# Niveles de razonamiento probabilístico de estudiantes de bachillerato frente a una situación básica de variable aleatoria y distribución

*García, Jaime y Sánchez, Ernesto*

Cinvestav

**Resumen**

En el presente artículo se estudia el razonamiento probabilístico de estudiantes de bachillerato a través de sus respuestas a preguntas referidas a una situación básica inscrita en el tema de variable aleatoria y distribución de probabilidades. Un cuestionario de 13 preguntas, sobre una situación-problema llamada *A la Suerte*, se aplicó a 24 estudiantes. El problema era predecir el comportamiento de sorteos de tipo binomial (n = 3, p = ½) con el fin de observar si los estudiantes perciben y consideran aspectos de dos elementos principales, a saber: aleatoriedad y la distribución de los resultados (0, 1, 2). Las respuestas se analizaron desde la taxonomía SOLO, es decir, se organizan en cinco categorías: Prestructural, Uniestructural, Multiestructural, Relacional y Abstracto Extendido. El análisis muestra que la mayoría de los estudiantes perciben la característica de la binomial de tener mayor probabilidad en valores centrales que en sus valores extremos y poco menos de la mitad expresa en sus gráficas la aleatoriedad del fenómeno.

**Palabras clave**: Razonamiento Probabilístico; Distribución; Variable Aleatoria; Aleatoriedad.

.

## 1. Introducción

El aprendizaje del razonamiento probabilístico es uno de los objetivos importantes de la educación matemática de los niveles básicos y pre-universitarios (Jones, Langrall y Mooney, 2007). La probabilidad es una de las ideas fundamentales de la estadística (Burril y Bielher, 2011). Se han llevado a cabo una gran cantidad de investigaciones sobre razonamiento probabilístico que se centran en alguno de los conceptos o procedimientos que forman el campo probabilístico: aleatoriedad, espacio muestral, enfoque clásico y enfoque frecuencial de probabilidad, combinatoria, etc. No obstante, se han hecho pocas exploraciones sobre problemas que intenten desarrollar el razonamiento de probabilidad tomando como punto de partida la noción de distribución de probabilidad.

Pfannkuch y Reading (2006) formulan 11 preguntas para indicar hacia dónde debieran los investigadores dirigir sus esfuerzos en lo que concierne al desarrollo de la noción de distribución por los estudiantes, entre ellas se encuentran las siguientes: ¿Cuáles son las formas y representaciones más simples de una distribución que pueden entender los estudiantes? ¿Cómo se desarrolla el razonamiento acerca de distribuciones desde sus formas o aspectos más simples a unos más complejos? ¿Cuáles son las dificultades que los estudiantes encuentran cuando trabajan, analizan e interpretan distribuciones? En este estudio se propone una situación de probabilidad en la que se busca avanzar en la respuesta a la segunda de estas preguntas: el razonamiento en una situación simple de distribución binomial y con estudiantes de bachillerato. Su respuesta puede ofrecer claves para responder las otras.

## 2. Marco Conceptual

El *contenido* que se explora en este estudio es la noción de *distribución binomial* que abarca el de la *variable aleatoria* correspondiente. Ésta presenta un patrón que varía en los

En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria (pp. 417-424). Granada, Departamento de Didáctica de la Matemática de la Universidad de Granada, 2013.

valores enteros entre cero y N, y se caracteriza por tener pocas ocurrencias en sus valores extremos y más ocurrencias conforme la variable se desplaza hacia un valor intermedio. Esta distribución se forma al repetir N veces un experimento de Bernoulli y observar el número de éxitos que se presentan. El espacio muestral asociado a la distribución está formado por $2^N$ secuencias de $E^s$ (éxitos) y $F^s$ (fracasos) de tamaño N; la variable aleatoria “el número de éxitos” genera una partición de eventos del espacio muestral cuyas probabilidades se distribuyen en forma binomial.

Se entiende por *razonamiento* al proceso de formular juicios o aserciones a partir de otras proposiciones ya conocidas u observaciones de un fenómeno. Asumimos que la gente se forma conceptos como consecuencia de su participación en las actividades sociales de dar y pedir razones, es decir, gracias al razonamiento se configuran y consolidan los conceptos (Brandom, 2002, Bakker y Derry, 2011). En consecuencia, conviene proponer tareas o situaciones-problema que propicien el razonamiento de los estudiantes; por ejemplo, en el campo de la probabilidad, tareas en situaciones de incertidumbre que logren captar el interés de los estudiantes, que lo enfrenten a algunas ‘grandes ideas’ de la probabilidad (Gal, 2005) y que le exijan realizar razonamientos probabilísticos; tales tareas deben involucrar a los estudiantes en procesos de estimación de la propensión de ocurrencia de eventos y de su predicción y representación y no restringirse sólo a problemas de cálculo.

