



## Validez y sensibilidad de pruebas pedagógicas de ejecución máxima

Edgar Aníbal Cárdenas Ayala\*

### Resumen

En BANNER CORONAL REAL 50x150 cm.pdf el presente trabajo de investigación, se analizó estadísticamente e interpretó la relación entre los estadígrafos psicométricos de validez y sensibilidad 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 validez y sensibilidad; los coeficientes de correlación y de regresión, entre la validez y sensibilidad de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín. Y, como hipótesis: existe una relación entre los estadígrafos psicométricos de validez y sensibilidad, 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 arribó a la conclusión que solo el 3,47% de la validez está determinado por el valor de la sensibilidad y el 96,53% se debe a otros factores; y que se tienen mayores grados de dependencia de la sensibilidad con respecto de la validez y menores grados de influencia de la validez en la sensibilidad 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 e investigación educativa, Psicometría.

## Validity and sensitivity of pedagogical testing

### Abstrac

It was statistically analyzed and interpreted the relationship between psychometric statisticians of validity and sensitivity of pedagogical testing applied students in educational institutions of Junín Region.

It had specific objectives, calculate psychometric statisticians of validity and sensitivity; correlation and regression coefficients, between the validity and sensitivity of pedagogical testing applied students in educational institutions of Junín Region. It had like hypothesis: there is a relationship between psychometric statisticians of validity and sensitivity of pedagogical testing applied students in educational institutions of Junín Region.

### Keywords:

Edumetric, Statistics and educational investigation, Psychometric

Recibido: 30-09-2021

Aceptado: 30-11-2021

\*Filiación: Universidad Nacional del Centro del Perú

Datos del autor:

Edgar Aníbal Cárdenas Ayala. Investigador y docente de Ciencias Naturales, Métodos Cuantitativos para la Investigación Científica e Idiomas (inglés y portugués). Doctor en Ciencias de la Educación. Correo: anibalito.c@gmail.com, ecardenas@uncp.edu.pe ORCID: <https://orcid.org/0000-0002-8888-2401>

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

It came to the conclusion that only 3.47% of validity is determined by the value of sensitivity and 96,53% is due to other factors; and that the degree of dependence on the sensitivity regarding the validity of the tests is greater than the degree of influence of validity in the sensitivity of pedagogical testing applied students in educational institutions of Junín Region.

## **Introducción**

En estos últimos años, se realizaron investigaciones centradas en averiguar algunas de las relaciones existentes entre los estadígrafos psicométricos de confiabilidad, poder de discriminación, nivel de dificultad, sensibilidad 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; quedando por hacer, el estudio de muchas otras relaciones que pudieran existir entre los estadígrafos psicométricos ya mencionados.

Por otro lado, el desarrollo de la asignatura de Análisis cuantitativo y cualitativo de datos, en 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: validez y sensibilidad; quedando en el tintero la relación que existe entre estos dos; es decir, existe primacía de una de ellos sobre el otro, uno es causa o efecto del otro, por qué hay que calcular siempre los dos o es necesario solo uno de ellos.

Entonces, para poder subsanar, en parte, el inconveniente citado en el párrafo anterior, en el presente trabajo de investigación se determinó que, existe una correlación baja, directa y con poca significatividad; así también, el coeficiente de determinación pone de manifiesto que sólo el 3,47% de la validez está determinada por el valor de la sensibilidad y el 96,53% se debe a otros factores; y, que el grado de dependencia de la sensibilidad con respecto de la validez es mayor que el grado de influencia de la validez en la sensibilidad; de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.

## **Método**

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. El método de investigación fue el descriptivo.

El diseño que se empleó en esta investigación fue el descriptivo, porque resulta 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, validez y sensibilidad, 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 validez y sensibilidad, los que fueron analizados. Sánchez (2006:105).

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 primario, secundario y tecnológico de la Región Junín, los estudiantes de la Maestría en Educación, en la Asignatura de Análisis de Datos Cuantitativos y Cualitativos; las mismas que se consignan en los anexos del presente trabajo de investigación.

