de un grupo sesgado (alguna característica en particular)	8 (0.7)	29 (2.3)	34 (2.7)	71 (5.7)
Usan algún método de muestreo y justifica	122 (9.8)	275 (22.2)	108 (8.7)	505 (40.7)
Total	274	529	438	1241

En resumen, el *Ítem 3* evidencia un alto porcentaje de respuestas incorrectas (59.3%); sólo el 40.7% de los estudiantes responden correctamente, *indicando un método de muestreo y justificando su elección*. Si organizamos los resultados de este ítem por curso, se observa que los estudiantes de 2° año de secundaria (22.2%), seguidos por los estudiantes de 8° año de primaria (9.8%) y, por último, los estudiantes de 4° año de secundaria (8.7%), responden correctamente esta pregunta. Se puede argumentar que, la mayor dificultad para esta pregunta, surge al no indicar explícitamente el número total de estudiantes de la escuela (población desconocida), lo que les impide comprender la pregunta planteada. Además, como los últimos cambios curriculares para el 2° año de secundaria, ya no incluyen los métodos de muestreo, éste podría ser un factor que influye en el alto porcentaje de respuestas incorrectas.

El *Ítem 4* (ver Anexo), se ha adaptado a partir de una actividad extraída de un texto escolar (Del Valle, Muñoz y Santís, 2013; p. 247), en la que los estudiantes deben indicar, si la muestra dada es representativa o no, así como justificar sus respuestas. En la Tabla 12, para resumir los datos se presentan los porcentajes de respuestas correctas.

Tabla 12. Análisis respuestas Ítem 4

Muestra	¿Es representativa?			
	Respuesta Correcta			
Curso	8° Básico	2° Medio	4° Medio	Total (%)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	
1.	118 (9.5)	193 (15.6)	165 (13.3)	476 (38.4)
2.	99 (8)	217 (17.4)	198 (16)	514 (41.4)
3.	114 (9.2)	213 (17.2)	200 (16.1)	527 (42.5)

En la primera muestra, sólo el 38.4% (476 estudiantes) responde correctamente indicando que la muestra es representativa, en una situación que implica un método de *muestreo aleatorio*. En la segunda muestra, sólo el 41.4% (514 estudiantes) responde que la muestra *no es representativa*, porque se trata de un método de muestreo restringido, lo que implica que los sujetos se eligen de manera sesgada y no representativa de la población. Por último, en la tercera muestra, sólo el 42.5% (527 estudiantes) responde que la muestra no es representativa, ya que se trata de un método de muestreo autoseleccionado, lo que demuestra que la selección de la muestra plantea problemas para determinar si es representativa o no, ya que considera a los sujetos con características específicas, sin considerar a “*todos*” los sujetos que conforman la población de estudio.

Se observa que, los estudiantes reconocen mayoritariamente *muestras no representativas*; pero en la primera muestra, no identifican que es un método de *muestreo aleatorio*. Se centran únicamente en el tamaño de la muestra dada, pero no consideran que ha sido seleccionada al azar, lo que da lugar a respuestas incorrectas (61,6%, 765 estudiantes).

Entre las explicaciones para justificar las respuestas a la primera muestra, los estudiantes mencionan lo siguiente: *porque se eligió a un tercio de la población; porque es una muestra aleatoria; se debe elegir a la mitad de la población*. En cuanto a la segunda muestra, señalan razones como: *deberían encuestar a más personas; porque sólo se encuesta a un grupo de usuarios; deberían ser encuestados*

en distintos horarios. Por último, para la tercera muestra, las explicaciones indicadas son: *porque se debe tener en cuenta la diversidad de alturas; no se consideró la variabilidad; se debe medir todos los estudiantes para obtener un promedio.*

El *Ítem 5* (ver Anexo) se ha adaptado de una actividad de un texto escolar (Muñoz, Rupin y Jimenez, 2013; p. 275), en la que los estudiantes deben indicar la población y una posible muestra para cada uno de los enunciados. En la Tabla 13 se presentan los porcentajes de respuestas correctas e incorrectas para facilitar el análisis de los lectores.

Tabla 13. Análisis respuestas *Ítem 5*

	Población Respuesta		Muestra Respuesta	
	Incorrecta <i>n</i> (%)	Correcta <i>n</i> (%)	Incorrecta <i>n</i> (%)	Correcta <i>n</i> (%)
1.	856 (69)	385 (31)	897 (72.3)	344 (27.7)
2.	976 (78.6)	265 (21.4)	974 (78.5)	267 (21.5)
3.	785 (63.3)	456 (36.7)	879 (70.8)	362 (29.2)
4.	1114 (89.8)	127 (10.2)	1149 (92.6)	92 (7.4)

La mayoría de los estudiantes contesta incorrectamente (ver Tabla 13), tanto para identificar la población como la muestra, lo que demuestra que tienen problemas para reconocer estos conceptos en contextos más alejados de la realidad escolar. Además, no identifican la diferencia entre la relación parte-todo en los contextos mencionados en el enunciado del ítem. También surgió la inquietud de analizar la complejidad en la redacción de la pregunta, porque este ítem fue adaptado de una actividad propuesta en un texto escolar.