Una manera de describir el razonamiento acerca de la distribución binomial desde sus formas o aspectos más simples a otras algo más complejos es analizando y organizando en una *jerarquía* las respuestas de estudiantes a una tarea. La taxonomía SOLO es útil para tal propósito. Ésta se propuso para evaluar la calidad de los aprendizajes de los estudiantes, pero ha sido también una herramienta para elaborar jerarquías de desarrollo del razonamiento (Jones, Thornton, Langrall y Tarr, 1999). Permite identificar elementos de conocimiento que los estudiantes utilizan como recursos para formar razonamientos. El propósito del presente artículo es describir el razonamiento de los estudiantes frente a una tarea binomial simple a través de una jerarquía utilizando la taxonomía SOLO.

## 3. Metodología

*Participantes.* Veinticuatro estudiantes de un grupo de primer semestre (15-16 años) del CCH-UNAM, el profesor titular del grupo y dos investigadores (autores del presente artículo). Los estudiantes no habían recibido ningún tipo de enseñanza formal de probabilidad y estadística; además, no recibieron información respecto el propósito del estudio.

*Instrumento.* Un cuestionario de 13 preguntas, sobre una situación-problema inscrita en el tema de variable aleatoria y distribución de probabilidades llamada *“A la Suerte”*. El problema era predecir el comportamiento de sorteos de tipo binomial (n = 3, p = ½) con el fin de observar si los estudiantes perciben y consideran aspectos de dos elementos principales, a saber: aleatoriedad y la distribución de los resultados (0, 1, 2). La situación-problema estaba relacionada con el sorteo del control de la TV en una familia de tres integrantes (Ana, Carlos y Beto) a través de un mecanismo aleatorio, el lanzamiento de dos monedas, de tal forma que si no salían águilas ganaba Ana, si salía una águila ganaba Beto y si salían dos águilas ganaba Carlos. La finalidad del problema era predecir el comportamiento de los sorteos para decidir si todos los miembros de la familia podían tener la misma frecuencia de éxito a largo plazo. Por las limitaciones de espacio, en este informe sólo se expondrá y comentará la pregunta 6 de la situación-problema, la cual se presenta en la Figura 1.

6. ¿Cómo piensas que serían los resultados para 40 días? Dibuja las barras de los posibles resultados en cada inciso, teniendo en cuenta la escala dada. Escribe el número de veces o el porcentaje, según sea el caso, sobre cada barra.

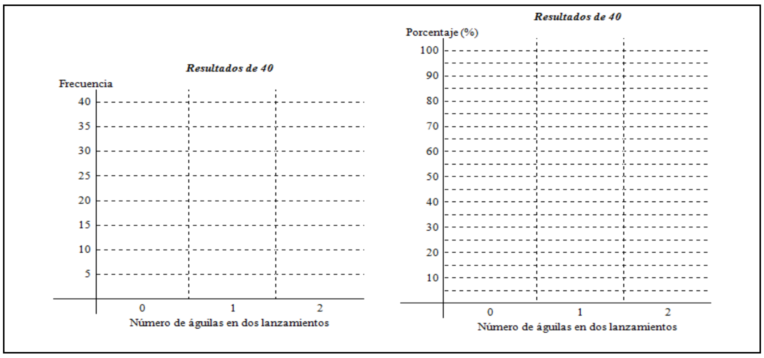



Figura 1. Pregunta 6

*Procedimientos.* El docente presenta a sus estudiantes la situación-problema, les pide leerlas en forma individual y grupal, les permite que realicen algunos ensayos para propiciar una familiarización con la situación aleatoria y puedan opinar, hacer predicciones, argumentar y proponer ideas en el cuestionario. La aplicación duró 1 hora aproximadamente.

## 4. Análisis y Resultados

Una forma de percibir y considerar aspectos de dos elementos principales, a saber: aleatoriedad y la distribución de resultados (0, 1, 2), es observar los gráficos propuestos por los estudiantes ante una predicción.

