

DISEÑO DE UNA PRUEBA PARA EVALUAR EL RAZONAMIENTO ESTADÍSTICO EN ESTUDIANTES UNIVERSITARIOS

Iris Alicia Meza Muñoz

Universidad Autónoma de Aguascalientes
Aguascalientes, México
ORCID: 0000-0003-1037-4316.

Daniel Eudave Muñoz

Universidad Autónoma de Aguascalientes
Aguascalientes, México
ORCID: 0000-0003-4070-3109

Adán Moisés García Medina

Comisión Nacional para la Mejora
Continua de la Educación (Mejoredu).
Ciudad de México, México,
ORCID: 0000-0002-2614-8471

Rogelio Salinas Gutiérrez

Universidad Autónoma de Aguascalientes
Aguascalientes, México
ORCID: 0000-0002-1669-4460

All content in this magazine is licensed under a Creative Commons Attribution License. Attribution-Non-Commercial-Non-Derivatives 4.0 International (CC BY-NC-ND 4.0).



Resumen: Este artículo presenta los resultados de una investigación de corte metodológico donde el objetivo, fundamentación, procedimiento y resultados tiene como propósito el diseño de una prueba que permita evaluar el razonamiento estadístico en estudiantes universitarios. La metodología que se adaptó para el diseño del instrumento es la de construcción de pruebas estandarizadas propuesta por Lane, Raymond, Haladyna y Downing (2016), en ella se establece el proceso para poder diseñar pruebas a gran escala que cumplan con los lineamientos calidad establecidos. El resultado es el Instrumento de Evaluación Diagnóstica del Razonamiento Estadístico (IEDRE), validado de acuerdo con los criterios metodológicos establecidos.

Palabras clave: Razonamiento estadístico; instrumentos de evaluación; diseño de pruebas estandarizadas.

INTRODUCCIÓN

La estadística es la ciencia de aprender de los datos y de medir, controlar y comunicar la incertidumbre, además que, la mayoría de las disciplinas científicas formales con el apoyo de la estadística describen y analizan elementos particulares de la vida y el mundo, desde la naturaleza física del universo hasta el cómo funcionan las sociedades; por lo tanto, la estadística es considerada como una metadisciplina que incorpora principios metodológicos que permiten convertir los datos en ideas reales en cualquier contexto (Wild, Utts y Horton, 2018).

Con base en lo anterior, el razonamiento estadístico puede definirse como la forma en que las personas razonan con ideas estadísticas y le dan sentido a la información estadística. Esto implica realizar interpretaciones basadas en conjuntos de datos, representaciones de datos o resúmenes estadísticos de datos. El razonamiento estadístico puede implicar la conexión de un concepto con otro (por ejemplo,

centro y dispersión), o puede combinar ideas sobre los datos y el azar. Razonar estadísticamente significa comprender y ser capaz de explicar los procesos estadísticos para interpretar plenamente los resultados estadísticos (Ben-Zvi y Garfield, 2004, p.7).

La ASA (2016) generó una versión de las Directrices de la Evaluación e Instrucción en Educación Estadística (Guidelines for Assessment and Instruction in Statistics Education, GAISE). En dicho texto, la ASA (2016) reconoce la existencia de múltiples programas de introducción a la estadística en el nivel universitario, por lo que establece nueve metas a alcanzar por los estudiantes al finalizar un curso introductorio de estadística en el nivel superior. Brevemente se describen cada una de las metas propuestas por la ASA (2016, pp. 8-11) y que resumen lo que un estudiante universitario debería saber y comprender al final de un primer curso de estadística:

1. Deben convertirse en consumidores críticos de los resultados estadísticos informados en los medios populares, reconociendo si los resultados informados se derivan razonablemente del estudio y análisis realizado.
2. Deben ser capaces de reconocer preguntas para las que el proceso de investigación en estadística sea útil y deben poder responder a preguntas utilizando el proceso de investigación.
3. Deben poder producir presentaciones gráficas y resúmenes numéricos e interpretar lo que los gráficos revelan y no revelan.
4. Deben reconocer y ser capaces de explicar el papel central de la variabilidad en el campo de la estadística.
5. Deben reconocer y ser capaces de explicar el papel central de la aleatoriedad al diseñar estudios y sacar conclusiones.