Si analizamos las respuestas por curso, podemos ver que los estudiantes de 4° año de secundaria obtienen mejores resultados al identificar la población, lo que disminuye en los otros cursos. Lo mismo ocurre para la cuarta muestra, mientras que, en las tres primeras muestras, los que obtienen más respuestas correctas son los estudiantes de 2° año de secundaria, seguidos por los de 4° año de secundaria, siendo los estudiantes de 8° año de primaria los que obtienen resultados más bajos.

Finalmente, en el *Ítem 6* (ver Anexo), sólo el 22.4% respondió correctamente a esta pregunta, escribiendo ambas listas de muestras solicitadas; el 25.8% dio respuestas parcialmente correcta, escribiendo sólo una de las muestras solicitadas, es decir, anotando la muestra con reemplazo o la muestra sin reemplazo; el 14% escribió la muestra con (o sin) reemplazo de forma incompleta, con algunos elementos faltantes; y el 37.8% no respondió (ver Tabla 14).

Tabla 14. Análisis respuestas *Ítem 6*

Curso Respuesta	8° Básico <i>n</i> (%)	2° Medio <i>n</i> (%)	4° Medio <i>n</i> (%)	Total (%)
No contesta	104 (8.4)	221 (17.8)	144 (11.6)	469 (37.8)
Anota listado con reemplazo	70 (5.6)	86 (6.9)	116 (9.4)	272 (21.9)
Anota listado con reemplazo incompleto	24 (1.9)	20 (1.6)	29 (2.3)	73 (5.9)
Anota listado sin reemplazo	10 (0.8)	22 (1.8)	16 (1.3)	48 (3.9)
Anota listado sin reemplazo incompleto	25 (2)	38 (3.1)	37 (3)	100 (8.1)
Anota ambos listados	41(3.3)	142 (11.4)	96 (7.7)	279 (22.4)
Total	274	529	438	1241 (100)

Al analizar los resultados por respuesta correcta y curso, se percibe que un mayor número de estudiantes de 2° año de secundaria responden correctamente (11.4%), seguidos por los estudiantes de 4° año de secundaria (7.7%); y por último, los estudiantes de 8° año de primaria (3.3%).

Como esta pregunta involucra combinaciones y permutaciones, observamos que algunos estudiantes de 4° año de secundaria (12° grado) responden incorrectamente porque confunden la fórmula de la técnica de conteo respectiva que debe utilizarse, ya que no identifican que esta fórmula sólo le da la cantidad de elementos que debe contener cada muestra.

6. DISCUSIÓN

Los hallazgos de este estudio exploratorio, han permitido identificar y caracterizar qué comprenden los estudiantes de secundaria chilenos sobre muestras y muestreo, desde la perspectiva del Razonamiento Inferencial Informal (RII). Los estudiantes, al momento de elaborar sus argumentos, no necesariamente utilizan métodos estadísticos formales, además, sus respuestas pueden (o no) conter conceptos o lenguaje estadístico formal (Zieffler et al., 2008), como se ha expuesto en los análisis presentados en la sección de Resultados. Las respuestas obtenidas por la muestra de 1241 estudiantes de secundaria chilenos, proporcionan valiosa información sobre su comprensión del muestreo, así como sobre las principales dificultades relacionadas a ésta.

Los resultados indican que los estudiantes distinguen el concepto de *muestra* en contextos cercanos a sus experiencias (salud, ciencia, tienda, etc.), siendo muy pocos los que dan una respuesta más elaborada y más relacionada a un contexto matemático; es decir, utilizan el concepto de *aleatoriedad* en sus argumentos. Sólo 15 estudiantes definen muestra como un *subconjunto de la población*. Resultados similares se pueden verificar en los trabajos de Watson y Moritz (2000) y Meletiou-Mavrotheris y Papparistodemou (2015).

Otro aspecto a considerar, es que los estudiantes son susceptibles a la *insensibilidad al tamaño de la muestra* (Tversky & Kahneman, 1974), es decir, asumen que la muestra seleccionada, independientemente de su tamaño, siempre representa a la población a la que pertenece, lo que induce a graves errores de interpretación en los análisis estadísticos. Además, como indican Kahneman, Slovic y Tversky (1982), este tipo de personas, también cree en la *ley de los pequeños números*; en otras palabras, creen que, al seleccionar muestras que no son suficientemente grandes, la distribución de la muestra se distribuye de la misma manera que la distribución de la población, independientemente del tamaño de la muestra.