Para la pregunta 6, referente a predicciones a corto plazo sobre distribuciones de frecuencia en términos de valores absolutos y relativos (%) del experimento del lanzamiento de dos monedas para 40 sorteos, se tuvieron en cuenta las siguientes componentes para evaluar la respuesta de cada estudiante:

- *Componente (a).* Distribuye las frecuencias absolutas y relativas teniendo en cuenta el número de sorteos (la suma de frecuencias es igual al tamaño de la muestra) y el 100% (la suma de las frecuencias relativas es igual al 100%), respectivamente.
- *Componente (b).* Asigna mayor frecuencia absoluta y relativa al evento con mayor probabilidad (1 águila) (mediante un enfoque clásico o mediante una apreciación subjetiva).
- *Componente (c).* Asigna cuando menos dos distintas frecuencias a los valores esperados de los eventos, en ambos gráficos, considerando la variabilidad. Esta componente es controvertida porque los estudiantes pueden asignar valores distintos a los esperados por causas diferentes a la de considerar la variabilidad. Hay quienes considerarían una expresión de mayor conocimiento aquellas respuestas que reflejaran los valores esperados, pero como ha puesto en evidencia Shaughnessy (1997) ésta no es necesariamente una respuesta que tenga en cuenta la naturaleza aleatoria del experimento.
- *Componente (d).* Representa adecuadamente los valores de frecuencia entre ambos gráficos, conservando la proporción de los resultados en las dos escalas.

Con el fin de analizar y clasificar las respuestas dadas por los estudiantes a la pregunta 6 (consideraron las cuatro componentes antes descritas) se tienen los siguientes niveles de la taxonomía SOLO:

- *Prestructural (P)*: No tiene en cuenta ninguna de las componentes de la pregunta.
- *Uniestructural (U)*: Tiene en cuenta una sola de las cuatro componentes de la pregunta: “a”, “b”, “c” o “d”.
- *Multiestructural (M)*: Tiene en cuenta sólo dos de las cuatro componentes de la pregunta: “ab”, “ac”, “ad”, “bc”, “bd” o “cd”.

- *Relacional (R)*: Tiene en cuenta sólo tres de las cuatro componentes de la pregunta: “abc”, “abd”, “acd” o “bcd”.
- *Abstracto Extendido (AE)*: Tiene en cuenta las cuatro componentes de la pregunta: “abcd”.

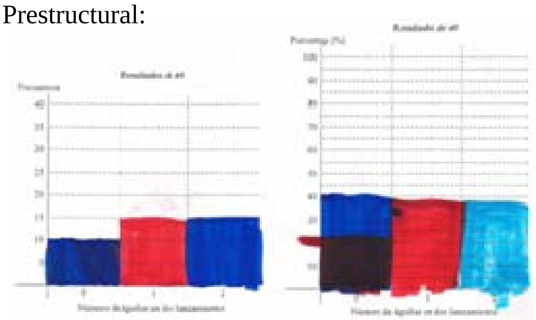
La Tabla 1 presenta la clasificación de los estudiantes de acuerdo a su nivel y a las componentes involucradas en su respuesta. En la Tabla 2 se presenta una ejemplificación del tipo de respuestas dadas por los estudiantes de acuerdo al nivel de razonamiento, seguida de una breve justificación sobre esta clasificación.

Tabla 1. Clasificación de acuerdo a nivel y componente

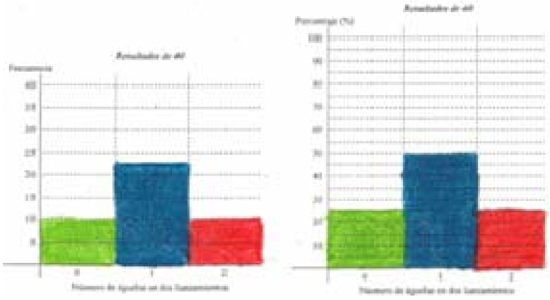
| Nivel | Componente | Participantes | Total |
|---|---|---|---|
| Prestructural | - | 2 | 2 |
| Uniestructural | b | 5 | 8 |
| | c | 3 | |
| Multiestructural | ab | 1 | 4 |
| | bc | 2 | |
| | ac | 1 | |
| Relacional | abd | 6 | 9 |
| | abc | 2 | |
| | bcd | 1 | |
| Abstracto Extendido | abcd | 1 | 1 |
| Total | | 24 | 24 |

