

ALGUNAS HERRAMIENTAS ESTADÍSTICAS PARA UNA EVALUACIÓN PLURIMETÓDICA

Teresita E. Terán

Universidad Nacional de Rosario

teresitateran@hotmail.com

Campo de investigación: Otros – Evaluación

Argentina

Nivel: Superior

Resumen. *Posicionados en lo que Medina Rivilla y Castillo Arredondo (1998) llaman “perspectiva plurimetódica”, y considerando a la evaluación como un proceso reflexivo, sistemático y riguroso de indagación sobre la realidad, en un contexto regido por principios de validez, fiabilidad, participación y ética, nos planteamos la necesidad de analizar las propiedades y los requisitos métricos de una evaluación y las respuestas de los alumnos. Para ello, proponemos utilizar herramientas estadísticas que permitan un análisis exhaustivo de cómo se ha formulado la evaluación y cuál ha sido el grado de comprensión por parte de los alumnos.*

Los estudios de índices de análisis de ítems y de fiabilidad son útiles para generar instancias de reflexión sobre la práctica docente que permitan rever no sólo el proceso de enseñanza-aprendizaje sino la confección de los instrumentos de evaluación. Además, una encuesta de opinión basada en la metacognición puede ser de ayuda para analizar el control y la conciencia que los alumnos tienen sobre sus procesos cognitivos.

Palabras clave: instrumentos, evaluación, consistencia interna, generalizabilidad

Introducción

Posicionados en lo que Medina Rivilla y Castillo Arredondo (1998) llaman “perspectiva plurimetódica”, Cardona Moltó (1998) considera al enfoque científico de investigación como un procedimiento de búsqueda de conocimiento aplicable a cualquier campo de estudio, por lo que no hay razón para pensar que vaya asociado a determinados campos o disciplinas.

Batanero y Godino (2000) señalan que en los trabajos sobre investigaciones en Ciencias Sociales y Experimentales se ha puesto de manifiesto la existencia de dificultades y errores en la aplicación de los conceptos y procedimientos estadísticos. Acotan además, que se van abandonando las controversias en torno a lo cuantitativo versus lo cualitativo y cada vez con más frecuencia las investigaciones se encuentran en un punto intermedio entre los paradigmas cuantitativo y cualitativo. El enfrentamiento entre ambos paradigmas marcó el desarrollo de la sociología en los años setenta.

Según D’Arcona (1999) se reconoce la pluralidad de vías para acceder a la realidad social, pero no se trata de afirmar un paradigma sobre otro, sino de buscar compatibilidades entre ellos. Reichardt y Cook (1979) sostienen la necesidad de construir puentes entre métodos a partir de la

triangulación en una misma investigación. En el caso de la educación matemática, esta idea de complementariedad de los métodos cuantitativos y cualitativos es sugerida ya, por Kilpatrick (1982) cuando dice que en lugar de abandonar los métodos cuantitativos a favor de los cualitativos deberíamos dirigir nuestros esfuerzos en la dirección de enriquecerlos.

Además, Wittmann (1995) enfatiza que en la Educación Matemática se debe priorizar la investigación sobre el diseño y evaluación de los cuestionarios. Surge, así, la necesidad de construir instrumentos confiables para la evaluación de los alumnos.

Desde esta perspectiva se formulan las siguientes preguntas: ¿Qué criterios seguir para evaluar los cuestionarios en Estadística? ¿Se cuenta con instrumentos de evaluación fiables y válidos en el tema? ¿Cómo analizar el grado de comprensión de los alumnos?

Es por ello, que se propone realizar una investigación sobre métodos mixtos a través del análisis de la concordancia entre el uso de herramientas estadísticas que permitan un análisis exhaustivo de cómo se ha formulado la evaluación y cuál ha sido el grado de comprensión por parte de los alumnos y una encuesta de opinión sobre las dificultades en el proceso de aprendizaje y el control que tienen los alumnos sobre sus procesos cognitivos.

A partir de los interrogantes planteados, se formula el objetivo siguiente:

Objetivo

Analizar algunas herramientas estadísticas para una evaluación plurimetódica.