Por otra parte, cuando se pidió a los estudiantes que *identificaran la muestra de un estudio* en un contexto más cercano a su realidad, se observan algunas incongruencias en sus respuestas. Por ejemplo, pueden identificar las limitaciones de un investigador a intentar “encuestar a todos los sujetos” de una determinada población, pero cuando tienen que sugerir una muestra o preguntarles cuál tamaño de muestra elegirían, el 43.4% de los estudiantes de nuestro estudio, responde incorrectamente.

Cuando intentan *sugerir un método de muestreo*, sólo el 38.9%, dio una respuesta, pero *no justificó* su método de selección. Cabe señalar que la mayoría de los estudiantes no proporcionan un método de muestreo porque en el enunciado del problema (Ítem 3, ver Anexo) no se menciona la cantidad total de estudiantes (*población desconocida*).

Al decidir si una muestra *es representativa*, los estudiantes suelen distinguir los casos en que la muestra dada no lo es; pero en el primer caso, no distinguen que se trata de un *método de muestreo aleatorio*, llevándolos al error de creer que esa muestra no es representativa de la población. En resumen, se centraron en el tamaño de la muestra, pero no reconocieron que *el azar* es un elemento importante que debe considerarse en el método de muestreo. En este sentido, sobre el papel del contexto en el razonamiento del muestreo, Wroughton, McGowan, Weiss y Cope (2013) sugieren que los estudiantes que tienen opiniones firmes sobre un tema, evaluarán la validez de las conclusiones de un estudio basándose en si esa conclusión está o no de acuerdo con su opinión, en lugar de analizar la calidad del método de muestreo utilizado por medio de principios estadísticos.

Cuando se les preguntó que *identificaran la población y la muestra*, los estudiantes respondieron incorrectamente. Creemos que, debido a que los contextos expuestos en esta pregunta se adaptaron a partir de un texto escolar, los cuales eran desconocidos para los estudiantes, lo que impidió una fácil identificación de la relación parte-todo entre los elementos que conforman la población o la muestra de cada caso.

Por último, cuando se les pidió que obtuvieran todas las muestras posibles (con y sin reemplazo), de una población finita, sólo el 22.4% de los estudiantes respondieron correctamente, ya que otro 25.8% anota solamente uno de los listados solicitados.

Como limitaciones de este estudio, podemos señalar que se debe mejorar la redacción del Ítem 5 del cuestionario (ver Anexo), empleando otros contextos más cercanos a las experiencias de los estudiantes.

Por otro lado, sería interesante ahondar en el análisis de los argumentos proporcionados por los estudiantes, utilizando algunas de las categorías mencionadas por Watson (2004), como también, complementar los resultados aquí expuestos con comparaciones entre los distintos tipos de centros educativos.

Para futuras investigaciones se podría considerar el diseño e implementación de una unidad didáctica sobre el muestreo, para promover el Razonamiento Inferencial Informal (RII) empleando los tres componentes sugeridos por Zieffler et al. (2008), como también, incluir otros tipos de tareas sugeridos para trabajar el RII. Por ejemplo, se podrían incorporar actividades con apoyo de simulaciones como “*Growing Samples*”, ya que con este enfoque, se introduce gradualmente a los estudiantes en el crecimiento del tamaño de las muestras que se toman de la misma población; y para cada muestra, se les pide que le den sentido y que hagan una inferencia informal (Ben-Zvi, Bakker & Makar, 2015).

Además, en dicha unidad didáctica se podría incorporar una sesión para la aplicación de este cuestionario; y otra para el análisis de las respuestas en una puesta en común con los estudiantes, como también, ampliar el estudio con algunas entrevistas, para ver más de cerca las ideas que poseen los estudiantes de estos conceptos, pues a veces, no expresan todos sus conocimientos en una tarea escrita.

Sería interesante aplicar del cuestionario en otros contextos latinoamericanos, que nos permitan comparar los hallazgos aquí expuestos.

Así mismo, se podría estudiar la comprensión del muestreo con futuros profesores de matemática, puesto que, al ser el muestreo un concepto básico para la inferencia estadística, y al estar presente en el currículo escolar chileno, es de gran importancia que los futuros profesores de matemática dominen dicho tema.

Considerando que estos resultados, aunque se centran en un contexto particular de educación secundaria, forman un precedente para seguir aportando al campo de investigación en educación estadística en el contexto latinoamericano, y específicamente, en la inferencia estadística informal, ya que esta línea de investigación aún es emergente en el contexto chileno, en comparación con lo ya desarrollado en el ámbito internacional.

