



XVI
Congreso Nacional de
Investigación Educativa
CNIE-2021

Diseño y validación de una prueba diagnóstica para evaluar el Razonamiento Estadístico de estudiantes universitarios

Iris Alicia Meza Muñoz
irismeza984@gmail.com

Área temática 12. Evaluación educativa.

Línea temática: Diseño y validación de instrumentos.

Porcentaje de avance: 60%.

Trabajo de investigación educativa asociada a tesis de grado.

Programa de posgrado: Doctorado en Investigación Educativa, 4º semestre.

Institución donde realiza los estudios de posgrado: Universidad Autónoma de Aguascalientes.



Resumen

La estadística es un conocimiento que se enseñan en todos los niveles de formación básica obligatoria en México, es decir, desde la primaria hasta el bachillerato se introducen conceptos propios de estadística por ser una herramienta que ayuda a validar gran parte del conocimiento de cualquier ciencia. Más aún, desde hace ya varias décadas atrás la mayoría de los programas de formación universitaria pertenezcan o no al área de ciencias exactas, llevan en su malla curricular por lo menos una asignatura dedicada al estudio de estadística básica; bajo esta situación se podría intuir que las Instituciones de Educación Superior consideran que la estadística es una herramienta esencial para la vida profesional y personal de los estudiantes universitarios, no obstante, se ha observado que muchos de ellos concluyen sus cursos de estadística sin comprender y aplicar correctamente los conceptos en situaciones relacionadas con sus carreras o en la vida cotidiana (Batanero, Díaz, Contreras y Roa, 2013). El estudio tiene como propósito central diseñar y validar una prueba diagnóstica que permita evaluar el razonamiento estadístico en estudiantes universitarios después de haber tomado como mínimo un curso de estadística dentro de su programa de licenciatura. Teóricamente se aborda desde el campo de la educación estadística, específicamente en el área de razonamiento estadístico. Metodológicamente, se trata de una investigación con base en la construcción de instrumentos de evaluación, específicamente en el diseño de pruebas. Participarán en el estudio una muestra de estudiantes de educación superior de diversas universidades del estado de Aguascalientes.

Palabras clave: *Evaluación educativa, pruebas, estadística y educación superior.*

Introducción

Para Ben-Zvi y Garfield (2004) saber, comprender y manejar conceptos estadísticos proporciona herramientas útiles a los ciudadanos informados para responder y tomar decisiones de manera inteligente en el mundo que los rodea; sin embargo, también indican que los estudiantes de varios niveles escolares están incapacitados para pensar estadísticamente sobre situaciones importantes que afectan sus vidas.

Wild, Utts y Horton (2018) retoman la definición de la Asociación Americana de Estadística (American Statistical Association, ASA), al mencionar que la estadística es la ciencia de aprender de los datos y de medir, controlar y comunicar la incertidumbre, y agregan que la mayoría de las disciplinas científicas formales describen y analizan elementos particulares de la vida y el mundo, desde la naturaleza física del universo hasta el cómo funcionan las sociedades; sin embargo, la estadística es considerada como una metadisciplina que incorpora principios metodológicos que permiten convertir los datos en ideas reales en cualquier contexto.

El objetivo general de la investigación está centrado en diseñar una prueba diagnóstica y obtener evidencias de validez, confiabilidad y equidad; que permita evaluar el razonamiento estadístico en estudiantes universitarios después de haber tomado como mínimo un curso de estadística dentro de su programa de licenciatura. Con base en el objetivo general se desprenden los siguientes objetivos específicos:

- a) Definir el dominio, las dimensiones y contenidos específicos a evaluar del razonamiento estadístico para el diseño de la prueba diagnóstica.
- b) Diseñar reactivos que conformen la prueba diagnóstica para evaluar las dimensiones del razonamiento estadístico.
- c) Obtener evidencias de validez, confiabilidad y equidad en el proceso de diseño de la prueba diagnóstica.

Del objetivo general se desprenden tres preguntas de investigación, la primera está relacionada con el diseño y desarrollo de la prueba, mientras que la segunda y tercera se orientan a la validez y confiabilidad de esta.

1. ¿Qué dimensiones y contenidos específicos del dominio de razonamiento estadístico son más pertinentes evaluar a través de una prueba diagnóstica en estudiantes universitarios?
2. ¿En qué medida la prueba diseñada permite identificar de manera consistente las dimensiones y contenidos específicos del razonamiento estadístico?
3. ¿Qué evidencias de validez, confiabilidad y equidad sustentan la prueba diagnóstica?