6. Deben adquirir experiencia sobre cómo se utilizan los modelos estadísticos, incluidos los modelos multivariantes.

7. Deben demostrar comprensión y capacidad para utilizar las ideas básicas de inferencia estadística, tanto pruebas de hipótesis como estimación de intervalos, en una variedad de entornos.

8. Deben poder interpretar y sacar conclusiones de los resultados estándar de los paquetes de software estadístico.

9. Deben demostrar conocimiento de las cuestiones éticas asociadas con una práctica estadística sólida.

En este sentido, el objetivo general de la presente investigación fue el desarrollar un instrumento de calidad que cumpliera con la metodología que rige el diseño de las pruebas estandarizadas; dicha prueba puede ser utilizada por los profesores de estadística como un recurso complementario al de su propio proceso de evaluación ya que la prueba es una herramienta que puede ayudarlos para identificar áreas de mejora y tomar decisiones que favorezcan la enseñanza y aprendizaje de la estadística en el nivel superior.

METODOLOGÍA

Para el presente estudio se retomó la propuesta de Lane et al. (2016) quienes describen a través de doce fases el proceso para el desarrollo de pruebas educativas y psicológicas. Adicionalmente se revisaron los trabajos realizados por Backhoff, Peón y Sánchez (2005), Backhoff, Peón, Sánchez y Andrade (2006), Backhoff, Solano, Peón, Sánchez y Prieto (2007) y Backhoff et al. (2009) para acotar los alcances del presente estudio y establecer las fases directrices en el desarrollo de la prueba para evaluar el razonamiento estadístico en estudiantes universitarios. La Tabla 1 muestra el proceso general para diseñar, construir y validar el IEDRE, la tabla

se compone de cada una de las fases con sus etapas, participantes externos, así como los productos que se obtuvieron durante todo el proceso.

RESULTADOS

A. EL PLAN GENERAL

El plan general es la primera fase para el diseño de pruebas y se considera la más importante de todo el proceso de diseño, porque proporciona un marco sistemático para todas las actividades principales asociadas con el desarrollo de pruebas. Para Lane et al. (2016) los elementos que guían el diseño del plan general son los que se definen en la Tabla 2, así mismo en dicha tabla se describen los elementos del plan general que orientaron el presente estudio.

Backhoff et al. (2009) señalan que para el desarrollo de pruebas de calidad es necesario contar con grupos de expertos denominados comités técnicos. Para los fines de esta prueba se consideraron tres comités técnicos, el primero de ellos el comité académico, el segundo el comité de construcción de reactivos y finalmente el comité de validación y sesgo.

B. DEFINICIÓN DEL DOMINIO Y OPERACIONALIZACIÓN DEL CONSTRUCTO

Llevar a cabo la segunda fase en el desarrollo del IEDRE implicó definir el dominio del constructo a evaluar en la prueba diagnóstica. Para obtener la tabla de contenidos de la prueba, primero fue necesario realizar una tabla de operacionalización del concepto de razonamiento estadístico mismo que se presentó al comité académico para su análisis, revisión y valoración.

Elaborar la tabla de operacionalización del constructo requirió el considerar diferentes elementos, entre ellos, la revisión de autores acerca de la definición del razonamiento

Fases	Etapas	Participantes	Productos
I. Plan general	1) Plan general		A. Plan general del IEDRE
II. Definición del dominio y operacionalización del constructo	2) Definición de la tabla de operacionalización del constructo		B. Tabla de operacionalización del constructo.
	3) Definición de la tabla de contenidos de la prueba	Comité técnico académico	C. Tabla de contenidos del IEDRE
III. Especificaciones de reactivos	4) Elaboración de las especificaciones de reactivos		D. Especificaciones de reactivos de la prueba
IV. Diseño de reactivos	5) Elaboración de reactivos	Comité técnico de construcción de reactivos	E. Manual técnico para la construcción de reactivos
	6) Prueba empírica de los reactivos a pequeña escala (micropilotaje)		F. Reactivos por especificación
	7) Jueceo de los reactivos por parte del comité de validación y sesgo	Comité técnico de validación y sesgo	G. Reporte de valoración para los reactivos por sesión de micropilotaje
V. Conformación del IEDRE	8) Ensamblaje del instrumento		H. Manual técnico para el jueceo de reactivos
	9) Pilotaje de reactivos		I. Dos reportes de validación por cada reactivo
	10) Análisis de los resultados del pilotaje	Asesor especialista en medición	J. Revisión colegiada de los reactivos
	11) Selección de reactivos e integración de módulos		K. Cuadernillos y hojas de respuestas impresos del IEDRE
			L. Bases de datos del pilotaje (por caso/sustentante y por reactivo)
			M. Informe técnico sobre el comportamiento psicométrico de los reactivos (análisis TRI)
			N. IEDRE con módulos de reactivos para su aplicación masiva