Como se observa, respecto a la primera componente (a), 11/24 estudiantes distribuyen adecuadamente las frecuencias teniendo en cuenta que la suma de frecuencias absolutas y relativas asignadas corresponden al tamaño de la muestra (40 sorteos) y al 100%, respectivamente; respecto a la segunda componente (b), 18/24 estudiantes asignaron mayor frecuencia al evento más probable (consciente o inconscientemente) en las dos distribuciones; respecto al tercer componente (c), 10/34 estudiantes tuvieron en cuenta en asignar cuando menos dos distintas frecuencias a los valores esperados de los eventos en ambas distribuciones considerando la variabilidad; finalmente, respecto a la cuarta componente (d), sólo 8/24 presentan adecuadamente los valores de frecuencia entre ambos gráficos, conservando la proporción adecuada de sus resultados en ambas escalas. Esto refleja una gran dificultad en el manejo de esta última componente.

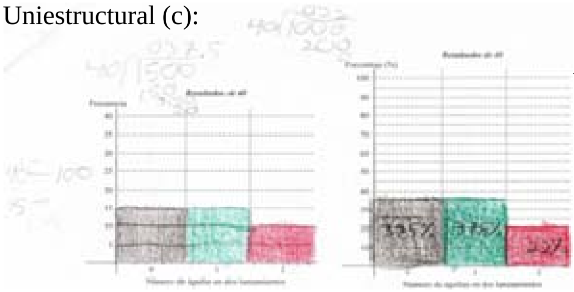
Tabla 1. Ejemplificación de acuerdo a nivel y componente

| Nivel y Componente con Respuesta del Estudiante | Descripción |
|---|---|
| Prestructural:  | La suma de las frecuencias absolutas es el número total de sorteos; sin embargo, no tiene en cuenta que la suma de frecuencias relativas es el 100%. No considera la variabilidad. Manifiesta sesgo de equiprobabilidad al asignar mismas frecuencias relativas. No asigna mayor frecuencia absoluta y relativa al evento con mayor probabilidad. El segundo gráfico no es proporcional al primero.<br>Se favorece al evento más probable (b), pero no tiene en cuenta que la suma de las frecuencias sea el número total de sorteos, ni la segundo gráfico es proporcional al primero. No ofrece rasgos que indiquen la variabilidad. |

Uniestructural (b):

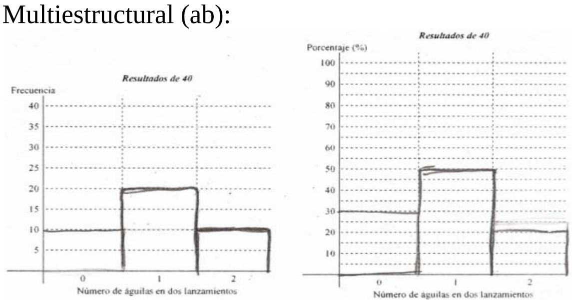


Uniestructural (c):



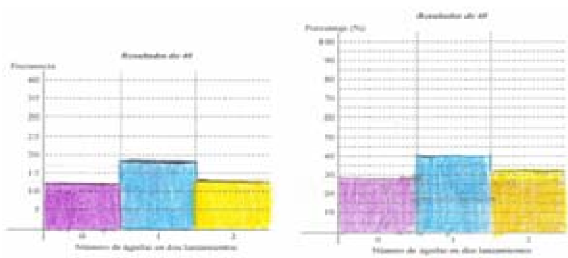
Considera la variabilidad al manifestar al menos dos distintas frecuencias absolutas y relativas a los valores esperados (c). No tiene en cuenta que la suma de frecuencias relativas es el 100%. No favorece al evento más probable. El segundo gráfico no es proporcional al primero.

Multiestructural (ab):



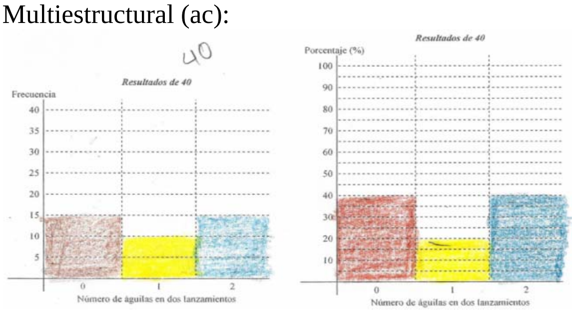

La suma de frecuencias absolutas y relativas es correcta (a), y favorece al evento más probable (b). Considera la variabilidad sólo en el gráfico de frecuencias relativas. El segundo gráfico no es proporcional al primero.