Desarrollo

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

1. Los estudiantes 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. Los estudiantes deben ser capaces de reconocer preguntas para las que el proceso de investigación en estadística sería útil y deben poder responder a preguntas utilizando el proceso de investigación.
3. Los estudiantes deben poder producir presentaciones gráficas y resúmenes numéricos e interpretar lo que los gráficos revelan y no revelan.
4. Los estudiantes deben reconocer y ser capaces de explicar el papel central de la variabilidad en el campo de la estadística.
5. Los estudiantes deben reconocer y ser capaces de explicar el papel central de la aleatoriedad al diseñar estudios y sacar conclusiones.
6. Los estudiantes deben adquirir experiencia sobre cómo se utilizan los modelos estadísticos, incluidos los modelos multivariantes.
7. Los estudiantes 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. Los estudiantes deben poder interpretar y sacar conclusiones de los resultados estándar de los paquetes de software estadístico.
9. Los estudiantes deben demostrar conocimiento de las cuestiones éticas asociadas con una práctica estadística sólida. (2016, pp. 8-11).

Por otra parte, la Asociación Americana de Investigación Educativa (American Educational Research Association, AERA), la Asociación Americana de Psicología (American Psychological Association, APA) y el Consejo Nacional de Medición en Educación (National Council on Measurement in Education, NCME) definen

a las pruebas de la siguiente forma “una prueba es un dispositivo o procedimiento en el que se obtiene una muestra del comportamiento de un examinado en un dominio específico y, posteriormente, se evalúa y puntúa mediante un proceso estandarizado” (p.2). También, mencionan que las pruebas de evaluación educativa y psicológica son algunas de las contribuciones más importantes de las ciencias cognitivas y conductuales a nuestra sociedad ya que son fuentes fundamentales e importantes de información sobre individuos y grupos.

Sin embargo, señalan la necesidad de contar con pruebas bien construidas y desarrolladas que son válidas para los propósitos previstos, ya que estas pruebas tienen el potencial de proveer beneficios sustanciales para los examinados y los usuarios. Es así, como los Estándares para las Pruebas Educativas y Psicológicas (AERA, APA Y NCME, 2014) proporcionan criterios para el desarrollo y evaluación de pruebas y prácticas de prueba, y aportan pautas para evaluar la validez de las interpretaciones de los puntajes de las pruebas para los usos previstos de las pruebas.

A través de la revisión de la literatura se ha observado una evolución en el diseño de instrumentos de medición educativa tipo pruebas de rendimiento académico, sin que este deje de ser un proceso ordenado y preciso que se va dando a través de diferentes etapas tomando en consideración el constructo a evaluar. En este sentido, Lane, Raymond, Haladyna y Dowing, (2016) afirman que “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” (p.3). De forma análoga, la AERA, APA y NCME, (2014) a través de los Estándares para Pruebas Educativas y Psicológicas, brindan directrices para el proceso de desarrollo y obtención de pruebas educativas que cuenten con las evidencias de validez, confiabilidad y equidad.

Diversos autores (Backhoff, Peon y Sanchez, 2005; Backhoff, Peon, Sánchez y Andrade, 2006; Backhoff, Solano, Peon, Sanchez y Prieto, 2009; Lane et al, 2016) describen un proceso para el desarrollo de pruebas segregado en diversas fases, etapas o componentes; estas pueden ser de mayor o menor número dependiendo del acercamiento que cada uno de los autores hace. Para el presente estudio se trabajará con la propuesta de Lane et al. (2016) que se muestra en la Tabla 1, en ella se observan las doce fases del proceso para el desarrollo de pruebas y algunas recomendaciones para el correcto desarrollo de cada una de ellas.

