



# Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria

## Secondary school students' understanding the expected value and variability of sample proportion

Nuria Begué

*Doctorado en Ciencias de la Educación, Facultad de Ciencias de la Educación, Universidad de Granada*  
nbegue@correo.ugr.es

Carmen Batanero, María M. Gea

*Departamento de Didáctica de la Matemática, Facultad de Ciencias de la Educación, Universidad de Granada*  
batanero@ugr.es, mmgea@ugr.es

**RESUMEN** • Se analiza la comprensión intuitiva de la relación entre la proporción en una población y el valor esperado de la proporción muestral, y de la variabilidad de dicha proporción, en función del tamaño de la muestra, de estudiantes de educación secundaria obligatoria. Se propusieron a 302 estudiantes cuatro ítems en los que se piden cuatro valores probables del número de ocurrencias de un suceso, variando la proporción poblacional y el tamaño muestral. El análisis estadístico de los valores proporcionados por los estudiantes indica una buena comprensión de la relación entre proporción muestral y poblacional. La variabilidad de la proporción muestral se sobreestima en muestras grandes y depende del contexto del problema en muestras pequeñas. Se observan los sesgos de equiprobabilidad, recencia positiva y negativa.

**PALABRAS CLAVES:** muestreo; proporción muestral; variabilidad y valor esperado; comprensión; educación secundaria obligatoria.

**ABSTRACT** • We analyse secondary school students' intuitive understanding of the relationships between the population proportion and the expected value of a sample proportion, as well as its variability in relation to the sample size. We propose to 302 students four items in each of which four probable values for the number of outcomes for a given event are requested and in which the proportion population and sample size are varied. The statistical analysis of the values provided by the students suggests a good understanding of the relationships between the population and sample proportions. The variability of the sample proportion is overestimated in big samples and depends on the problem context in small samples. We also observed the equiprobability, positive and negative recency biases.

**KEYWORDS:** sampling; sample proportion; variability and expected value; understanding; secondary school.

Recepción: abril 2018 • Aceptación: abril 2018 • Publicación: junio 2018

Begué, N., Batanero, C., & Gea, M. (2018). Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria. *Enseñanza de las ciencias*, 36(2), 63-79.

## INTRODUCCIÓN

El muestreo recibe actualmente una gran atención en la investigación didáctica, pues las ideas asociadas a este subyacen en el trabajo con la simulación, cuya utilización en el aula se recomienda para mejorar la comprensión de la probabilidad y la inferencia estadística (Eichler y Vogel, 2014; Huerta, 2015). Este concepto, según Heitele (1975), juega un gran papel en el estudio de ciertos temas de probabilidad (leyes de los grandes números y enfoque frecuencial). Burrill y Biehler (2011) lo consideran como una de las ideas fundamentales en estadística, por ser la base de la inferencia.

Además, nuestro conocimiento sobre el mundo está basado en el muestreo, ya que, usualmente, solo podemos observar una parcela de la realidad en la que estamos interesados. Por otro lado, los resultados obtenidos mediante muestreo en diferentes campos se publican con frecuencia en los medios de comunicación. Por tanto, es importante preparar a los estudiantes para comprender estas noticias, y hacerles conscientes de sus posibles sesgos en el razonamiento sobre muestreo.

El muestreo aparece por primera vez en el currículo, en España, en los dos primeros cursos de la educación secundaria obligatoria (ESO), donde se indican los siguientes contenidos (MECD, 2015):

Población e individuo. Muestra. Variables estadísticas. Variables cualitativas y cuantitativas. Frecuencias absolutas y relativas. Medidas de tendencia central. Medidas de dispersión.

Frecuencia relativa de un suceso y su aproximación a la probabilidad mediante la simulación o experimentación (p. 413).

Para el tercer curso se introducen los siguientes contenidos y estándares de aprendizaje (MECD, 2015):

Fases y tareas de un estudio estadístico. Población, muestra. Métodos de selección de una muestra estadística. Representatividad de una muestra (pp. 394 y 403).

Distingue población y muestra justificando las diferencias en problemas contextualizados. Valora la representatividad de una muestra a través del procedimiento de selección, en casos sencillos (pp. 394 y 403).

El objetivo de este trabajo es evaluar la comprensión de algunas propiedades del muestreo por parte de los estudiantes de ESO. Más concretamente, se evalúa su comprensión intuitiva de la relación entre el valor de la proporción de un suceso en la población y la frecuencia esperada de dicho suceso en muestras tomadas de dicha población (que se deduce de la proporción muestral, es decir, la proporción del mismo suceso en una muestra de la población). Igualmente, analizamos la comprensión de la variabilidad de la proporción en el muestreo y del efecto del tamaño de la muestra sobre dicha variabilidad.

## MARCO TEÓRICO

La comprensión del muestreo requiere relacionar dos ideas aparentemente antagónicas: representatividad y variabilidad muestral (Rubin, Bruce y Tenney, 1991; Saldahna y Thompson, 2002). La representatividad indica que una muestra aleatoria de tamaño suficiente tendrá características parecidas a las correspondientes características en la población, mientras que la variabilidad supone que las muestras pueden cambiar en su composición y características (como ocurre con la proporción muestral).

La comprensión del muestreo exige trabajar con tres tipos de distribuciones que los estudiantes deben comprender y relacionar (Harradine, Batanero y Rossman, 2011):

- La *distribución teórica de probabilidad* que modela los valores de una variable aleatoria en una población y depende de algún *parámetro* o característica estadística en dicha población. En los ítems de nuestro cuestionario se considera una variable aleatoria dicotómica, siendo  $p$  la proporción de elementos que comparten un cierto rasgo (éxitos) dentro de la población.
- La *distribución de datos en una muestra*. Es una parte de la población donde podemos calcular la proporción de éxitos  $\hat{p}$  (que es un resumen estadístico en la muestra) para estimar el parámetro  $p$ . Mientras que dicho parámetro es constante, el valor del estadístico  $\hat{p}$  varía en las diferentes muestras.
- La *distribución muestral del estadístico* o distribución de probabilidad de todos los valores que puede tomar  $\hat{p}$  en el conjunto de las posibles muestras de la población de un tamaño dado. Puesto que hay una correspondencia biunívoca entre el número de éxitos en una muestra de tamaño  $n$  y la proporción de éxitos en la muestra, al igual que en otras investigaciones sobre comprensión del muestreo, se pedirá a los estudiantes valores del número de éxitos en la muestra. El modelo probabilístico que se aplica a esta variable es la distribución Binomial  $B(n, p)$ , donde  $p$  es la proporción en la población de dicho rasgo y  $n$  el tamaño de la muestra.

## ANTECEDENTES

Pasamos a describir la investigación previa sobre muestreo directamente relacionada con nuestro trabajo; una síntesis de otras investigaciones se describe en Harradine, Batanero y Rossman (2011).