Nota: Elaboración propia a partir de la revisión Lane et al. (2016, pp. 3-18) y Backhoof et al. (2006, pp. 5-7).

Tabla 1. Proceso general para diseñar, construir y validar el IEDRE

Elemento del plan general	Descripción
1. Constructo por medir	Razonamiento estadístico.
2. Población a la que está destinada la prueba	Estudiantes universitarios que han cursado por lo menos una asignatura de estadística en su formación superior.
3. Usuarios de la prueba	Profesores y/o autoridades educativas del área de estadística.
4. Interpretaciones y usos previstos de las puntuaciones de las pruebas.	Toma de decisiones para la mejora de los procesos de enseñanza-aprendizaje de los contenidos relacionados con el dominio del razonamiento estadístico.
5. Los contenidos de la prueba, demandas cognitivas y formato que apoyarán las interpretaciones y usos previsto.	Taxonomía de Anderson y Krathwohl (2001). Se establecen niveles de logro de acuerdo con el puntaje obtenido en la prueba.
6. Tipo de reactivos	Reactivos de respuesta seleccionada (RS) del tipo de opción múltiple (MC).
7. Procedimientos de administración de la prueba.	Aplicación en papel y lápiz.
8. Procedimientos e informe de las puntuaciones.	Informe a los profesores y/o autoridades del área de estadística de un puntaje total desagregado por niveles de logro.
9. Especificaciones de las interpretaciones y usos previstos de la prueba.	Al establecer niveles de logro se puede identificar el alcance que tienen los estudiantes al concluir un primer curso de estadística, con la finalidad de reconocer áreas de bajo desarrollo para que los maestros puedan tomar la decisión de mejorar aspectos de la enseñanza-aprendizaje de contenidos que se relacionan con el razonamiento estadístico.

Nota: Elaboración propia a partir de la revisión Lane et al. (2016, pp. 3-6)

Tabla 2. Plan general para el IEDRE

estadístico (Garfield y Gal, 1997, Garfield y Gal, 1999; Garfield, 2002; Ben-Zvi y Garfield, 2004; DelMas, 2004), las directrices que aporta el GAISE (ASA, 2016) para los aprendizajes esperados en estudiantes universitarios después de haber tomado un primer curso de estadística; los tópicos que abordan los instrumentos que han evaluado el razonamiento estadístico y la revisión de programas de materia de estadística de varias universidades.

La Tabla 3 muestra la tabla de operacionalización del constructo del razonamiento estadístico que se presentó al Comité Académico para su revisión, valoración y aprobación, en ella se describen las competencias del razonamiento estadístico y las dimensiones que son posible evaluar.

Los resultados que se obtuvieron del trabajo con el comité académico fueron los siguientes:

a) El comité académico no agregó ningún otro contenido a evaluar a la tabla de operacionalización del constructo.

b) El comité académico determinó con base en su experiencia docente, que solo es posible evaluar a través del IEDRE la dimensión de *Estadística Descriptiva*, tomando en consideración los contenidos que se abordan con estudiantes que solo llevan una materia de estadística en su formación universitaria, sin embargo, es posible dejar las dimensiones de *Probabilidad y Estadística Inferencial* como módulos adicionales para programas educativos que llevan más de una materia de estadística.

c) El comité académico acordó en consenso que la dimensión *El papel de la estadística en el proceso de la investigación científica* debería ser eliminada ya que solo algunos programas educativos contienen materias del tipo metodología de la investigación en su malla curricular.