Multiestructural (bc):



Propone gráficos muy similares. Favorece al evento más probable (b) y considera la variabilidad al asignar dos distintas frecuencias a los valores esperados (c). No tiene en cuenta que la suma de frecuencias asignadas corresponda al tamaño de la muestra (40 sorteos). El segundo gráfico no es proporcional al primero.

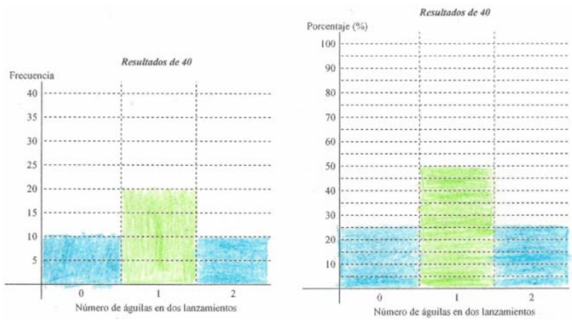
Multiestructural (ac):



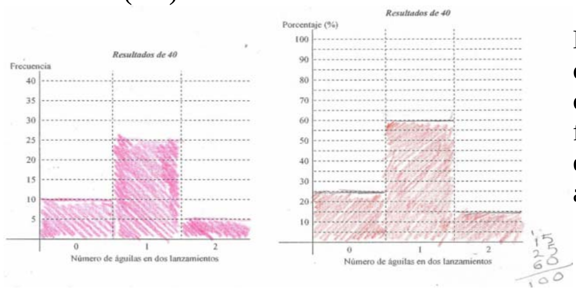

La suma de frecuencias absolutas y relativas es correcta (a), y considera la variabilidad al asignar distintas frecuencias a los valores esperados (c). No favorece el evento más probable (1 águila), al contrario favorece a los eventos 0 águilas y 2 águilas. Sus gráficos no conservan proporción en sus resultados en ambas escalas.

Relacional (abd):

La suma de frecuencias absolutas y relativas es correcta (a), favorece al evento más probable (b) y sus gráficos conservan una proporción adecuada en sus resultados en ambas escalas (d). Sin embargo, no considera la variabilidad al no asignar frecuencias distintas a los valores esperados.

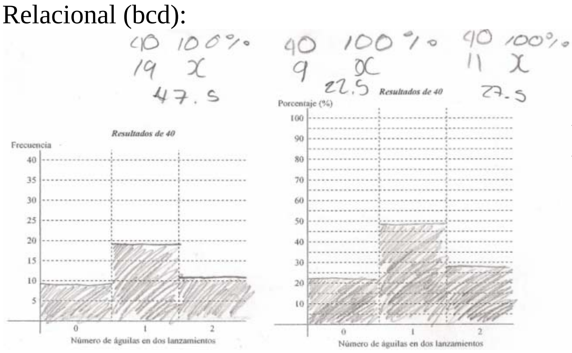



Relacional (abc):



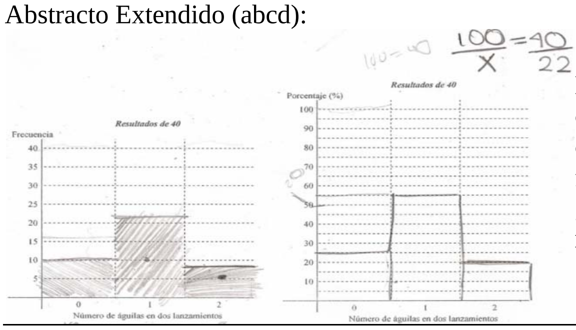

La suma de frecuencias absolutas y relativas es correcta (a), favorece al evento más probable (b) y considera la variabilidad al asignar distintas frecuencias a los valores esperados (c). Sin embargo, sus gráficos no conservan proporción adecuada en sus resultados en ambas escalas.

Relacional (bcd):




Favorece al evento más probable (b), considera la variabilidad al asignar distintas frecuencias a los valores esperados y sus gráficos conservan una proporción adecuada en sus resultados en ambas escalas (d). Sin embargo, la suma de frecuencias relativas es incorrecta.

Abstracto Extendido (abcd):




La suma de frecuencias absolutas y relativas es correcta (a), favorece al evento más probable (b), considera la variabilidad al asignar distintas frecuencias a los valores esperados (c) y sus gráficos conservan una proporción adecuada en sus resultados en ambas escalas (d).