### Heurísticas y sesgos en el razonamiento sobre muestreo

Las primeras investigaciones sobre el muestreo se realizaron dentro del programa denominado *heurísticas y sesgos*, donde la tarea típica consiste en preguntar por la probabilidad de un valor de la proporción muestral en muestras de diferente tamaño, y se observan las siguientes heurísticas:

- La *heurística de representatividad* (Tversky y Kahneman, 1982), que consiste en estimar la probabilidad pedida, teniendo en cuenta únicamente la similitud que guarda la proporción muestral con la poblacional. Un sesgo asociado es la *insensibilidad al tamaño de la muestra*, que implica no tener en cuenta el tamaño de la muestra para juzgar la variabilidad de la proporción muestral. Otro sesgo (*falacia del jugador*) es considerar que el resultado de un experimento aleatorio afectará en la probabilidad de sucesos futuros. Si se supone que los siguientes resultados seguirán el patrón observado, se habla de *recencia positiva* y si se piensa que se compensarán los resultados futuros con los observados se denomina *recencia negativa*.
- La *heurística de la disponibilidad* (Tversky y Kahneman, 1974) consiste en estimar la probabilidad de un suceso basándose en la facilidad para encontrar ejemplos de situaciones similares. Un sesgo asociado es la *equiprobabilidad* (Lecoutre, 1992), mediante la cual se supone que los resultados de cualquier fenómeno aleatorio son igualmente probables. Este sesgo ha sido recientemente explicado por Chernoff y Russel (2012) mediante la *falacia de la composición*, que consiste en extender indebidamente la equiprobabilidad de los sucesos de un experimento aleatorio simple a cualquier muestra de resultados del experimento.

### Comprensión del concepto de muestra y del sesgo en el método de muestreo

Algunas investigaciones se centran en la comprensión del concepto de muestra y del sesgo en el método de muestreo por estudiantes de varias edades. Por ejemplo, Watson, en varios trabajos (Watson,

2004; Watson y Moritz, 2000*a* y *b*), realiza estudios con alumnos de entre 8 y 15 años. La autora pregunta en entrevistas y cuestionarios qué significa una muestra y pide proporcionar ejemplos de muestras. Desde su análisis, diferencia tres niveles de desarrollo en la comprensión del concepto: en el primero solo se comprende la terminología del muestreo; en el segundo, se comprenden además las aplicaciones del muestreo y en el tercero se adquiere una capacidad crítica para discutir conclusiones a partir de muestras que no sean adecuadas desde el punto de vista estadístico.

En Watson y Kelly (2005) se pide a 639 estudiantes de entre 8 y 14 años decidir entre varios ejemplos de muestreo aquellos que consideren adecuados. Los autores concluyen que muchos estudiantes prefieren métodos sesgados y no confían en el muestreo aleatorio para producir muestras representativas. A una conclusión similar llegan Saldahna y Thompson (2002) en un experimento de enseñanza con 26 estudiantes de 16 a 18 años, la mayoría de los cuales no llegan a diferenciar los tres tipos de distribución presentes en el muestreo (distribución de la población, distribución de la muestra y distribución del estadístico en el muestreo).

### **Comprensión de la distribución muestral**

La comprensión de la distribución muestral implica coordinar las ideas de valor esperado y variabilidad de un cierto estadístico en muestras repetidas de la misma población. La investigación relacionada suele utilizar poblaciones binomiales cuya proporción conoce el estudiante. Por ejemplo, Shaughnessy, Ciancetta y Canada (2004) investigaron la comprensión del muestreo de 272 estudiantes (10-19 años), pidiéndoles dar el número de sucesos de un cierto tipo en una muestra de 10 elementos y otra de 100 elementos. También les preguntaron si esperarían repetición de los resultados en una segunda muestra. La cuarta parte de estos estudiantes esperaban obtener el mismo resultado en las dos muestras repetidas del mismo tamaño y algunos propusieron en sus muestras resultados muy poco probables (como todos los sucesos idénticos). Por otro lado, la mayoría de los estudiantes sobreestimaron la variabilidad de la distribución muestral, con independencia del tamaño de la muestra, mientras algunos le otorgaron muy poca variabilidad. Los autores identifican tres niveles progresivos en el razonamiento sobre el muestreo: 1) el nivel de razonamiento aditivo (el más frecuente), que consiste en considerar las diferentes muestras como subconjuntos disjuntos de la población y utilizar en las estimaciones únicamente la frecuencia absoluta, sin tener en cuenta la proporción del suceso; 2) el nivel de razonamiento proporcional, en el que se utilizan proporciones al realizar estimaciones y se comprende el valor esperado de la distribución muestral, y 3) el nivel de razonamiento distribucional (el menos frecuente), donde se integran las ideas de valor esperado y de variabilidad, al realizar estimaciones.

Serrano (1996) propuso algunas preguntas a 147 alumnos españoles de 13 años y 130 alumnos de 17 años sobre variabilidad de la proporción en muestras grandes y pequeñas. Al igual que en estudios previos, muchos estudiantes suponían mayor variabilidad en las muestras grandes.

Otra investigación que analiza la comprensión de la distribución muestral, desde un punto de vista formal y después de haber realizado un experimento de enseñanza con estudiantes universitarios, es la de Alvarado, Galindo y Retamal (2013). Los autores informan del éxito de su propuesta para la comprensión de la convergencia de la distribución muestral hacia la distribución normal y el valor esperado de dicha distribución. Sin embargo, las propiedades de la variabilidad de dicha distribución y la relación con el tamaño de la muestra continuaron siendo complejas.

### **Comprensión de la probabilidad, desde el punto de vista frecuencial**

En el significado frecuencial, la probabilidad se estima a partir de las frecuencias relativas de aparición del suceso, apoyándose en la ley de los grandes números. Las ideas de representatividad y variabilidad

muestral están implícitas en este significado, y la precisión de la estimación de la probabilidad depende del tamaño de la muestra. En consecuencia, podemos reinterpretar en términos de muestreo algunas investigaciones relacionadas con la comprensión de la probabilidad desde el punto de vista frecuencial, en las que se inspiran las tareas de nuestro cuestionario.

Uno de los ítems utilizados por Green (1983), en su estudio del razonamiento probabilístico de 2930 estudiantes (11-16 años), pide estimar el número de chinchetas que caerán con la punta hacia arriba cuando se lanzan 100 chinchetas sobre una mesa. El enunciado de la tarea indica que, en la realización previa del experimento, 68 de 100 chinchetas cayeron con la punta hacia arriba. Aunque el autor esperaba que los estudiantes diesen un resultado parecido, el 64 % mostraron el sesgo de equiprobabilidad, sugiriendo que la mitad de las chinchetas, aproximadamente, caerían hacia arriba. Otro 15 % mostraron preferencia respecto a la punta hacia arriba, pero dando una cantidad muy alejada de la esperada. Solamente el 17 % dieron una estimación correcta. Resultados muy similares fueron obtenidos con el mismo ítem por Cañizares (1997), en una muestra de 253 estudiantes españoles (11-14 años), de los que solo el 15 % responden correctamente.

Gómez, Batanero y Contreras (2014), en una investigación con 202 futuros profesores de educación primaria, incluyen una adaptación del ítem sobre lanzamiento de chinchetas propuesto por Green (1983), que consiste en pedir cuatro resultados diferentes al lanzar 100 chinchetas. Solo una tercera parte de los sujetos de su estudio tuvieron una intuición simultánea de la convergencia al valor esperado y la variabilidad muestral. Además, los autores identifican en los sujetos diferentes sesgos como la equiprobabilidad o la heurística de la representatividad.