Métodos de análisis cuantitativos

Desde esta posición, se presenta un detalle de los métodos de análisis cuantitativos que sugerimos emplear al realizar una evaluación continua de los alumnos.

Índice de Dificultad

Se considera de interés analizar qué tipos de ítems resultan de fácil resolución para los alumnos, y cuáles presentan las mayores dificultades. Para ello se aplica un Índice de Dificultad.

Para la elección de este índice se tiene en cuenta la definición de Muñiz (1994) sobre Índice de Dificultad (ID) de un ítem. Muñiz define ID de un ítem, a la proporción de sujetos que lo aciertan (A) de aquellos que han intentado resolverlo (N).

$$\text{Simbólicamente } ID = \frac{A}{N}$$

Como se observa este índice revela mayor dificultad en el ítem cuando es menor el número de respuestas correctas.

Índice de Discriminación

Muñiz (1994) define que un ítem tiene poder discriminativo si distingue, discrimina, entre aquellos sujetos que puntúan alto en el test y los que puntúan bajo, es decir, si discrimina entre los eficaces en el test y los ineficaces. Para Muñiz, el índice de discriminación de un ítem (ρ) en una prueba es el grado en que diferencia a los examinados respecto al carácter que se pretende medir. Se mide mediante el coeficiente de correlación de la puntuación de cada ítem con la puntuación total de la prueba. Un ítem discrimina mejor el nivel de apropiación de los alumnos con respecto a los conceptos relacionados con un tema en evaluación, cuando su correlación con respecto a la puntuación total en la prueba es mayor.

Para ello, se utiliza un coeficiente de correlación biserial-puntual, que es una aplicación de la correlación de Pearson cuando una de las variables es dicotómica y la otra cuantitativa.

$$\text{Simbólicamente: } \rho = \frac{\mu_p - \mu_x}{\sigma_x} \sqrt{\frac{p}{q}}$$

donde:

μ_p : Media en el test de los sujetos que aciertan el ítem.

μ_x : Media del test.

σ_x : Desviación típica del test.

P : Proporción de sujetos que aciertan el ítem.

$q : (1 - p)$

Fiabilidad de la prueba

Muñiz (1994) basándose en la teoría clásica de los tests considera que los errores de medida de los que se ocupa la fiabilidad son aquellos no sometidos a control e inevitables en todo proceso de medir, sea físico, químico o psicológico.

Se llama fiabilidad o consistencia a la extensión por la cual un experimento, test u otro procedimiento de medida produce los mismos resultados en ensayos repetidos. La medida siempre produce un cierto error aleatorio, pero dos medidas del mismo fenómeno sobre un mismo individuo suelen ser consistentes.

La fiabilidad de los tests se estima a través de diversos métodos entre los cuales se destaca el método de consistencia interna. Se mide a través del coeficiente alfa de Cronbach (α)

Este coeficiente refleja el grado en el que covarían los ítems que constituyen el test.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n \sigma_j^2}{\sigma_x^2} \right)$$

Simbólicamente:

donde:

n : Número de ítems del test.

$\sum \sigma_j^2$: Suma de las variancias de los n ítems.

σ_x^2 : Variancia de las puntuaciones en el test.

El rango de variación de este coeficiente es de 0 a 1 (cuanto mayor es su valor, mayor es la fiabilidad del cuestionario).

Si la prueba es homogénea, tiene una alta coherencia interna y mide la misma habilidad en todos sus ítems, pero, si la prueba es heterogénea no se puede esperar un índice de consistencia interna muy alto lo que nos indica que el alfa obtenido es un coeficiente significativo.

Asimismo se ha complementado este estudio a través del cálculo de dos coeficientes de generalizabilidad. La ventaja más clara que presenta la teoría de la generalizabilidad frente a la teoría clásica de la fiabilidad es que permite estimar la fiabilidad de un instrumento de medida en situaciones en las que intervienen múltiples fuentes de error o variabilidad de las puntuaciones.

La teoría de la generalizabilidad tiene en cuenta los múltiples factores que pueden producir variaciones en las puntuaciones de los sujetos y mediante la aplicación de un diseño multivariado y los procedimientos clásicos del Análisis de Variancia (ANOVA), permite estimar la variancia atribuible a cada uno de ellos, así como a sus interacciones.