C. ESPECIFICACIÓN DE REACTIVOS

Las especificaciones de reactivos proporcionan una definición operativa del contenido de la prueba y las demandas cognitivas de las tareas de evaluación (AERA et al., 2018). A partir de obtener la tabla de contenidos aprobada por el comité académico, el siguiente paso consistió en diseñar cada una de las especificaciones de contenido, el IEDRE contó con 20 especificaciones, 13 para la dimensión de Estadística Descriptiva, 4 para la dimensión de Probabilidad y finalmente 3 para la dimensión de Estadística Inferencial, como se puede apreciar en la Tabla 4 a través de algunos ejemplos de especificaciones.

D. DISEÑO DE REACTIVOS

La cuarta fase en el desarrollo de pruebas es uno de los elementos más importantes para el desarrollo de la prueba diagnóstico, porque es en ella donde se desarrollan los reactivos que contendrá. Para ello se requirió la conformación de un comité de construcción de reactivos quienes diseñaron los reactivos del IEDRE, para llevar a cabo la prueba empírica de los reactivos elaborados a pequeña escala (micropilotaje) y finalmente el jueceo de los reactivos.

Para conocer si los estudiantes que aplicarán la prueba tendrán alguna dificultad o no con los reactivos de esta, es necesario probarlos empíricamente. La prueba empírica de los reactivos del IEDRE se realizó en sesiones de trabajo con grupos de estudiantes que asistían de forma presencial a las clases. A partir de los comentarios y resultados obtenidos en las sesiones de aplicación de micropilotaje se corrigieron los reactivos y con ello se obtuvieron 96 reactivos que pasaron a la siguiente etapa de jueceo de reactivos.

El jueceo de los reactivos fue un procedimiento que implicó la emisión de un juicio de valor a partir uno criterio previamente definido, esta actividad la realizó

	Competencias del razonamiento estadístico	Dimensiones
Desarrollar una posición crítica ante la información estadística (Garfield y Gal, 1997; ASA, 2016).	Dominar las habilidades procedimentales estadísticas (Garfield y Gal, 1997).	Estadística descriptiva
	Comprender los elementos básicos del azar y la probabilidad (Garfield y Gal, 1997).	Probabilidad
	Comprender elementos esenciales de inferencia estadística (ASA, 2016).	Estadística inferencial
	Comprender el propósito, la lógica y el proceso de las investigaciones científicas que involucran elementos estadísticos (Garfield y Gal, 1997; ASA, 2016).	El papel de la estadística en el proceso de investigación científica

Nota: Elaboración propia

Tabla 3. Tabla de operacionalización del dominio razonamiento estadístico

Dimensión	Nombre de la especificación
Estadística descriptiva	1. Analizar variables por su tipo.
	2. Comprender la diferencia entre población y muestra, así como las características de una muestra representativa.
	3. Analizar el tipo de muestro probabilístico más apropiado para la recolección de información.
	4. Analizar correctamente la información presente en tablas tomando en consideración los datos representados.
Probabilidad	1. Comprender la probabilidad, como una medida de ocurrencia de un evento aleatorio, así como verificar sus propiedades en ejemplos contextualizados.
	2. Analizar la distribución de probabilidad de una variable aleatoria discreta.
Estadística inferencial	1. Aplicar el proceso para la obtención de distribución de estadísticos muestrales.
	2. Analizar parámetros poblacionales mediante la información contenida en una muestra utilizando el proceso de inferencia estadística.

Nota: Elaboración propia

Tabla 4. Ejemplos de Especificaciones de reactivos por dimensión para el dominio del Razonamiento Estadístico

Dimensión	Distribución de reactivos	Porcentaje en el examen
Estadística descriptiva (módulo 1)	43	62.32%
Probabilidad (módulo 2)	13	18.84%
Estadística inferencial (módulo 3)	13	18.84%
Total	69 reactivos	100%

Nota: Elaboración propia

Tabla 5. Estructura general de IEDRE por dimensión (módulo) para el pilotaje

Dimensión	Porcentaje en el examen	Distribución de reactivos
Estadística descriptiva (módulo 1)	65%	39
Probabilidad (módulo 2)	16.7%	10
Estadística inferencial (módulo 3)	18.3%	11
Total	100%	60

Nota: Elaboración propia

Tabla 6. Estructura general de IEDRE por dimensión para su aplicación a gran escala.