Los trabajos anteriores analizan la comprensión de diferentes propiedades del muestreo; no obstante, salvo en el trabajo de Gómez *et al.* (2014), no analizan la variabilidad de las diferentes estimaciones. En nuestro estudio se realiza este análisis, y se comparan los resultados obtenidos, teniendo en cuenta algunas variables de tareas que se describen en la sección siguiente.

## MÉTODO

Nuestro trabajo tiene carácter exploratorio y está basado en el análisis estadístico de las respuestas a un cuestionario formado por cuatro ítems de respuesta abierta.

### Participantes y contexto

La muestra estuvo constituida por un total de 302 estudiantes, 157 de segundo curso de ESO (13-14 años; en total 9 grupos) y 145 de cuarto curso de ESO (15-16 años; en total 8 grupos). Todos los alumnos cursaban sus estudios en centros públicos diferentes en la ciudad de Huesca. Se solicitó permiso a los directores de los centros, a la Consejería de Educación y a los profesores de los grupos.

Los estudiantes de segundo curso no habían recibido instrucción previa sobre probabilidad y estadística, aparte de los conocimientos que pudieran haber obtenido en la educación primaria, mientras que los estudiantes de cuarto habían estudiado la probabilidad desde el significado clásico y estadística descriptiva en el curso anterior, siguiendo los contenidos del currículo que hemos descrito, que incluye los conceptos de población y muestra (MECD, 2015). Esta información la proporcionaron los profesores de los estudiantes.

Los datos fueron recogidos como una actividad de la clase de matemáticas, completándose por escrito en un tiempo de 15-25 minutos. Durante la aplicación del cuestionario, estuvieron presentes el docente responsable del grupo y una de las investigadoras. Esta última describió la finalidad del cuestionario, resolviendo posibles dudas sobre la forma de completarlo.

## Cuestionario y conocimientos evaluados

El cuestionario se presenta en el «Anexo» y tiene como finalidad evaluar la comprensión intuitiva del estudiante de las siguientes propiedades del muestreo:

1. Proporción muestral esperada y relación que establece el estudiante entre dicha proporción muestral y la proporción en la población.
2. Variabilidad de la proporción en el muestreo y efecto del tamaño de la muestra sobre dicha variabilidad.

El primer ítem del cuestionario se ha tomado de Gómez *et al.* (2014) y los otros tres son variantes del primero. En todas las tareas se pide al estudiante escribir cuatro resultados probables del valor esperado de un suceso dado, en muestras del mismo tamaño de la población descrita en el enunciado. A continuación, se analiza detalladamente la primera tarea, lo que nos permite caracterizar las respuestas de los estudiantes.

En el primer ítem, el número de chinchetas  $X$  que caen con la punta hacia arriba sigue una distribución binomial,  $B(n, p)$ , donde  $n = 100$  es el tamaño de la muestra, y  $p$  es la proporción desconocida de chinchetas que caen con la punta hacia arriba en la población. La inferencia estadística indica que la mejor estimación de esta es la proporción muestral dada en el enunciado,  $\hat{p} = 0,68$ , por sus propiedades de insesguez y mínima varianza (Zacks, 2014). A partir de ella, se puede también estimar el número esperado de chinchetas que caerán con la punta hacia arriba en la muestra,  $n\hat{p} = 68$  y su desviación típica,  $\sqrt{n\hat{p}(1-\hat{p})} = 4,66$ . Por tanto, se considera que el sujeto presenta una buena comprensión intuitiva de la proporción muestral, si el valor medio del número de chinchetas con la punta hacia arriba en las cuatro muestras que proporciona es cercano a 68 (media teórica). Es decir, si dicho valor medio se localiza dentro del intervalo  $[63,3-72,7]$ , que contiene el 68 % de las observaciones centrales en la distribución normal, que aproxima a la distribución binomial considerada. En nuestro estudio, se ha conservado el valor  $n = 100$  del ítem usado por Green (1983) y adaptado por Gómez *et al.* (2014).

Para analizar la comprensión de la variabilidad, se estudia el rango de los cuatro valores proporcionados por cada estudiante. Puesto que en una distribución normal, en el intervalo  $\mu \pm 2\sigma$  se encuentran el 95 % de las observaciones, y siguiendo a Gómez *et al.* (2014), se considera normativa la variabilidad cuando el rango calculado en la forma anterior está a distancia entre dos y cuatro desviaciones típicas de la media. Para el primer ítem, en el 95 % de las muestras, el número de chinchetas que caen con la punta hacia arriba oscilarían entre 58 y 78, que se corresponde con un rango entre 10 y 20. Si el rango se incluye entre cuatro y seis desviaciones típicas (que contiene el 99 % de observaciones en la distribución normal; en el ítem 1, entre 21 y 30) se considera aceptable y si es mayor, excesivo. Si es menor que dos desviaciones típicas (10 en el ítem 1), se considera demasiada concentración y se interpreta como falta de comprensión de la variabilidad muestral.

Realizando un análisis similar con los otros tres ítems, la tabla 1 muestra las variables consideradas en el cuestionario, donde se han redondeado los valores obtenidos para calcular los intervalos en que el rango es normativo o aceptable. Los dos primeros ítems corresponden a muestras grandes (100 elementos) y los dos últimos a muestras pequeñas (10 elementos). El primero y último de los ítems describen situaciones en las que el fenómeno aleatorio presenta sucesos no equiprobables, cuya probabilidad se estima a partir de la frecuencia dada en el enunciado del ítem. Las otras dos tareas hacen referencia a sucesos equiprobables, cuya probabilidad el estudiante puede conocer a través de su experiencia o considerando la regla de Laplace.



Tabla 1.  
Variables consideradas en el diseño del cuestionario

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Tamaño de la muestra	100	100	10	10
Probabilidad del suceso de interés	0,68	0,5	0,5	0,7
Estimación de la probabilidad	Frecuencial	Clásica	Clásica	Frecuencial
Número esperado de éxitos	68	50	5	7
Desviación típica	4,66	5	1,58	1,45
Intervalo que contiene el 68 % de medias muestrales	[63,3-72,7]	[45-55]	[3,4-6,6]	[5,5-8,4]
Intervalo que contiene el 95 % de medias muestrales	[58,6-77,4]	[40-60]	[1,8-8,2]	[4,1-9,9]
Rango normativo	[10-20]	[10-20]	[3-6]	[3-6]
Rango aceptable	[21-30]	[21-30]	[7-9]	[7-9]

Además, en la tabla 1 incluimos el valor esperado y la desviación típica de la distribución binomial que modeliza la situación descrita en cada ítem, los intervalos que contienen el 68 y 95 % de las medias muestrales, y los valores del rango considerados como normativo y aceptable.

## RESULTADOS Y DISCUSIÓN

Tras la recogida de los datos, se calcularon la media y el rango de los cuatro valores proporcionados por el estudiante para cada ítem. El análisis de la distribución de los valores medios así obtenidos informa de la comprensión intuitiva de los estudiantes del valor esperado de la proporción muestral, mientras que el estudio de la distribución de los rangos permite identificar la variabilidad que los estudiantes conceden a las proporciones de las muestras proporcionadas. Para cada uno de estos aspectos se comparan los resultados en los dos cursos que componen la muestra. Finalmente, se comparan los resultados por ítem, para analizar las posibles diferencias según el tamaño de la muestra y según si los sucesos considerados son o no equiprobables.

