

Confiabilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima

Reliability and power of discrimination of the pedagogical tests of maximum execution

Aníbal C. Ayala ⁽¹⁾

⁽¹⁾ Facultad de Educación, Universidad Nacional del Centro del Perú

Email: Anibalito.c@gmail.com

Resumen

En el presente trabajo de investigación, se analizó estadísticamente e interpretó la relación entre los estadígrafos psicométricos de confiabilidad y poder de discriminación de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.

Tuvo como objetivos específicos, calcular los estadígrafos psicométricos de confiabilidad y poder de discriminación; los coeficientes de correlación y de regresión, entre la confiabilidad y el poder de discriminación de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.

Verificando la hipótesis: existe una relación entre los estadígrafos psicométricos, confiabilidad y poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín. La investigación es del tipo tecnológico-aplicada con un diseño descriptivo correlacional.

Se llevó a cabo en la Universidad Nacional del Centro del Perú, Facultad de Educación, a lo largo de un año calendario comprendido entre el 2014-07-01 y el 2015-06-30.

Arribó a la conclusión que, hasta el 57 % del valor de la confiabilidad está determinado por el poder de discriminación, el 43% restante se debe a otros factores; y que el grado de dependencia del poder de discriminación con respecto de la confiabilidad es mayor que el grado de influencia de la confiabilidad en el poder de discriminación; de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín

Palabras clave: edumetría. estadística aplicada a la investigación educativa. psicometría

Abstract

It was statistically analyzed and interpreted the relationship between psychometric statistics of reliability and discrimination power of pedagogical testing items applied to educational institutions students of Junin Region.

It had specific objectives; calculate psychometric statistics of reliability and discrimination power; correlation and regression coefficients between reliability and discrimination power of pedagogical testing items applied to educational institutions students of Junin Region.

Verifying the hypothesis: there is a relationship between psychometric statistics, reliability and discrimination power, pedagogical testing items applied to educational institutions students of Junin Region.

The research is technology-applied type with a descriptive correlational design.

It was held at the National University of Central Peru, Education Faculty, over a year from 2014-07-01 until 2015-06-30.

It came to the conclusion, up to 57% of reliability value is determined by discrimination power, the remaining 43% is due to other factors; and dependence degree of discrimination power respect of reliability is greater than influence degree of reliability in discrimination power; of pedagogical testing applied to educational institutions students of Junín Region.

Keywords: edumetric. statistic applied to the educational investigation. psychometric

Introducción

Entre los años 2004 al 2013, se realizaron investigaciones centradas en el cálculo de los estadígrafos psicométricos y edumétricos, tales como, la validez, confiabilidad, dificultad, sensibilidad e índice de discriminación de los diferentes tipos de instrumentos de recolección de datos; y quedaron como sugerencias reiterativas, averiguar la relación entre los estadígrafos psicométricos de confiabilidad y poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.

Por otro lado, el contenido de la asignatura de Técnicas del Procesamiento de Datos e Información, de los Planes de Estudio de la Unidad de Post Grado de la Facultad de Educación, implica, entre otros temas, el cálculo de los estadígrafos psicométricos de los instrumentos de recolección de datos: confiabilidad y poder de discriminación; quedando pendiente siempre la relación que existe entre la confiabilidad y el poder de discriminación; es decir, existe primacía de una de ellas sobre la otra, una es causa o efecto de la otra, por qué hay que calcular siempre los dos o es necesaria solo uno de ellas.

Entonces, para poder subsanar, en parte, el inconveniente citado líneas arriba, en el presente trabajo de investigación se explica, que existe una correlación de Pearson moderada, significativa e inversa y el coeficiente de determinación que dice que hasta el 57 % del valor de la confiabilidad está determinado por el poder de discriminación y que el grado de dependencia del poder de discriminación con respecto de la confiabilidad es mayor que el grado de influencia de la confiabilidad en el poder de discriminación.

Metodología de la investigación

Población y muestra de la investigación

POBLACIÓN

Se tomó como población de la investigación a todas las pruebas pedagógicas objetivas de ejecución máxima que aplicaron a sus alumnos, en las diferentes instituciones educativas del nivel secundario de la Región Junín, los estudiantes de la Maestría en Educación, hasta el año académico 2014, así como, los egresados del Pregrado de la Facultad de Educación hasta el año 2014; las mismas que están consignadas en los anexos de sus respectivas tesis de grado.