Institución educativa	Frecuencia	Porcentaje
Universidad Autónoma de Aguascalientes (UAA)	217	49.3
Instituto Tecnológico de Aguascalientes (ITA)	114	25.9
Universidad Cuauhtémoc Aguascalientes (UCA)	60	13.6
Normal Superior Federal de Aguascalientes (ENSFA)	49	11.1
Total	440	100

Nota: Elaboración propia

Tabla 7. Porcentaje de participación por institución educativa

el comité de validación y sesgo. Para el caso del IEDRE, el trabajo que realizaron los jueces fue una de las tareas más importantes en el proceso de construcción del instrumento, pues se encargaron de valorar a los reactivos diseñados por el comité de construcción de reactivos y dictaminar si son idóneos para conformar el examen.

El procedimiento de jueceo se dividió en dos etapas de trabajo. La primera etapa consistió en una valoración individual e independiente que realizaron los jueces sobre los 96 reactivos corregidos después del proceso de micropilotaje. La segunda etapa radicó en el trabajo colegiado del comité de jueces, es decir, en una sesión plenaria se reunieron a los jueces en parejas y se revisó cada uno de los reactivos que presentaron observaciones en la etapa anterior. Posterior a las sesiones de trabajo colegiado con los integrantes del comité de validación y sesgo, se corrigieron los reactivos con base en las observaciones y nuevos dictámenes de los reactivos revisados en ellas. También se ajustaron los reactivos restantes que contaban con comentarios con el propósito de conformar el IEDRE en su versión para pilotaje. La primera versión del IEDRE quedó ensamblado como se observa en la Tabla 5, donde se visualiza la cantidad de reactivos por dimensión o módulo en los cuadernillos y el porcentaje de distribución de cada uno de ellos en el examen.

El pilotaje de reactivos fue una etapa primordial en el proceso de diseño del IEDRE, en primer lugar, se gestionó el permiso y apoyo de varias instituciones de educación superior del estado de Aguascalientes. Como requisito para la aplicación del IEDRE se pidió a las instituciones que identificasen grupos de estudiantes que hubiesen cursado como mínimo una materia de estadística en su formación a nivel superior. Finalmente, cuatro de ellas accedieron a la aplicación del pilotaje de algunos de los módulos que

componen el instrumento.

Por cuestiones de la pandemia COVID-19 se tomó en consideración las condiciones de trabajo que las instituciones tenían en el periodo de aplicación y se realizaron las adaptaciones a lo que establecieron con respecto a la duración de la aplicación (máximo 60 minutos) y la existencia de grupos incompletos (asistencia de la mitad de los grupos).

Posterior al proceso de pilotaje el IEDRE quedó conformado por un total de 60 reactivos distribuidos por dimensión como lo muestra la Tabla 6 para su aplicación a gran escala.

E. PARTICIPANTES

El total de la muestra que participó en el pilotaje del IEDRE fue de 440 estudiantes que aplicaron uno o dos módulos de la prueba como se puede observar en la Tabla 7.

Con respecto al sexo de la muestra, se observa una ligera mayor participación de mujeres con un 53.41%.

DISCUSIONES

La presente investigación de corte metodológico permitió plasmar el proceso de diseño de un instrumento de evaluación del razonamiento estadístico. El desarrollo de pruebas requiere un enfoque sistemático para garantizar la validez de las interpretaciones y usos de las puntuaciones de las pruebas basadas en esas interpretaciones. Las pruebas bien construidas a partir de las cuales se obtiene información válida de acuerdo con sus fines previstos presentan el potencial de brindar beneficios sustanciales para los examinados y usuarios de la prueba (Lane et al., 2016). AERA et al. (2018) añaden “su uso adecuado puede dar lugar a mejores decisiones sobre los individuos y programas que las que se generarían sin su uso y también pueden proporcionar un camino hacia un acceso más amplio y equitativo de la educación” (p. 1).

Por otra parte, el razonamiento estadístico es una competencia genérica y transversal presente en la mayoría de los programas de pregrado de distintas instituciones de educación superior a nivel nacional. Los cursos o materias de estadística deberían de estar orientados a desarrollar competencias en los futuros profesionistas.