### Comprensión intuitiva del valor esperado

En la tabla 2 se presentan las medias teóricas (valor esperado en la distribución binomial que modela el enunciado de cada ítem) y las obtenidas en la muestra total de alumnos.

Tabla 2.  
Media teórica y observada de las distribuciones del valor medio de las cuatro respuestas proporcionadas por cada estudiante a cada ítem en el global de la muestra

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Media teórica	68	50	5	7
Media observada	57,9	51,2	5,1	6,6

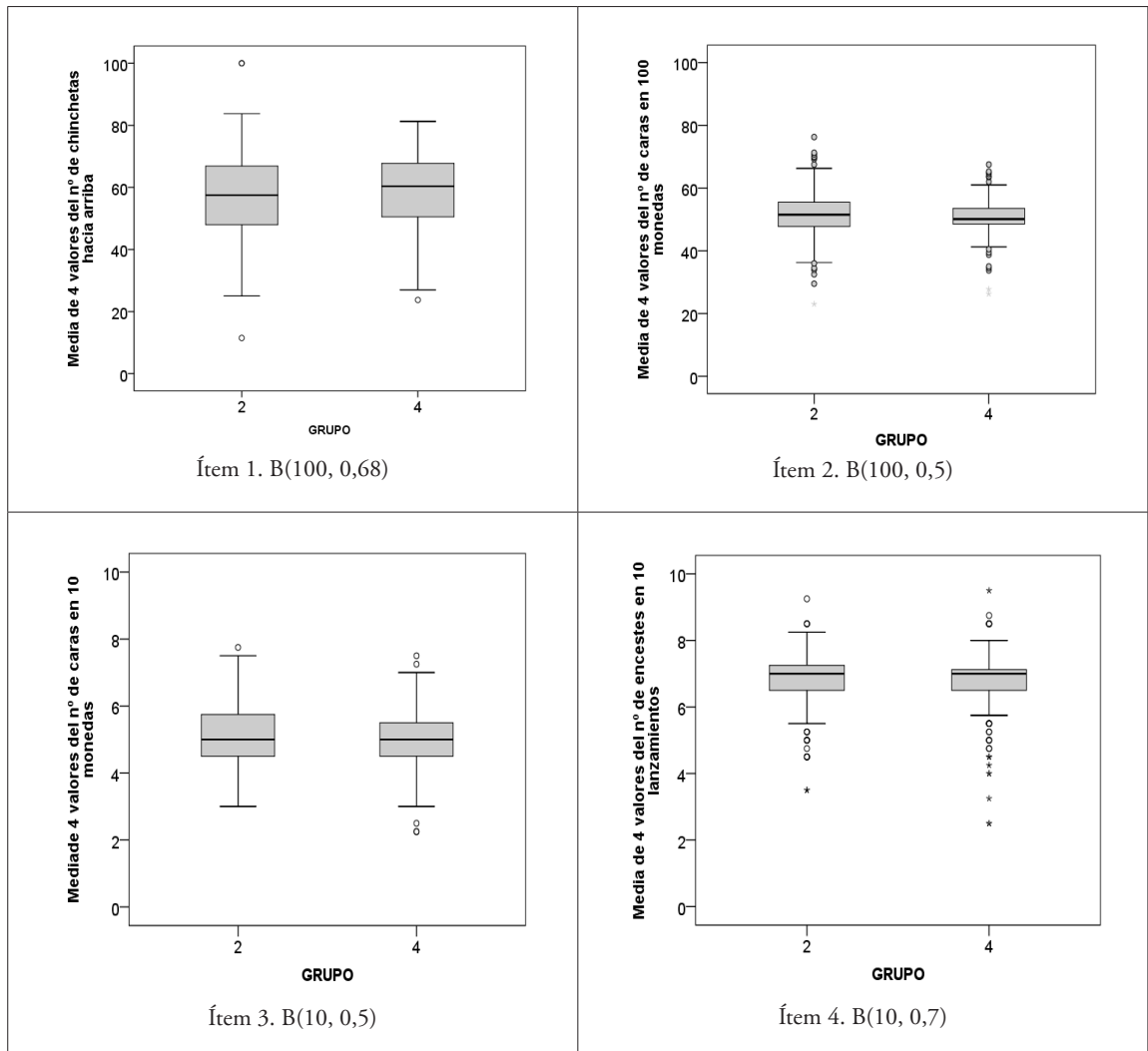


Fig. 1. Gráficos de caja de las distribuciones del valor medio en las cuatro estimaciones proporcionadas por los estudiantes en cada ítem según curso.

Los estudiantes han sido bastante precisos en su estimación del valor esperado de la proporción muestral en los tres últimos ítems, puesto que la media observada en la distribución obtenida de sus respuestas se acerca mucho a la media teórica en la distribución binomial correspondiente. Por el contrario, en el primer ítem encontramos una diferencia de 10 puntos entre estos dos valores. En este ítem, la probabilidad del suceso de interés (punta hacia arriba) se proporciona al estudiante a partir de una información frecuencial y los estudiantes también pueden razonar, desde su experiencia, que es más probable que la chincheta caiga hacia arriba, debido al peso de su superficie plana. Sin embargo, se obtuvo un valor medio de 57,9 (muy similar al obtenido en el trabajo de Gómez *et al.* [2014] con futuros profesores, que fue 57,7). Este valor, inferior al valor medio esperado (68), es debido a que parte de los estudiantes proporcionan una estimación del número de chinchetas que caen hacia arriba sesgada hacia el valor 50, pues no tienen en cuenta la información frecuencial. Dichos estudiantes razonarían de acuerdo con el sesgo de equiprobabilidad, que es uno de los sesgos atribuidos a la heurística de la disponibilidad. En nuestro trabajo, no es posible aceptar la teoría de Chernoff y Russell (2012), por la que los estudiantes razonan según la equiprobabilidad debido a la falacia de la composición, trans-



firiendo a la muestra una propiedad de la población, puesto que en este ítem los sucesos elementales no son equiprobables.

Para comparar las respuestas en los dos cursos, se presentan en la figura 1 los gráficos de caja de las distribuciones del valor medio de las cuatro estimaciones proporcionadas por los estudiantes. Observamos un comportamiento muy similar en los dos grupos en todos los ítems, en lo que se refiere a las medianas y cuartiles. Además, más de la mitad de los estudiantes, salvo en el ítem 1, presentan una buena comprensión intuitiva de la proporción muestral, pues los cuartiles se encuentran incluidos en el intervalo aceptable de estimación (tabla 1). También advertimos la existencia de valores atípicos (que se separan mucho del centro de los datos).