MUESTRA

Para la elección de la muestra se utilizó un diseño muestral no aleatorio dirigido, de tal manera que, la muestra estuvo conformada por seis pruebas pedagógicas de ejecución máxima (dos policotómicas y cuatro dicotómicas).

Nota: Las pruebas pedagógicas objetivas policotómicas de ejecución máxima confiables y con poder de discriminación aceptable son algo escasas

Materiales y métodos

TIPO DE INVESTIGACIÓN

Dentro del campo de la investigación educativa y de acuerdo a las características de la hipótesis y los objetivos se enmarcó dentro del tipo de investigación tecnológico-aplicada.

MÉTODO BÁSICO DE LA INVESTIGACIÓN

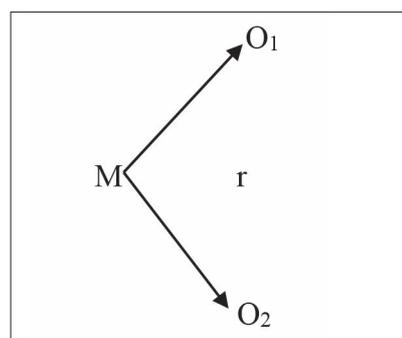
El método de investigación fue el descriptivo.

DISEÑO DE LA INVESTIGACIÓN

El diseño a emplearse en esta investigación fue el descriptivo, porque resultó imposible el control experimental riguroso. Además, este diseño se adecua perfectamente a la presente investigación.

Dentro de los diseños descriptivos se encuentra el diseño descriptivo correlacional, es el que se aplicó para determinar la relación entre los estadígrafos psicométricos, confiabilidad y validez, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.

En el caso concreto de la presente investigación, se aplicaron las pruebas pedagógicas, luego con los datos recolectados, para cada ítem, se calcularon los estadígrafos psicométricos de confiabilidad y validez, los que fueron analizados. Sánchez (2006:105). El diseño puede diagramarse de la siguiente manera:



dónde:

M: Pruebas pedagógicas que se aplican a los alumnos de las instituciones educativas de la Región Junín.

O1, O2: Estadígrafos psicométricos, de confiabilidad y poder de discriminación, respectivamente.

r: Coeficiente de correlación.

VARIABLES DE LA INVESTIGACIÓN

VARIABLE 01: (V-01)

Ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de Región Junín.

VARIABLE 02: (V-02)

Estadígrafos psicométricos, confiabilidad y poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de Región Junín.

INSTRUMENTOS DE RECOPIACIÓN DE DATOS

Las seis pruebas pedagógicas que se seleccionaron como muestra fueron las que aplicaron a sus alumnos las Profesoras:

- Pianto /Tinoco, (inédito); en la I.E. "La Victoria" – El Tambo; en el Área de C.T.A. 4º grado de secundaria. Durante el III bimestre del Año Académico Escolar 2006. Prueba policotómica N°01.
- Gamero, (inédito); en la I.E. Mariscal Castilla – El Tambo; en la asignatura de Matemática, Nivel Secundario, Tercer Grado. Durante el II bimestre del Año Académico Escolar 2010. Prueba policotómica N°02.
- Ilizarbe, (inédito); en la I.E. Fe y Alegría – Chilca; en la asignatura Habilidades Matemáticas, Nivel Inicial, 4 años. Durante el III bimestre del Año Académico Escolar 2009. Prueba dicotómica N°01.
- Lavado, (inédito); en la I.E. Ricardo Menéndez – El Tambo; en la asignatura de Comunicación, Nivel Primario, Tercer Grado. Durante el III bimestre del Año Académico Escolar 2009. Prueba dicotómica N°02.
- Meza, (inédito); en la I.E. Carl F. Gauss – Huancaayo; en la asignatura de Comunicación, Nivel Primario, Quinto Grado. Durante el III bimestre del Año Académico Escolar 2009. Prueba dicotómica N°03.
- Serrano, (inédito); en la I. E. Rafael Gastélua – Satipo; en el área de lógico matemático, Nivel Inicial, 5 años. Durante el IV bimestre del Año Académico Escolar 2008. Prueba dicotómica N°04.

Se consideraron a estos instrumentos de recolección de datos, debido a que los resultados de los estadígrafos psicométricos calculados –confiabilidad y poder discriminativo– fueron óptimos para los objetivos del presente estudio.