El coeficiente de generalizabilidad indica el grado en que se pueden generalizar los resultados obtenidos a otras situaciones en que muestras aleatorias de n ítems sean aplicadas en una o más ocasiones aleatorias.

Se calcula, en primer lugar el coeficiente de generalizabilidad (G) que se define como el cociente entre la variancia verdadera en las puntuaciones de la prueba y la variancia observada que es la suma de la variancia verdadera más la variancia debida al error aleatorio. Simbólicamente:

$$G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}$$

donde:

σ_v^2 : Variancia verdadera en las puntuaciones de la prueba.

σ_e^2 : Variancia debida al error aleatorio.

Thorndike (1989) plantea que la variancia del error depende de cómo definimos el universo de puntuaciones verdaderas y en el análisis de la generalizabilidad considera ciertas fuentes como parte de la variancia de error en unas condiciones y otras fuentes en otras.

Se diferencian dos fuentes para el error aleatorio, por lo que se calculan dos coeficientes de generalizabilidad:

Coeficiente de generalizabilidad a otros alumnos de la misma prueba.

Coeficiente de generalizabilidad a otros problemas similares a los incluidos en la prueba a los mismos alumnos.

A través del análisis del modelo de estimación de Dunn y Clarck (1987) y del programa SPSS, se puede calcular para el análisis de variancia de medida repetida las siguientes componentes de la variancia:

Variancia dentro de los sujetos σ_s^2

Variancia dentro de los ítems σ_i^2

Variancia residual σ_e^2

Sustituyendo estos valores en la fórmula y teniendo en cuenta los tamaños de muestra (número de alumnos y número de ítems) se obtienen las siguientes estimaciones según qué fuente de variación se considere.

Coefficiente de generalizabilidad a otros alumnos de la misma prueba.

$$G = \frac{\sigma_i^2}{\sigma_i^2 + \frac{\sigma_e^2}{n}} \quad n: \text{número de alumnos}$$

Este valor obtenido si es alto indica que se pueden generalizar los resultados a otros alumnos conservando el mismo cuestionario de evaluación, suponiendo condiciones uniformes en el tipo de alumno y en el tipo de enseñanza impartida.

Coefficiente de generalizabilidad a otros problemas similares a los incluidos en la prueba a los mismos alumnos.

$$G = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_e^2}{n}} \quad n: \text{número de ítems de la prueba}$$

Se observa que el valor de este coeficiente es similar al coeficiente α de Cronbach, ya que estos dos coeficientes deben acercarse puesto que se considera como fuente de variación la de los problemas, y los alumnos fijos.

Métodos de análisis cualitativos

La metacognición como estrategia didáctica: La encuesta de opinión

Para poder realizar una interpretación más exacta de la relación entre la opinión de los alumnos sobre dificultades en el aprendizaje de un tema y el resultado de una evaluación escrita, se incorporan algunas reflexiones teóricas sobre la metacognición.

Cuando se habla de metacognición se refiere a la conciencia y el control que los individuos tienen sobre sus procesos cognitivos. Garner (1987) sostiene que, durante la última década, una considerable cantidad de estudios han demostrado que la metacognición desempeña un papel importante en la efectiva comprensión.

El término metacognición de acuerdo a la mayoría de los autores alude a dos componentes básicos, el saber acerca de la cognición y la regulación de la cognición. El primer componente se refiere a la capacidad de reflexionar sobre nuestros propios procesos cognitivos, y la regulación metacognitiva implica el uso de estrategias que nos permiten controlar esfuerzos cognitivos.

Fischer y Lipson (1986) expresan que uno de los objetivos de la enseñanza universitaria de la ciencia es que los estudiantes aprendan a reconocer y corregir sus propios errores. Sostienen que los propios estudiantes deben adquirir la habilidad de manejarse frente al error y ser capaces de desmontar sus propios “programas”.

En base a lo señalado, se observa que el reconocimiento y la corrección de los propios errores son operaciones metacognitivas fundamentales.