Para analizar mejor estas diferencias, las respuestas de los estudiantes se han clasificado en las siguientes categorías:

- *Estimación normativa del valor esperado.* Cuando el valor medio de las cuatro estimaciones proporcionadas por el estudiante en el ítem es muy próximo al valor teórico (pertenecer al intervalo central que contendría el 68 % de los valores de las medias muestrales). Los estudiantes que dan estas respuestas comprenden la propiedad de representatividad muestral.
- *Estimación aceptable del valor esperado.* Si el valor medio de las cuatro estimaciones dadas por el estudiante se encuentra fuera del intervalo central que contendría el 68 % de los valores de las medias muestrales, pero dentro del intervalo que contendría el 95 % de dichos valores.
- *Recencia negativa.* En los experimentos descritos en los ítems 1 y 4 el enunciado indica que se obtuvo una frecuencia mayor de uno de los sucesos. Algunos estudiantes tratan de compensar este resultado, dando estimaciones cuyo valor medio favorece el suceso contrario (como ejemplo, en el ítem 1 dan muestras con un número de chinchetas hacia arriba igual a 38 o 40). Esta tendencia muestra una recencia negativa, conducta que se ha explicado mediante la heurística de la representatividad y que implica una pobre comprensión de la independencia de los ensayos en un experimento aleatorio (Batanero y Serrano, 1999).
- *Recencia positiva.* Ante un enunciado en el que la frecuencia de los dos sucesos implicados es diferente, se tiende a proporcionar estimaciones en las que se exagera tal diferencia, conducta también explicada por la heurística de la representatividad. En el caso del ítem 1, sería, por ejemplo, proporcionar muestras con 90 o incluso con 100 chinchetas hacia arriba. Estos casos extremos también fueron informados por Shaughnessy *et al.* (2004), aunque los autores no los relacionan con los sesgos de recencia positiva o negativa.
- *Sesgo de equiprobabilidad.* En los ítems 1 y 4, que describen experimentos con sucesos elementales no equiprobables, algunos estudiantes proporcionan estimaciones cuyos valores medios se localizan muy próximos al 50 % del tamaño de la muestra, sin considerar la información frecuencial del enunciado. Estos estudiantes presentan el sesgo de equiprobabilidad que se atribuye a la heurística de disponibilidad. Algunos estudiantes incluso proporcionan cuatro estimaciones exactamente iguales al 50 %; por ejemplo, en el ítem 1, escriben una cuaterna (50, 50, 50, 50), respuesta que también aparece en el trabajo de Gómez *et al.* (2014).
- *No completa.* Consideramos en esta categoría a aquellos estudiantes que no responden a un ítem.

Tabla 3.  
Porcentaje de alumnos por grupo según valor medio en las cuatro estimaciones en cada ítem

	2.º ESO				4.º ESO			
	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Estimación normativa del valor esperado	29,0	50,3	83,4	77,7	33,8	69,7	78,6	77,9
Estimación aceptable del valor esperado	12,7	22,3	7,7	10,1	15,1	14,5	13,1	11,1
Recencia negativa	15,9	7,6		1,4	11,0	4,8		4,1
Recencia positiva	3,5	12,8			2,9	6,2		
Equiprobabilidad	32,5				32,4			
No completa	6,4	7,0	8,9	10,8	4,8	4,8	8,3	6,9

En la tabla 3 observamos que la estimación normativa es la más frecuente, salvo en el ítem 1, lo que muestra una buena percepción del valor esperado en los estudiantes y una adecuada relación entre el valor de la proporción en la población y en la muestra. Además, salvo en el ítem 1, la segunda respuesta más frecuente es la estimación aceptable (intervalo del 95 %). El porcentaje de estudiantes que da una de estas dos respuestas es muy alto, y superior al 70 % en los dos grupos y en los tres últimos ítems. En consecuencia, la mayoría de los estudiantes de nuestra muestra han llegado, al menos, al nivel de razonamiento proporcional sobre el muestreo descrito por Shaughnessy *et al.* (2004) al estar realizando sus estimaciones, que tienen en cuenta la proporción en la población teórica.

El ítem que tiene peor resultado es el primero, donde alrededor del 30 % en cada grupo muestran el *sesgo de equiprobabilidad*, dando estimaciones cuyo valor medio se acerca al 50 %, sin tener en cuenta los datos frecuenciales (apareció en el 33,8 % en el trabajo de Gómez *et al.*, 2014). Los *sesgos de recencia positiva y negativa* tienen una frecuencia similar a la observada en el trabajo de Gómez *et al.* (2014). Por tanto, estos estudiantes responden según las heurísticas de disponibilidad y representatividad.

Al comparar los dos grupos, considerando las estimaciones normativas y aceptables como correctas y la disminución de los sesgos, observamos que los resultados son mejores en los alumnos de cuarto curso en todos los ítems, aunque los dos grupos muestran muy buenos resultados.

### Percepción de la variabilidad en el muestreo

La percepción de los estudiantes de la variabilidad en el muestreo se resume en la figura 2, donde se muestra la distribución del rango de las cuatro estimaciones proporcionadas por los alumnos en cada ítem. Al comparar con los valores normativos y aceptables para el rango en cada ítem (tabla 1), vemos que ambos grupos de estudiantes proporcionan estimaciones con una variabilidad excesiva (alrededor de la mitad de los estudiantes) en los dos primeros ítems, correspondientes a muestras de tamaño  $n = 100$ . Por el contrario, en los ítems 3 y 4, correspondientes a muestras pequeñas ( $n = 10$ ), el rango es, generalmente, igual (en el ítem 3) o inferior (alta concentración, en el ítem 4) al valor normativo. En este sentido, nuestros resultados se diferencian de los de Shaughnessy *et al.* (2004), ya que en la investigación de estos autores, la mayoría de estudiantes produjeron una variabilidad excesiva, con independencia del tamaño de la muestra.

Para analizar mejor la percepción de la variabilidad, por parte de los participantes, sus respuestas se han clasificado en las siguientes categorías:

- *Estimación normativa de la variabilidad muestral.* Es la respuesta óptima; la da el estudiante que ofrece cuatro estimaciones cuyo rango se encuentra entre dos y cuatro veces la desviación típica teórica. Dicho rango corresponde a los valores más probables en la distribución binomial considerada en el enunciado.

- *Estimación aceptable de la variabilidad muestral.* El estudiante da cuatro estimaciones cuyo rango se encuentra entre cuatro y seis veces la desviación típica teórica. Estos valores tienen menor probabilidad en la distribución, pero no serían extremadamente raros.
- *Estimación excesiva de la variabilidad muestral.* El estudiante proporciona cuatro estimaciones cuyo rango es mayor a seis veces la desviación típica de la distribución. El estudiante no percibe que la variabilidad de la distribución muestral es menor que la de la población.
- *Alta concentración.* Cuando el estudiante proporciona cuatro estimaciones con el mismo valor o muy parecido. En este caso, subyace una concepción determinista del muestreo, esperando la replicación de los resultados en las diferentes muestras. No se muestra comprensión intuitiva de la variabilidad del muestreo.

Los resultados se presentan en la tabla 4, donde observamos que pocos estudiantes proporcionan estimaciones con variabilidad normativa o aceptable, excepto en el ítem 3, siendo este porcentaje menor que el obtenido por Gómez *et al.* (2014) (45 % de futuros profesores). En concreto, en el ítem 3 (lanzamiento de 10 monedas), la mayoría sugieren una variabilidad normativa y, por tanto, los resultados contradicen los de Shaughnessy *et al.* (2004). En los dos primeros ítems lo más frecuente es que las estimaciones tengan una variabilidad excesiva y en el cuarto ítem, una alta concentración.