CONCLUSIONES

El objetivo general de la investigación estuvo centrado en diseñar una prueba que permitiera evaluar el razonamiento estadístico en estudiantes universitarios después de haber tomado como mínimo un curso de estadística dentro de su programa de licenciatura. Uno de los objetivos específicos del estudio estuvo enfocado en definir las dimensiones del razonamiento estadístico para evaluar un corpus común de conocimientos sobre el razonamiento estadístico en estudiantes universitarios. Es decir, medir de forma objetiva en qué medida (nivel) los estudiantes universitarios dominan una serie de conocimientos comunes sobre estadística delimitados por el razonamiento estadístico. En este sentido, la prueba no evalúa la calidad del trabajo docente, ni su desempeño en el aula o si cumple con un programa de materia, así como tampoco evalúa la pertinencia de los programas de materias de estadística de las instituciones de educación superior, sin embargo, el IEDRE contiene en su construcción los suficientes elementos técnicos que las pruebas estandarizadas requieren para medir el grado en que los estudiantes universitarios dominan un conjunto común de saberes sobre el razonamiento estadístico.

Es muy importante señalar que, debido a la naturaleza metodológica del objeto de investigación, no es posible incluir en el presente artículo el instrumento completo (IEDRE), porque al hacerlo de dominio de público se degradaría la validez y confiabilidad

de los resultados que se podrían obtener en su aplicación masiva. En este sentido, el IEDRE puede ser aplicado con un previo acercamiento de los interesados en ello a partir de que conozcan el propósito de su diseño y sobre todo que en su aplicación se garantice la validez de las interpretaciones y que los usos de las puntuaciones del IEDRE estén basadas en esas interpretaciones.

REFERENCIAS

- AERA, APA y NCME. (2018). *Estándares para pruebas educativas y psicológicas*. Washintong, EUA: AERA
- Anderson, L. y Krathwohl, D. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Ney York, EUA: Pearson.
- ASA (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE), College Report 2016*. EUA: ASA.
- Backhoff, E., Peón, M. y Sánchez, A. (2005). *Manual Técnico. Construcción de reactivos*. CDMX, México: INEE.
- Backhoff, E., Peón, M. y Sánchez, A. (2009). *Manual Técnico. Diseño de exámenes de la calidad y el logro educativo*. CDMX, México: INEE.
- Backhoff, E., Peón, M., Sánchez, A. y Andrade, E. (2006). *Excale. Manual Técnico. Validación de reactivos*. CDMX, México: INEE.
- Backhoff, E., Solano, G., Peón, M., Sánchez, A. y Prieto, L. (2007). *Manual Técnico. Especificaciones de Reactivos*. CDMX, México: INEE.
- Ben-Zvi, D. y Garfield, J. (2004). **Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges**. En D. Ben-Zvi, y J. Garfield, *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-18). EUA: Kluwer Academic Publishers Dordrecht.
- DelMas, R. (2004). **A Comparison of Mathematical and Statistical Reasoning**. En D. Ben-Zvi, y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 79-96). EUA: Springer.
- Garfield, J. y Gal, I. (1997). **Curricular Goals and Assessment Challenges in Statistics Education**. En J. Garfield, e I. Gal (Eds.), *The Assessment Challenge in Statistics Education*. (pp. 1-13). IOS Press. Recuperado de <https://iase-web.org/documents/book1/chapter01.pdf?1402524893>
- Garfield, J. y Gal, I. (1999). **Teaching and Assessing Statistical Reasoning**. En L. Stiff, y F. Curcio (Eds.), *Developing Mathematical Reasoning in Grades K-12*. (pp. 207-221).
- Garfield, J. (2002). **The Challenge of Developing Statistical Reasoning**. *Journal of Statistics Education*, 10(3), 1-12. Recuperado de <https://doi.org/10.1080/10691898.2002.11910676>
- Lane, S., Raymond, M., Haladyna, T. y Downing, S. (2016). **Test Development Process**. En S. Lane, M. Raymond y T. Haladyna (Eds.), *Handbook of Test Development*. EUA: Routledge.
- Wild, C., Utss, J. y Horton, N. (2018). **What is statistics?**. En D. Ben-Zvi, K. Makar, y J. Garfield, J. (Eds.), *International Handbook of Research in Statistics Education*. EUA: Springer.