PROCEDIMIENTOS DE RECOPIACIÓN DE DATOS

Los docentes –mencionados líneas arriba– fueron capacitados (oportunamente) en el diseño, elaboración y aplicación de pruebas. Así como, en el análisis estadístico y el cálculo de los estadígrafos psicométricos de confiabilidad y poder discriminativo. Los seis instrumentos de recolección de datos, fueron aplicados por los docentes a sus estudiantes, en sus respectivas instituciones educativas

TÉCNICAS Y PROCEDIMIENTOS DE PROCESAMIENTO Y ANÁLISIS DE DATOS

Siendo la investigación de carácter cuantitativo, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics 20, con el menú correspondiente: escala–análisis de la fiabilidad. El cálculo de los estadígrafos se realizó para cada uno de los ítems de las seis pruebas. Los datos, pueden observarse en los Anexos respectivos.

- El cálculo de la confiabilidad (lo realiza el computador con el paquete mencionado líneas arriba) pero está basado en la fórmula: Alfa de Cronbach. Los resultados de los coeficientes de confiabilidad de cada uno de los ítems las pruebas, fueron mayores o iguales que 0,60; por lo que se concluyó que las seis pruebas son confiables, Shiefelbein (1999:296).

- El cálculo del poder de discriminación (se realizó con el computador y el paquete mencionado líneas arriba) pero además se aplicó la fórmula: de correlación punto biserial. Todos los coeficientes de poder de discriminación, de cada uno de los ítems de las pruebas, fueron mayores o iguales que 0,30, Santibáñez (2001:187); por lo que los ítems de las seis pruebas tienen poder discriminativo aceptable o muy buena.

Los coeficientes de confiabilidad y poder de discriminación fueron sometidos a un análisis de correlación y regresión; y, finalmente interpretados y discutidos.

Resultados y discusión**CUADRO RESUMEN DE CORRELACIONES DE PEARSON ENTRE LOS ESTADÍGRAFOS PSICOMÉTRICOS DE PODER DE DISCRIMINACIÓN Y CONFIABILIDAD**

PRUEBA PEDAGÓGICA	Correlación de Pearson r	Coefficiente de determinación r ²	Variabilidad explicada r ² (%)
POLICOTÓMICA N°01	0,123	0,020	2%
POLICOTÓMICA N°02	-0,929	0,863	86%
DICOTÓMICA N°01	-0,985	0,970	97%
DICOTÓMICA N°02	-0,224	0,050	5%
DICOTÓMICA N°03	-0,872	0,760	76%
DICOTÓMICA N°04	-0,874	0,764	76%
PROMEDIO	-0,627	0,566	57%

Para las seis pruebas pedagógicas (dos policotómicas y cuatro dicotómicas), se tiene una correlación de Pearson promedio $r = -0.627$; correlación moderada, significativa a un nivel de 0.01, por ser negativa, existe una correlación inversa entre los estadígrafos psicométricos de poder de discriminación y confiabilidad; es decir, a mayor valor del poder de discriminación menor valor de la confiabilidad de la pruebas. Por otro lado, el coeficiente de determinación promedio es $r^2 = 0.566$; lo que significa –también en promedio mediante la variabilidad explicada– que el 57% del valor del poder de discrimi-

nación de las pruebas está determinado por el valor de la confiabilidad, y el 43% se debe a otros factores.

CUADRO RESUMEN DE REGRESIONES LINEALES EN-

PRUEBA PEDAGÓGICA	Grado de dependencia del poder de discriminación con respecto de la confiabilidad	Grado de influencia de la confiabilidad en el poder de discriminación
	b	k
POLICOTÓMICA N°01	0,276	0,055
POLICOTÓMICA N°02	-4,999	-0,173
DICOTÓMICA N°01	-14,141	-0,069
DICOTÓMICA N°02	-1,308	-0,038
DICOTÓMICA N°03	-6,375	-0,119
DICOTÓMICA N°04	-13,306	-0,057
PROMEDIO	-6,642	-0,067

Para las seis pruebas pedagógicas (dos policotómicas y cuatro dicotómicas), se tienen mayores grados de dependencia, tanto directos como inversos (positivos o negativos), del poder de discriminación con respecto de la confiabilidad; y menores grados de influencia, tanto directos como inversos (positivos o negativos), de la confiabilidad en el poder de discriminación.