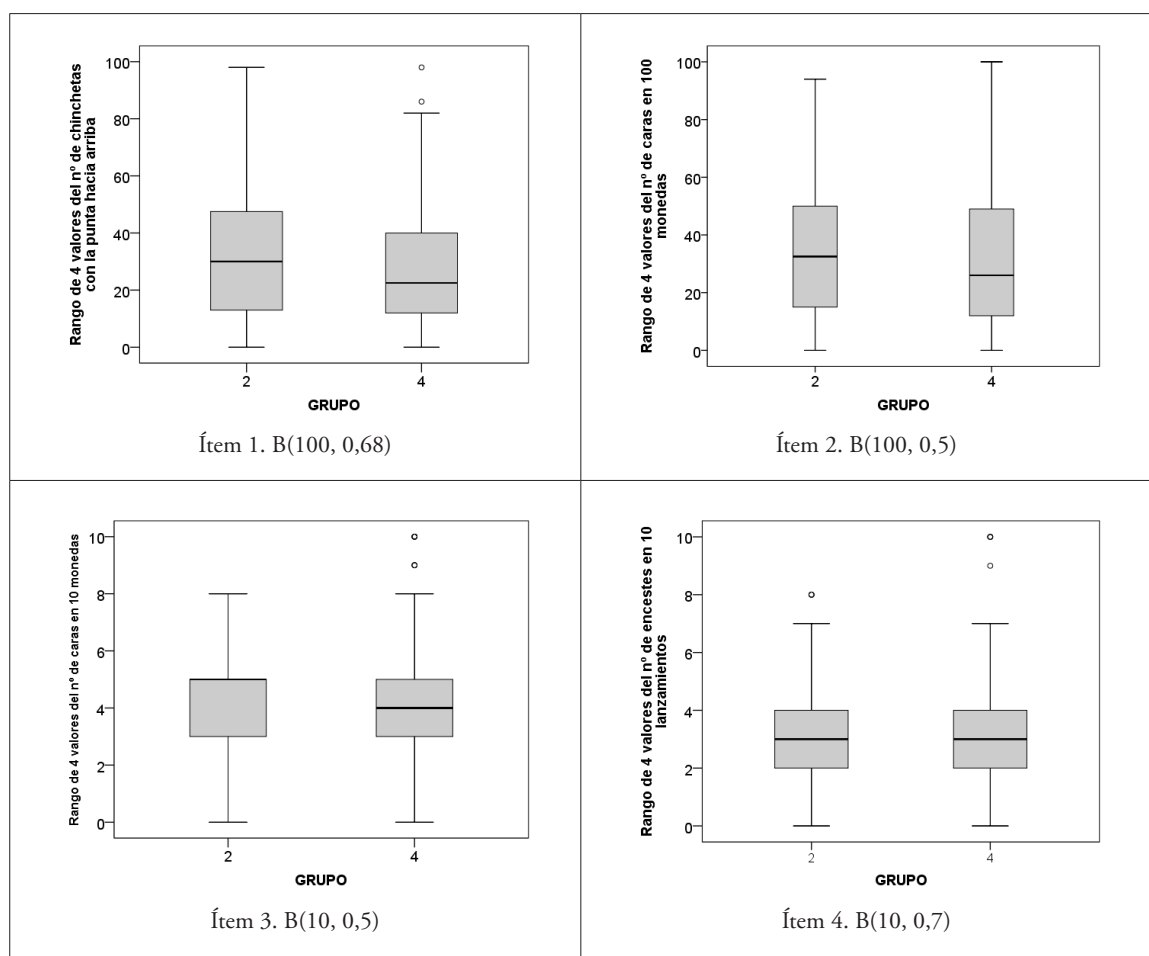


Fig. 2. Gráficos de caja de las distribuciones del rango en las cuatro estimaciones proporcionadas por los estudiantes en cada ítem según curso.

Salvo en el caso del lanzamiento de 10 monedas (ítem 3), los estudiantes no aprecian la variabilidad del muestreo y el efecto del tamaño de la muestra sobre esta. Además, en las muestras grandes (ítems 1 y 2) se percibe más variabilidad que en las pequeñas, cuando la teoría estadística indica justamente lo contrario. Estos resultados coinciden con la investigación de Serrano (1996), cuyos estudiantes también concedían mayor variabilidad a las muestras grandes frente a las pequeñas. En consecuencia, son pocos los estudiantes que producen una variabilidad adecuada y llegan al nivel de razonamiento distribucional sobre muestreo descrito por Shaughnessy *et al.* (2004).

Tabla 4.  
Porcentaje de alumnos por grupo según intervalo  
en que se sitúa el rango de las cuatro estimaciones en cada ítem

Variabilidad	2.º ESO				4.º ESO			
	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Estimación normativa de la variabilidad muestral	18,5	20,4	68,8	21,0	26,0	20,7	69,0	14,5
Estimación aceptable de la variabilidad muestral	15,3	10,8	10,2		16,6	17,9	10,3	0,7
Estimación de una variabilidad muestral excesiva	45,2	47,8			34,3	37,3	3,4	2,1
Alta concentración	14,6	14,0	12,1	68,2	18,3	19,3	9,0	75,8
No completa	6,4	7,0	8,9	10,8	4,8	4,8	8,3	6,9

En el último ítem, la excesiva concentración apunta de nuevo a la consideración de la situación como cercana al determinismo. También apreciamos un efecto del contexto pues, aunque se trata de un juego donde interviene el azar, la habilidad del lanzador de la canasta parece percibirse como determinista. Las diferencias por curso son ahora menores y prácticamente se restringen al caso de muestras grandes (ítems 1 y 2), donde los alumnos de cuarto curso muestran una mejor comprensión intuitiva de la variabilidad en el muestreo.

## CONCLUSIONES E IMPLICACIONES PARA LA ENSEÑANZA

Los resultados sugieren, en general, una buena percepción del valor esperado de la proporción en las situaciones de muestreo propuestas, dada la proximidad entre la media teórica y la observada en las cuatro estimaciones ofrecidas en cada ítem. Sin embargo, menos de la cuarta parte de los estudiantes parecen percibir correctamente el efecto del tamaño de la muestra sobre la variabilidad muestral. Por consiguiente, la mayoría de los estudiantes, en especial los de cuarto curso, alcanzan el nivel proporcional de razonamiento sobre el muestreo descrito por Shaughnessy *et al.* (2004), pero pocos llegan al nivel distribucional de razonamiento sobre muestreo descrito por estos autores.

Nuestro trabajo proporciona algunas conclusiones nuevas respecto a investigaciones previas. En primer lugar, el comportamiento de una parte de los estudiantes en el primer ítem, olvidando la información frecuencial, lo que indica que todavía en estos niveles educativos persiste el *sesgo de equiprobabilidad* (Lecoutre, 1992). No se sostiene para estos estudiantes la explicación de dicho sesgo mediante la *falacia de la composición* debida a Chernoff y Russell (2012), ya que aplican la equiprobabilidad a muestras en que los sucesos de la población no son equiprobables. Por otro lado, aunque la media global de las cuatro estimaciones dadas se acerque a la teórica, observamos a estudiantes que razonan de acuerdo con la recencia negativa o recencia positiva, resultado observado en Gómez *et al.* (2014).