El problema del preconcepto erróneo es de orden metacognitivo. Si los estudiantes no toman conciencia de que no poseen el conocimiento correcto, no pueden clarificar su comprensión, a ese fin la encuesta de opinión constituye una estrategia para esa toma de conciencia. Es por ello, que se propone la realización de una encuesta cuyo objetivo es conocer la opinión de los alumnos sobre las dificultades en el proceso de aprendizaje del tema y a través de esa opinión conocer la conciencia y el control que tiene el alumno sobre sus procesos cognitivos.

Para confeccionar el protocolo de la encuesta de opinión de los alumnos se debe tener en cuenta no sólo la propia experiencia docente sobre el tema sino estudios realizados por especialistas sobre las dificultades más observadas.

Luego de seleccionadas estas dificultades se agrupan en variables didácticas. En general, se mencionan: comprensión de la teoría, comprensión de la simbología, comprensión de las consignas, planteo del problema, cálculo numérico, especificación de las soluciones utilizando la simbología y las gráficas utilizadas (si las hubiere), interpretación de los resultados.

Una vez realizada la encuesta se debe analizar la proporción de alumnos que dicen no tener problemas en la evaluación de los distintos ítems que comprenden las dificultades más comunes observadas y se debe comparar con las calificaciones obtenidas en cada uno de los ítems en la evaluación escrita, a través de tablas de contingencia. Las conclusiones se deben reforzar mediante los coeficientes de asociación y a través de las pruebas donde se analiza la coherencia entre la opinión de los alumnos sobre las dificultades que se le presentan en el tema considerado y los resultados obtenidos por ellos en la evaluación realizada.

Conclusión

Investigar las herramientas estadísticas para una evaluación plurimetódica permite al docente una investigación- acción sobre su propia práctica docente, generando instancias de reflexión sobre su accionar, complementando el desarrollo de los contenidos estadísticos con la evaluación continua del proceso de enseñanza y aprendizaje y de las herramientas utilizadas para que dicha evaluación sea confiable y válida. La evaluación de los instrumentos es para Wittmann (1995) un proceso de interacción activa entre diferentes áreas y disciplinas relacionadas como lo son la Matemática, la Didáctica, la Pedagogía, la Psicología, entre otras.

Este trabajo es una contribución al mejoramiento de la calidad educativa que propone el estudio exhaustivo de herramientas estadísticas que redundarán en el análisis de los ítems que permitan la confección de un cuestionario de evaluación válido y confiable, que tenga en cuenta todos los factores que pueden afectar el conocer el grado de comprensión de los alumnos sobre los temas propuestos en la evaluación y su triangulación con una encuesta de opinión donde el alumno utilice la metacognición como una estrategia didáctica de su propio grado de comprensión y aprendizaje del tema planteado. .

Referencias bibliográficas

- D'Arcona, M.A. (1999) *Metodología cuantitativa: Estrategias y técnicas de investigación social*. Madrid: Síntesis.
- Batanero, C. y Godino, J. D. (2000). *Análisis de datos y su didáctica*. Granada: Departamento de Didáctica de la Matemática.
- Cardona Moltó, C. (1998). *Pedagogía diferencial. Educación especial. Proyecto docente y de investigación inéditos*. Alicante: Dpto. Psicología de la Salud, Universidad de Alicante.
- Dunn, O. J. y Clarck, V. A. (1987). *Applied statistics: Analysis of variance and regression*. New York: Wiley.
- Fischer, K. M. y Lipson, J.K. (1986). Twenty questions about student errors. *Journal of Research in Science Teaching* 23, 783-803.
- Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.
- Kilpatrick, J. (1982). Research on mathematical learning and thinking in the United States". *Recherches en Didactique des Mathématiques* 2(3), 393-379.
- Medina Rivilla, A. y Castillo Arredondo, S. (1998). *Evaluación de los Procesos y resultados del aprendizaje de los estudiantes*. Madrid: UNED.
- Muñiz, J. (1994). *Teoría clásica de los tests*. Madrid: Pirámide.
- Reichardt y Cook (1979). *Qualitative and Quantitative Methods in Social Sciences*. Beverly Hills: Sage.
- Thorndike, R. L. (1989). *Psicometría aplicada*. México: Limusa.
- Wittmann, R. L (1995). Mathematics Education as a Design Science. *Educational Studies in Mathematics*. 29, 355-374