DOCIMASIA DE HIPÓTESIS "t" DE STUDENT DE LA CORRELACIÓN PROMEDIO

- Hipótesis nula (H_0): La correlación promedio entre los estadígrafos psicométricos de poder de discriminación y confiabilidad de las pruebas pedagógicas de ejecución máxima no e

$$H_0: \rho = 0$$

- Hipótesis alterna (H_1): La correlación promedio entre los estadígrafos psicométricos de poder de discriminación y confiabilidad de las pruebas pedagógicas de ejecución má

$$H_1: \rho \neq 0$$

- Distribución de la muestra: la distribución de la muestra es la distribución t de student con grados de libertad $gl = N - 2$; teniendo en cuenta que son 6 pruebas pedagógicas consideradas en el espacio muestral, será:

$$gl = 6 - 2 = 4$$

- Prueba estadística: Se elige la prueba t de student bilateral, en función de la r de Pearson promedio para las seis instrumentos $r = -0,627$.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{N-2}}}$$

$$t = \frac{-0,627}{\sqrt{\frac{1-0,566}{4}}} = \frac{-0,627}{0,3294} = -2,9035$$

- Región crítica: t tabulada = +/- 2,776. Puesto que H_1 es bilateral, la región crítica abarca todos los valores de $t \geq +2,776$ o $t \leq -2,776$.



- Decisión: Puesto que la t calculada está ubicada en la región de rechazo de la H_0 ; es decir, la t calculada $> t$ tabulada, se acepta la H_1 .

CONCLUSIÓN ESTADÍSTICA

La correlación promedio entre los estadígrafos psicométricos de poder de discriminación y confiabilidad de las pruebas pedagógicas de ejecución máxima es significativa.

Discusión

Para responder a la interrogante de la investigación: ¿Qué relación existe entre los estadígrafos psicométricos de confiabilidad y poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín?; en primer lugar hay que poner de manifiesto que Schiefelbein (1999) pone de manifiesto que, la confiabilidad como estadígrafo psicométrico debe tener un valor superior o igual a 0,60 tanto para el ítem como para todo el instrumento de recolección de datos; de igual forma Santibáñez (2001) manifiesta que el poder de discriminación debe tener un valor superior o igual a 0,30 para cada uno de los ítems que conforman el instrumento de recolección de datos.

En la presente investigación, se han observado estas consideraciones teóricas y los valores calculados de la confiabilidad y el poder de discriminación de los ítems de las pruebas pedagógicas de ejecución máxima conformantes del espacio muestral; están enmarcados

entre los mínimos requeridos.

El valor del coeficiente de correlación de Pearson promedio $r = -0,627$, entre los estadígrafos psicométricos de confiabilidad y poder de discriminación es alto y significativo a un nivel de 0.01, que por ser negativo se dice que existe una relación inversa entre los estadígrafos; es decir, a mayor valor de la confiabilidad menor valor del poder de discriminación, de las pruebas y viceversa. Esta forma de interpretar, es estrictamente matemática, y no estaría revelando relaciones de causa-efecto, como demarca el espíritu de la investigación. Que existe relación entre la confiabilidad y el poder de discriminación, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es $r^2 = 0,566$; lo que quiere decir es que, en promedio el 57 % del valor de la confiabilidad de las pruebas está determinado por el valor del poder de discriminación, y el 43% se debe a otros factores. Por cierto que para emitir cualquier juicio valorativo sobre lo expuesto, habría que analizar no solo seis pruebas pedagógicas, como es el caso de la presente investigación, sino muchas (muchísimas más); no perder de vista que a la presente investigación es exploratoria y sería un punto de partida para otras investigaciones que se pudieran realizar.

Observando que los resultados de los coeficientes de correlación de Pearson —a excepción de la prueba policotómica N°01— son negativos y altos; se puede aseverar que a mayor confiabilidad menor poder de discriminación, y viceversa; pero esta aseveración si se la generaliza de manera contundente para todas las pruebas pedagógicas de ejecución máxima que pudieran haber; y, por cierto que las hay; sería riesgosa y temeraria, porque la muestra para la presente investigación no es tan vasta; es más, no han sido tomada de manera totalmente aleatoria, sino intencional.

Por otro lado, observando los resultados de los coeficientes de regresión lineal, para las seis pruebas pedagógicas (dos policotómicas y cuatro dicotómicas); en promedio, el grado de dependencia del poder de discriminación con respecto de la confiabilidad es mayor que el grado de influencia de la confiabilidad en el poder de discriminación.