La variabilidad excesiva de las estimaciones en las muestras grandes, mayor incluso que en las pequeñas, indica una falta de intuición sobre la ley de los grandes números y confirma los resultados de Serrano (1996). También contradice otras investigaciones, como la de Watson y Moritz (2000*a* y *b*), quienes observaron en su estudio que los estudiantes producen muestras sin apenas variabilidad, y la de Shaughnessy *et al.* (2004), cuyos alumnos también sobreestiman la variabilidad en las muestras pequeñas. La razón de las diferencias comentadas radica en el tipo de tarea, dado que los autores pedían a sus estudiantes generar una o dos muestras en cada situación, mientras que nosotros les hemos pedido cuatro, lo que nos ha permitido analizar la variabilidad de las cuatro estimaciones proporcionadas mediante un estudio estadístico, en lugar de juzgar la variabilidad de forma cualitativa. En nuestro estudio añadimos información sobre el efecto en la respuesta de las variables consideradas en los ítems y se muestra que una parte importante del grupo produce muestras de variabilidad extrema, o de patrón determinista, como ocurrió en el de Shaughnessy *et al.* (2004).

En general, se observan mejores resultados en los alumnos de cuarto curso de la ESO, en todos los ítems y características analizadas. Este resultado puede ser debido a la enseñanza previa de diferentes contenidos de estadística y probabilidad, así como a la mayor madurez de los estudiantes. Sin embargo, la diferencia es pequeña y todavía en el cuarto curso la mayoría de los estudiantes no comprenden bien la variabilidad en el muestreo. Los estudiantes que cursen posteriormente el bachillerato de Ciencias Sociales o que en la universidad tengan algún curso de estadística se encontrarán poco preparados para abordar la inferencia estadística, pues pocos han alcanzado el nivel de razonamiento distribucional (Shaughnessy *et al.*, 2004).

Como hemos descrito en la introducción, el currículo actual en la ESO pone el énfasis en la diferencia entre población y muestra, los métodos de selección de las muestras y la idea de representatividad. La variabilidad en el muestreo es un contenido implícito que podría resaltarse cuando se introduce el enfoque frecuencial de la probabilidad. En dicho enfoque, la actividad más frecuente es el cálculo de la frecuencia relativa de un suceso de interés, aumentando progresivamente el tamaño de la muestra, para mostrar empíricamente la convergencia de la frecuencia relativa a la probabilidad teórica. Por tanto, sería importante completar estas actividades con otras relacionadas con la extracción de un gran número de muestras del mismo tamaño de una población, para analizar la variabilidad de dicho estadístico y llegar progresivamente a la idea de distribución muestral del estadístico.

La simulación, cuyo interés es resaltado por Huerta (2015), permite adquirir experiencia con el muestreo repetido y la distribución muestral, esenciales para la posterior comprensión de la inferencia. Como indican Saldahna y Thompson (2002), aunque la inferencia (por ejemplo, el cálculo de un intervalo de confianza) se realiza a partir de una única muestra, el razonamiento inferencial implica imaginar todas las posibles muestras del mismo tamaño que podrían tomarse de la población dada. Con ello, el estudiante llegaría a diferenciar claramente las tres distribuciones implícitas en el muestreo, tomando conciencia de que la distribución de datos de la muestra que se ha tomado permite realizar predicciones sobre la distribución de probabilidad de la población. Sin embargo, el estadístico de la muestra concreta de la que se dispone es solo un elemento de la distribución muestral de dicho estadístico, que es la que permite completar los márgenes de error y las probabilidades asociadas en las inferencias sobre el parámetro en la población.

## AGRADECIMIENTOS

Proyecto EDU2016-74848-P (AEI, FEDER) y Grupo FQM126 (Junta de Andalucía).

## REFERENCIAS

- ALVARADO, H., GALINDO, M. y RETAMAL, L. (2013). Comprensión de la distribución muestral mediante configuraciones didácticas y su implicación en la inferencia estadística. *Enseñanza de las Ciencias*, 31(2), pp. 75-91.  
<http://doi.org/10.5565/rev/ec/v31n2.803>
- BATANERO, C. y SERRANO, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, 30(5), pp. 558-567.
- BURRILL, G. y BIEHLER, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.). *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 57-69). Dordrecht, Holanda: Springer.  
[https://doi.org/10.1007/978-94-007-1131-0\\_10](https://doi.org/10.1007/978-94-007-1131-0_10)
- CAÑIZARES, M. J. (1997). *Influencia del razonamiento proporcional y combinatorio y de creencias subjetivas en las intuiciones probabilísticas primarias*. Tesis doctoral no publicada. Granada: Universidad de Granada.
- CHERNOFF, E. J. y RUSSELL, G. L. (2012). The fallacy of composition: Prospective mathematics teachers' use of logical fallacies. *Canadian Journal for Science, Mathematics and Technology Education*, 12(3), pp. 259-271.  
<http://doi.org/10.1080/14926156.2012.704128>
- EICHLER, A. y VOGEL, M. (2014). Three approaches for modelling situations with randomness. En E. J. Chernoff y B. Sriraman (Eds.). *Probabilistic thinking, presenting plural perspectives* (pp. 75-99). Dordrecht, Holanda.  
[https://doi.org/10.1007/978-94-007-7155-0\\_4](https://doi.org/10.1007/978-94-007-7155-0_4)
- GÓMEZ, E., BATANERO, C. y CONTRERAS, C. (2014). Conocimiento matemático de futuros profesores para la enseñanza de la probabilidad desde el enfoque frecuencial. *Bolema*, 28(48), pp. 209-229.  
<http://doi.org/10.1590/1980-4415v28n48a11>
- GREEN, D. R. (1983). A Survey of probabilistic concepts in 3000 pupils aged 11-16 years. En D. R. Grey, P. Holmes, V. Barnett y G. M. Constable (Eds.). *Proceedings of the First International Conference on Teaching Statistics* (vol. 2, pp. 766-783). Universidad de Sheffield: Teaching Statistics Trust.
- HARRADINE, A., BATANERO, C. y ROSSMAN, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.). *Teaching statistics in school mathematics. Challenges for teaching and teacher education*. (pp. 235-246). Dordrecht, Holanda: Springer.  
[https://doi.org/10.1007/978-94-007-1131-0\\_24](https://doi.org/10.1007/978-94-007-1131-0_24)
- HEITEL, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6(2), pp. 187-205.  
<https://doi.org/10.1007/BF00302543>
- HUERTA, M. P. (2015). La resolución de problemas de probabilidad con intención didáctica en la formación de maestros y profesores de matemáticas. En C. Fernández, M. Molina y N. Planas (Eds.). *Investigación en Educación Matemática XIX* (pp. 105-119). Alicante: Sociedad Española de Investigación en Educación Matemática.
- KAHNEMAN, D. (1982). Judgments of and by representativeness. En D. Kahneman, P. Slovic y A. Tversky (Eds.). *Judgement under uncertainty: Heuristics and biases* (pp. 117-128). Nueva York: Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511809477.007>
- KAHNEMAN, D., SLOVIC, P. y TVERSKY, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Nueva York: Cambridge University Press.