Tabla 1. Proceso para el desarrollo de pruebas

Fases para el desarrollo de pruebas	Recomendaciones para el desarrollo de pruebas
1) Plan general	Desarrolle un plan detallado para todo el proyecto de desarrollo de la prueba, que incluya información sobre todos los componentes de la prueba, una justificación para cada componente y los métodos específicos que se utilizarán para evaluar la validez de todas las interpretaciones y usos previstos de la puntuación de la prueba y la calidad psicométrica de la prueba.
2) Definición del dominio y declaración de las afirmaciones	Nombre y defina el dominio a medir. Proporcione una declaración clara de las afirmaciones que se harán sobre los conocimientos, habilidades y habilidades del examinado.
3) Especificaciones de contenido	Desarrolle especificaciones de contenido para guiar el desarrollo de elementos, ensamblaje de formularios, informes de calificaciones y otras actividades.
4) Desarrollo de reactivos/ítems	Identifique los formatos y materiales adecuados para los artículos. Desarrolle elementos para obtener evidencia de validez que respalden su uso.
5) Diseño de prueba y ensamblaje	Diseñe y cree formularios de prueba basados en especificaciones de prueba; atendiendo a los problemas relacionados con el contenido de la prueba, el formato, las reglas de puntuación, la escala y la equiparación.
6) Producción de prueba	Produzca un formulario de prueba claro, preciso y accesible.
7) Administración de la prueba	Administre la prueba de forma estandarizada. Evite las amenazas a la validez que puedan surgir durante la administración.
8) Puntaje	Establezca una política y procedimientos de control de calidad para calificar y tabular las respuestas a los ítems. Garantice una puntuación precisa y coherente cuando se requiera juicio.
9) Puntajes de corte	Establezca puntajes de corte defendibles consistentes con el propósito de la prueba.
10) Informes de calificaciones de la prueba	Desarrolle informes de calificaciones de pruebas accesibles y comprensibles.
11) Prueba de seguridad	Establezca políticas y procedimientos para garantizar la seguridad de las pruebas durante el desarrollo y la administración de estas.
12) Documentación de la prueba	Preparare informes técnicos y otra documentación que respalde la validez, imparcialidad y adecuación técnica de la prueba.

Nota: Traducción y adaptación de la Tabla 1.1 Test Development Process (Lane, et al, 2016, p.4).

Es importante señalar que es necesario seguir correctamente cada una de las doce fases propuestas al momento de diseñar pruebas como instrumentos de evaluación, ya que dicho proceso proporciona un marco para recopilar y organizar evidencia que respalden la calidad psicométrica de la prueba y la validez de las interpretaciones y usos de las puntuaciones de la prueba (Lane, et al., 2016, p.3).

Consideraciones finales

El plan general es la primera fase para el diseño de pruebas, 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 (véase tabla 1). Para Lane, et al. (2016) los elementos que guían el diseño del plan general son los que se definen en la tabla 2, en la cual se agregan los aspectos para el presente estudio.

Tabla 2. Elementos del plan general para el presente estudio

Elemento del plan general	Presente estudio
1. Constructo por medir	Razonamiento estadístico.
2. Población a la que está destinada la prueba	Estudiantes universitarios que hayan 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.	Tomar 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 establecerán 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 o en su defecto se traslada a una aplicación virtual por medio de un software de encuestas (Qualtrics, SurveyToGo, SurveyMonkey o Google Forms).
8. Procedimientos e informe de las puntuaciones.	Informe a los profesores y/o autoridades 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 que los maestros puedan tomar decisión de mejora sobre la enseñanza de contenidos que se relacionan con el razonamiento estadístico.

Nota: Elaboración propia a partir de la revisión Lane et al. (2016, pp. 5-17).

Con base en la tabla 2, se define el propósito de la prueba que consiste en tener un instrumento diagnóstico que cumpla con los criterios y evidencias que proporcionan las pruebas estandarizadas (validez, confiabilidad y equidad) para evaluar el razonamiento estadístico de estudiantes universitarios que han cursado por lo menos una asignatura de estadística en su formación superior y que permita a los profesores o autoridades educativas del área de estadística tomar decisiones para la mejora de los procesos de enseñanza-aprendizaje de los contenidos relacionados con el dominio del razonamiento estadístico.

Como lo señalan Backhoff et al. (2009) “para poder diseñar instrumentos nacionales de calidad se requiere el trabajo colegiado de expertos, mismos que se fundamente en un método sólido y consistente con los propósitos que se persiguen” (p.7). Los comités técnicos son un grupo de expertos especialistas en el campo o dominio a evaluar (razonamiento estadístico), así como en el área de enseñanza del dominio. Los integrantes de los comités técnicos participan realizando tareas en las distintas fases de construcción del instrumento. Para los fines de esta prueba se consideró el trabajo de dos comités técnicos, el comité académico (CA) y el comité de validación y sesgo (CVS).

Las funciones del CA para el diseño de la prueba de la diagnóstico fueron la de verificar si es necesario agregar contenidos adicionales a los mostrados en la tabla de contenidos específicos, así como de evaluar el nivel de importancia del contenido específico a evaluar.

La función del CVS para el diseño de la prueba de la diagnóstico consiste en el jueceo del diseño de los reactivos con la finalidad de revisar su construcción y redacción antes de ir al pilotaje y aplicación de la versión final de la prueba diagnóstico.

Llevar a cabo la segunda fase en el desarrollo de pruebas, implicó definir el dominio del constructo a evaluar en la prueba diagnóstica. En este sentido, es necesario que la tabla de operacionalización del constructo pase por la revisión del comité académico con la finalidad de obtener en consenso de los contenidos y habilidades que se engloban dentro del razonamiento estadístico y como producto se obtenga una tabla de contenidos de la prueba que servirá como insumo en el diseño de las especificaciones de reactivos.