AGRADECIMIENTOS

Trabajo realizado en el marco del Proyecto EDU2016-74848-P y la Beca CONICYT Chile (PFCHA 72160521).

REFERENCIAS

- Batanero, C. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática*, 11(8), 277–291.
- Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, 88(3), 291–303.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 57–69). Springer.
- CCSSI. (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices and the Council of Chief State School Officers.
- Del Valle, J., Muñoz, G., & Santís, M. A. (2013). *Matemática 1º Medio*. S.M.
- Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school-mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 235–246). Springer.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. McGraw Hill.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, 13, 1–4.

- MEC. (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. Madrid: Autor.
- Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics*, 88(3), 385–404.
- Merino, R., Muñoz, V., Pérez, B., & Rupin, P. (2015). *Matemática 7º Básico*. Santiago de Chile: S.M. MINEDUC.
- MINEDUC. (2009). *Objetivos Fundamentales y Contenidos Mínimos Obligatorios de la Enseñanza Básica y Media*. Santiago, Chile: Unidad de Currículum y Evaluación.
- MINEDUC. (2012). *Bases Curriculares: Matemática, Educación Básica*. Unidad de Currículum y Evaluación.
- MINEDUC. (2015). *Bases Curriculares: Matemática, Educación Media*. Unidad de Currículum y Evaluación.
- Muñoz, G., Rupín, P., & Jiménez, L. (2013). *Matemáticas 2º Medio Santiago de Chile*: S.M.
- Ruiz-Reyes, K., Begué, N., Batanero, C., & Contreras, J. M. (2017). Un estudio comparado de los contenidos de muestreo en la Educación Secundaria Obligatoria en Chile. *Educação Matemática Pesquisa*, 19(3), 67–83.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Watson, J. (2004). Developing reasoning about samples. In D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277–294). Kluwer.
- Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44–70.
- Wroughton, J. R., McGowan, H. M., Weiss, L. V., & Cope, T. M. (2013). Exploring the role of context in students' understanding of sampling. *Statistics Education Research Journal*, 12(2). <https://doi.org/10.52041/serj.v12i2.303>
- Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58. [https://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Zieffler.pdf](https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf)

KAREN RUIZ REYES
 Departamento de Didáctica de la Matemática
 Facultad de Ciencias de la Educación
 Campus Universitario de Cartuja, s/n. 18071.
 Granada, España.

ANEXO

Ítem 1. ¿Has oído la palabra *muestra* antes? Explica qué significa para ti.

Ítem 2. En una investigación sobre los hábitos de ejercicio de los “estudiantes de secundaria chilenos”, algunos investigadores entrevistaron a una *muestra* de ellos.

1. ¿Qué significa la palabra muestra en esta frase?
2. ¿Por qué piensas que los investigadores seleccionan una *muestra* de estudiantes en vez de preguntar a todos?
3. ¿Estás de acuerdo con que los investigadores seleccionen una muestra de 10 estudiantes?
¿Cuántos elegirías tú?

Ítem 3. Supongamos que quieres averiguar el porcentaje de niños que viene a tu escuela en diferentes medios de transporte: caminando, auto, micro, bicicleta, u otro. ¿Cómo tomarías una muestra de 50 niños de modo que los resultados sean representativos de toda la escuela?

Ítem 4. Analiza si las siguientes muestras son representativas de la población de estudio y completa la tabla con tus respuestas.

<i>Muestra</i>	<i>¿Es representativa?</i>	<i>Explica por qué</i>
1. Se seleccionan al azar 100 estudiantes, de un colegio de 300 alumnos, para calcular el promedio de notas de la población.		
2. Se les pregunta sobre la calidad del servicio del metro a las primeras 100 personas que toman el metro en la mañana.		
3. Se escogen los niños más altos y los más bajos de un curso para determinar el promedio de la estatura de los alumnos.		

Ítem 5. Determina en cada caso la población y una posible muestra de ella. Completa la tabla con tus respuestas.

<i>Datos</i>	<i>Población</i>	<i>Muestra</i>
1. Una fábrica de yogur quiere investigar sobre la calidad de sus productos.		
2. Diego necesita saber el precio de un kilogramo de carne, para una comida familiar.		
3. Ximena estudia respecto del tamaño de las hormigas que habitan en un insectario.		
4. Daniel desea saber si una ciudad cuenta con suficiente lluvia, para realizar una plantación.		

Ítem 6. Tenemos 4 libros numerados 1, 2, 3 y 4. Escribe todas las muestras posibles de dos libros *con* y *sin* reemplazamiento.