- LECOUTRE, M. P. (1992). Cognitive models and problem spaces in «purely random» situations. *Educational Studies in Mathematics*, 23(6), pp. 557-568.  
<https://doi.org/10.1007/BF00540060>
- MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE, MECD (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.
- RUBIN, A., BRUCE, B. y TENNEY, Y. (1991). Learning about sampling: trouble at the core of statistics. En D. Vere-Jones (Ed.). *Proceedings of the Third International Conference on Teaching Statistics*. Dunedin, Nueva Zelanda: International Statistical Institute. Disponible en línea: <[iase-web.org/documents/papers/icots3/BOOK1/A9-4.pdf](http://iase-web.org/documents/papers/icots3/BOOK1/A9-4.pdf)> (consultado el 19/06/2017).
- SALDANHA, L. y THOMPSON, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51(2), pp. 257-270.  
<http://doi.org/10.1023/A:1023692604014>
- SERRANO, L. (1996). *Significados institucionales y personales de objetos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad*. Tesis doctoral no publicada. Granada: Universidad de Granada.
- SHAUGHNESSY, J. M., CIANCETTA, M. y CANADA, D. (2004). Types of student reasoning on sampling tasks. En M. J. Hoines y A. B. Fuglestad (Eds.). *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 177-184). Bergen, Noruega: International Group for the Psychology of Mathematics Education.
- TVERSKY, A. y KAHNEMAN, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, pp. 1124-1131.  
<http://doi.org/10.1126/science.185.4157.1124>
- WATSON, J. M. (2004). Developing reasoning about samples. En D. Ben-Zvvi y J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, Holanda: Springer.  
<https://doi.org/10.1007/s10649-014-9551-5>
- WATSON, J. y KELLY, B. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal*, 17(1), pp. 24-57.  
<https://doi.org/10.1007/BF03217408>
- WATSON, J. M. y MORITZ, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), pp. 44-70.
- WATSON, J. M. y MORITZ, J. B. (2000b). Development of understanding of sampling for statistical literacy. *The Journal of Mathematical Behavior*, 19(1), pp. 109-136.  
[https://doi.org/10.1016/S0732-3123\(00\)00039-0](https://doi.org/10.1016/S0732-3123(00)00039-0)
- ZACKS, S. (2014). *Parametric statistical inference: basic theory and modern approaches*. Oxford, Reino Unido: Pergamon Press.



## ANEXO

### Cuestionario

*Ítem 1.* Un profesor vacía sobre la mesa un paquete de 100 chinchetas y obtiene los siguientes resultados: 68 caen con la punta para arriba y 32 caen hacia abajo. Supongamos que el profesor pide a 4 niños repetir el experimento, lanzando las 100 chinchetas. Cada niño vacía una caja de 100 chinchetas y obtendrá algunas con la punta hacia arriba y otras con la punta hacia abajo.

Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Daniel	Martín	Diana	María
Punta arriba:	Punta arriba:	Punta arriba:	Punta arriba:
Punta abajo:	Punta abajo:	Punta abajo:	Punta abajo:

*Ítem 2.* Un profesor vacía sobre la mesa un paquete de 100 monedas obteniendo los siguientes resultados: 53 caen con la cara hacia arriba y 47 caen con la cruz hacia arriba.

Supongamos que el profesor pide a 4 niños repetir el experimento. Cada niño lanza las 100 monedas y obtendrá algunas con la cara hacia arriba y otras con la cruz hacia arriba.

Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Elena	Clara	Matías	Rosa
Cara:	Cara:	Cara:	Cara:
Cruz:	Cruz:	Cruz:	Cruz:

*Ítem 3.* Un profesor pide a 4 niños lanzar 10 monedas sobre la mesa y contar el número de caras y cruces obtenidos.

Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Silvia	Javier	Miguel	Carmen
Cara:	Cara:	Cara:	Cara:
Cruz:	Cruz:	Cruz:	Cruz:

*Ítem 4.* Un jugador de baloncesto suele encestar 70 de cada 100 tiros a una canasta desde la posición de tiros libres.

Escribe en la siguiente tabla un resultado que sea probable para cuatro partidos en los que lanza 10 tiros desde el punto de lanzamientos libres.

Partido 1 (10 tiros)	Partido 2 (10 tiros)	Partido 3 (10 tiros)	Partido 4 (10 tiros)
Número de encestes:	Número de encestes:	Número de encestes:	Número de encestes:
Número de fallos:	Número de fallos:	Número de fallos:	Número de fallos:

---

# Secondary school students' understanding the expected value and variability of sample proportion

Nuria Begué

Doctorado en Ciencias de la Educación, Facultad de Ciencias de la Educación, Universidad de Granada  
nbugue@correo.ugr.es

Carmen Batanero, María M. Gea

Departamento de Didáctica de la Matemática, Facultad de Ciencias de la Educación, Universidad de Granada  
batanero@ugr.es, mmgea@ugr.es

Sampling is a fundamental idea in statistics, since it establishes a link between probability and variability, and is the base of statistical inference; moreover, ideas of sampling underlie simulation and the frequentist approach to probability, which are included in the secondary school curricula as well as sampling itself.

The aim of this research was to analyse secondary school students' intuitive understanding of some properties of sampling. More specifically, we intended to evaluate the students' understanding of the relationships between the population proportion and the expected value of a sample proportion, as well as their understanding of the sample proportion variability in relation to the sample size.

The data were collected through a questionnaire given to 302 students from different schools in Huesca, Spain; 157 of them were second grade students (13-14 years old) and 145 fourth grade students (15-16 years old). The questionnaire was composed of four open-ended items, in each of which four probable values for the number of outcomes for a given event were requested to the students. The proportion population and the sample size are systematically varied (in combining sample sizes of  $n=100$  and  $n=1$  with population proportions  $p=0,5$  and  $0,7$ ).

To evaluate the students' understanding of expected value, we analysed the distribution of average value of the four provided by the students in each item and classified these averages in normative, acceptable, and biased expected values. This classification took into account the values where the theoretical sample means range with different probabilities.

Students were almost precise in their estimation of the expected value in most items, and, according to previous research we interpreted that they achieved a proportional level of reasoning about sampling. This accuracy in the students' estimations also suggests a good understanding of the relationships between the population and sample proportions. Some students, however, provided estimations that reflected the equiprobability bias or else reasoned according to negative or positive recency when generating samples from populations with non-equiprobable events.

To evaluate the students' understanding of variability, we studied the range of the four estimates provided by the students in each item and classified these ranges in normative, acceptable, and excessive variability as well as excessive concentration, as compared with the magnitude of the range in a theoretical sampling distribution. The variability of the sample proportion was overestimated in big samples and depended on the problem context in small samples. Consequently, only a part of students achieved the distributional level of reasoning about sampling. This excessive variability of estimates for big samples suggests a lack of intuition of the Law of Large Numbers, which contradicts results from previous research, explained by the difference in the task proposed to the student. In addition, some students produced samples with a deterministic pattern. Fourth grade students achieve better results than those in second grade for all the items.