Para elaborar la tabla de contenidos fue necesario el análisis de diferentes elementos, entre ellos, la revisión de autores acerca de la definición del razonamiento estadístico; las directrices que aporta la 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; la revisión de los programas de materia de estadística básica seleccionados por centro Académico que imparte el Departamento de Estadística de la Universidad Autónoma de Aguascalientes (UAA) y finalmente la revisión de programas de materia propios de las carreras que corresponden al área de metodología de la investigación.

La tabla de contenidos se integró de varios elementos, el primero corresponde al razonamiento estadístico que es al dominio a evaluar. El segundo es la conceptualización del razonamiento estadístico, retomando los conceptos que proponen Garfield y Gal, (1999, p. 2), Ben-Zvi y Garfield (2004, p.7) y DelMas, (2004, p.85). El tercer elemento hace referencia a las competencias del razonamiento estadístico, estas surgen tomando como referencia las ocho submetas básicas de Garfield y Gal (1997) propuestas para la instrucción de la estadística en los niveles preuniversitario y universitarios. Adicionalmente se retoman las nueve metas a alcanzar por los estudiantes al finalizar un curso introductorio de estadística en el nivel superior que propone la ASA (2016, p.8) a través del GAISE, estas metas se diseñaron con base en una revisión exhaustiva de la literatura sobre educación estadística y resume lo que un estudiante debe saber y comprender al final de un primer curso de estadística. Finalmente, el cuarto elemento de la tabla corresponde a las dimensiones, temas y subtemas, para ello se tomó como referencia la temáticas a abordar en un primer curso de estadística en el nivel universitario que sugieren Garfield, DelMas y Chance (2003, pp. 6-7).

La tercera fase que corresponde al diseño de las especificaciones de contenido proporciona una definición operativa del contenido de la prueba y las demandas cognitivas de las tareas de evaluación (AERA, APA Y NCME, 2014), actualmente se está trabajando en el diseño de cada una de las especificaciones de contenido.

Referencias

American Statistical Association. (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE), College Report 2016 EUA: ASA.

American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. EUA: American Educational Research Association.

- Backhoff, E., Peon, M. y Sánchez, A. (2009). *Manual Técnico. Diseño de exámenes de la calidad y el logro educativo*. (D. d. INEE, Ed.) CDMX: INEE.
- Backhoff, E., Peon, M., Sánchez, A. y Andrade, E. (2006). *Manual Técnico. Validación de reactivos*. (D. d. INEE, Ed.) CDMX: INEE.
- Backhoff, E., Solano, G., Peon, M., Sánchez, A. y Prieto, L. (2005). *Manual Técnico. Especificaciones de Reactivos*. (D. d. INEE, Ed.) CDMX: INEE.
- Batanero, C., Díaz C., Contreras, J. y Roa, R. (2013). El sentido estadístico y su desarrollo. *Números, Revista de didáctica de las matemáticas*, 83, 7-18. Recuperado de www.sinewton.org › numeros › numeros › Monografico_01
- Ben-Zvi, D. y Garfield, J. (2004). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. En Ben-Zvi, D., Garfield, J. (Eds.), *The challenge of developing statistical, literacy, reasoning and thinking*. Nueva Zelanda: Springer.
- DelMas, R. (2004). A Comparison of Mathematical and Statistical Reasoning. En D. Ben-Zvi, y J. Garfield, *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (p. 79-96). EUA: Springer.
- Garfield, J. y Gal, I. (1997). Curricular Goals and Assessment Challenges in Statistics Education. En J. Garfield, y G. Iddo, *The Assessment Challenge in Statistics Education*. (p. 1-13). EUA: IOS Press.
- Garfield, J. y Gal, I. (1999). Teaching and Assessing Statistical Reasoning. En L. Stiff, y F. Curcio, *Developing Mathematical Reasoning in Grades K-12*. (p. 207-221).
- Garfield, J., DelMas, R. y Chance, B. (2003). The Web-based ARTIST: Assessment Resource Tools for Improving Statistical Thinking. *Assessment of Statistical Reasoning to Enhance Educational Quality* (p. 22). Chicago: AERA.
- Lane, S., Raymond, M., Haladyna, T. y Downing, S. (2016). Test Development Process. En Lane, S., Raymond, M. y Haladyna, T. (Eds.), *Handbook of Test Development*. EUA: Routledge.
- Wild, C., Utss, J. y Horton, N. (2018). What is statistics?. En Ben-Zvi, D., Makar, K., Garfield, J. (Eds.), *International Handbook of Research in Statistics Education*. EUA: Springer.