Todo lo expresado líneas arriba, se escapa de los alcances de una buena práctica pedagógica, porque ésta tiene otras características, y los estadígrafos psicométricos investigados no son los únicos, sino que existen otros que también influyen en la elaboración de una prueba pedagógica; tal como lo plantea Santibáñez (2001), quién pone de manifiesto que también se deben tener en cuenta a los estadígrafos edumétricos.

Conclusiones

1.- El estadígrafo psicométrico de confiabilidad de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, fluctuaron entre los valores 0,600

y 0,842; en términos cualitativos fueron todos confiables. 2.- El estadígrafo psicométrico de poder de discriminación de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, fluctuaron entre los valores 0,30 y 0,775; en términos cualitativos fueron todos aceptables o muy buenos.

3.- La correlación es moderada, significativa e inversa y el coeficiente de determinación pone de manifiesto que hasta el 57 % del valor de la confiabilidad de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín está determinado por el poder de discriminación, el 43% restante se debe a otros factores.

4.- El análisis de regresión lineal, de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, pone de manifiesto que el grado de dependencia del poder de discriminación con respecto de la confiabilidad es mayor que el grado de influencia de la confiabilidad en el poder de discriminación.

Referencias

- Ary; Jacobs y Razavieh. (1992). Introducción a la investigación pedagógica. México D.F., México: McGraw-Hill.
- Cárdenas Ayala, A. (2007). Pertinencia de los Instrumentos de recolección de datos con los métodos estadísticos de confiabilidad. (I.I. FPH-UNCP, trabajo de investigación inédito). Huancayo, Perú.
- (2008). Análisis estadístico de los instrumentos de recolección de datos a través de los estadígrafos de deformación y apuntamiento. (I.I. FPH-UNCP, trabajo de investigación inédito). Huancayo, Perú.
- (2009). Poder discriminativo, nivel de exigencia y capacidad predictiva de los exámenes de selección, I.I.FPH-UNCP, trabajo de investigación inédito, Huancayo, Perú.
- (2010). Relación entre la validez, el poder discriminativo y nivel de dificultad de las pruebas pedagógicas. Horizonte de la ciencia, 01, 109-119. Huancayo, Perú.
- (2011). Índice de discriminación de los ítems policotómicos en las pruebas pedagógicas. Prospectiva universitaria, V- 8, 01, 119-126. Huancayo, Perú.
- (2012). Análisis estadístico del índice de dificultad de los ítems policotómicos en las prue-

- bas pedagógicas. *Horizonte de la ciencia*, 02, 74-81. Huancayo, Perú.
- (2013). Índice de dificultad y asimetría de los ítems en las pruebas pedagógicas. *Horizonte de la ciencia*, 05, 65-74. Huancayo, Perú.
- Delgado; Escurra y Torres. (2006). *La medición en psicología y educación*. Lima, Perú: Hozlo S.R.L.
- Hambleton y Novik. (1989). *Toward an integration of theory and method for criterion referenced test*. New York, USA: Measurement.
- Nunnally y Bernstein. (1995). *Teoría psicométrica*. México: Mc Graw-Hill.
- Quezada, N. (2010). *Metodología de la investigación-Estadística aplicada en la investigación*. Lima, Perú: MACRO.
- Sánchez y Reyes. (2006). *Metodología y Diseños en la Investigación Científica*. Lima, Perú: Visión Universitaria.
- Schiefelbein, E. (1999). *Teoría, técnicas, procesos y casos en el planeamiento de la educación*. Buenos Aires, Argentina: Ateneo.
- Santibáñez, J. (2001). *Manual para la evaluación del aprendizaje estudiantil (conceptos, procedimientos, análisis e interpretación para el proceso evaluativo)*. México:Trillas.
- Tavella, N. (1991). *Análisis de los ítems en la construcción de instrumentosPsicométricos*. México:Trillas.
- Tembrick, T. (1981). *Evaluación*. Madrid, España:Narcea.
- Thorndikee y Hagen. (1991). *Medición y evaluación en psicología y educación*. México: Trillas.
- Visauta y Martori i Cañas. (2003). *Análisis estadístico con SPSS para Windows, (Volumen II)*. Madrid, España:McGraw-Hill.