Analizando los niveles de respuesta, se observa que la mayor parte de los estudiantes reflejan respuestas en el nivel relacional (38%) y en el nivel uniestructural (33%); pocos reflejan respuestas en el nivel multiestructural (17%) y en el nivel prestructural (8%); y sólo un estudiante (4%) refleja su respuesta en un nivel abstracto extendido.

## 5. Discusión y Conclusiones

En este estudio se puede apreciar cómo los estudiantes de nivel bachillerato expresan mediante un gráfico sus creencias o conocimientos sobre la situación en la que se pone en juego una distribución binomial simple. Se observa que 18 estudiantes del grupo encuestado tiene la noción de que "obtener un águila" o el valor "1" es el evento más probable, lo cual es una idea importante ligada a la distribución binomial. Sólo la mitad de los estudiantes tiene cuidado de reflejar en la gráfica que la suma de las frecuencias absolutas es 40; se deben tomar precauciones en la enseñanza para tratar con este descuido, pues puede dificultar la comprensión de que la suma de las probabilidades de todos los valores de una distribución debe ser 1. La manera en que observamos la componente que hemos llamado la 'consideración de la variación' puede suscitar controversias ya que su manifestación puede ser resultado de causas diferentes a la precepción del estudiante de aleatoriedad y la variación correspondiente. No obstante, pensamos que algunos estudiantes pueden expresar valores aproximados a los esperados teniendo en cuenta que es una experiencia aleatoria y no es natural esperar que en 40 repeticiones ocurran exactamente el número de eventos que predice la probabilidad. Con esto en mente, 10 estudiantes reflejaron en su propuesta variantes ligeras a los valores esperados. Finalmente, la componente en que menos acertaron los estudiantes, sólo 8 de 24, fue en controlar que sus dos gráficos fueran consistentes, una teniendo en cuenta como base 40 ensayos y la otra en porcentajes. Es probable que muchos hayan asumido que no tenían por qué hacerlas proporcionales si no sólo aproximadas, otros quizá sí tengan problemas para traducir las frecuencias a sus respectivas proporciones. Esta operación de normalizar las frecuencias absolutas es de gran importancia para entender las distribuciones de probabilidad.

La mayoría de los estudiantes perciben la característica de la binomial de tener mayor probabilidad en valores centrales que en sus valores extremos y poco menos de la mitad expresa en sus gráficas la aleatoriedad del fenómeno. Enfatizamos este aspecto porque consideramos que una comprensión adecuada de una distribución de probabilidades nunca debe separarse de la idea de que expresa un fenómeno de incertidumbre en cuyas manifestaciones (repeticiones de sorteos de acuerdo a la distribución) se debe reflejar la aleatoriedad y la variación. Las otras componentes también es importante señalarlas para que la noción de distribución se construya de manera coherente.

## Referencias


Bakker, A. y Gravemeijer, K. P. E. (2004). Leraning to reason about distribution. En D. Ben – Zvi & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking.* Dordrecht, The Netherlands: Kluwer Academic Press.

Bakker, A. y Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*(1-2), 15-26.

Burril, G. y Bielher, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burril, C. Reading (Eds.). *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study: The 18th ICMI Study.* New York: Springer.

Brandom, R.B. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge, MA: Harvard University Press.

Gal. I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. En G. A. Jones (Ed.), *Exploring probability in school. challenges for teaching and learning* (pp. 39-63). New York: Springer.

Jones, G.A., Langrall, C.W. y Mooney, E.S. (2007). Research in probability. En F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909-955). Charlotte, NC, USA: Information Age-NCTM.

Jones, G.A., Thornton, C.A., Langrall, C.W., y Tarr, J.E. (1999). Understanding students' probabilistic reasoning. En L.V. Stiff y F.R. Curcio, (Eds.), *Developing mathematical reasoning in grades K-12* [1999 Yearbook]. Reston, VA: NCTM.

Pfannkuch y Reading (2006). Reasoning about distribution: a complex process. *Statistical Education Research Journal*, *5*(2), pp. 4-9.

Shaughnessy, J.M. (1997). Missed opportunities in research on the teaching and learning of data and chance. En F. Biddulph y K. Carr (Eds.), *People in Mathematics Education* (Proceedings of the $20^{th}$ annual conference of Mathematics Education Research Group of Australasia, vol

1, pp. 6-22). Waikato, NZ: MERGA.