Para la elección de la muestra se utilizó un diseño muestral no aleatorio dirigido, de tal manera que, la muestra estuvo conformada por diez pruebas pedagógicas de ejecución máxima (todas dicotómicas),

que aplicaron --los estudiantes de la Maestría en Educación-- a sus alumnos en las instituciones educativas de la Región Junín en los dos últimos Años Académicos.

Nota: Las pruebas pedagógicas objetivas policotómicas de ejecución máxima para el desarrollo de la presente investigación fueron algo escasas.

Variable 01: (V-01)

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

Variable 02: (V-02)

Estadígrafos psicométricos, validez y sensibilidad, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de Región Junín.

Procedimientos de recopilación de datos

Los docentes --mencionadas 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 validez y sensibilidad. Los diez instrumentos de recolección de datos, fueron aplicados por los docentes a sus estudiantes, en sus respectivas instituciones educativas.

Instrumentos de recopilación de datos

Las diez pruebas pedagógicas que se seleccionaron como muestra fueron las que aplicaron a sus estudiantes los docentes capacitados para tal fin.

Se consideraron a estos instrumentos de recolección de datos, debido a que los resultados de los estadígrafos psicométricos calculados --validez y sensibilidad-- fueron óptimos para los objetivos del presente estudio.

Siendo la investigación de carácter cuantitativo, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics 20, con los menús correspondientes, --escala: análisis de la fiabilidad y estadísticos descriptivos: kurtosis--. El cálculo de los estadígrafos se realizó para cada uno de los ítems y para el puntaje total obtenido en cada una de las seis pruebas. Los datos, pueden observarse en los Anexos respectivos.

El cálculo de la validez (lo realizó el computador con el paquete mencionado líneas arriba) el que está basado en la fórmula: correlación ítem-total o correlación ítem-test. Los resultados de los coeficientes de validez de cada uno de los ítems las pruebas, fueron mayores o iguales que 0,2000; por lo que se concluyó que las diez pruebas son confiables, Shiefelbein (1999:296).

El cálculo de la sensibilidad (también lo realizó el computador con el paquete mencionado líneas arriba) mediante el estadígrafo kurtosis. Para el análisis de los valores de los coeficientes de sensibilidad de cada uno de los ítems de las pruebas, así como del puntaje total, se tuvo en cuenta el rango de valores: iguales a 0,263 --indican sensibilidad intermedia, mesokurtosis--; mayores que 0,263 indican --mucha sensibilidad--; menores que 0,263 indican --no mucha sensibilidad--; Shiefelbein (1999:286).

## Resultados

Tabla1. Resumen de correlaciones de pearson entre los estadígrafos psicométricos de validez y sensibilidad

Prueba Pedagógica	Correlación de Pearson R	Coefficiente de determinación r <sup>2</sup>	Variabilidad explicada r <sup>2</sup> (%)
Nº01	-0,005	0,0000	0,00
Nº02	-0,324	0,1049	10,49
Nº03	-0,141	0,0199	2,00
Nº04	-0,064	0,0041	0,41
Nº05	-0,346	0,1197	11,97
Nº06	0,085	0,0072	0,72
Nº07	-0,181	0,0327	3,27
Nº08	0,132	0,0174	1,74
Nº09	-0,071	0,0050	0,50
Nº10	-0,190	0,0361	3,61
Promedio	-0,111	0,0347	3,47

Para las diez pruebas pedagógicas (dicotómicas), se tiene una correlación de Pearson promedio  $r = -0,111$ ; correlación muy baja o débil, con significatividad bilateral promedio de 0,533, por ser negativa, existe una correlación inversa entre los estadígrafos psicométricos de sensibilidad y validez; es decir, a mayor valor de la sensibilidad menor valor de la validez de la prueba. Por otro lado, el coeficiente de determinación promedio es  $r^2 = 0,0347$ ; lo que se interpreta --mediante el promedio de la variabilidad explicada-- que hasta el 3,47% del valor de la sensibilidad de las pruebas está determinado por el valor de la validez, y el 96,53% se debe a otros factores.

Tabla 2. Resumen de regresiones lineales entre los estadígrafos psicométricos de validez y sensibilidad

Prueba Pedagógica	Grado de dependencia de la sensibilidad con respecto de la validez B	Grado de influencia de la validez en la sensibilidad k
Nº01	-0,0708	0,000
Nº02	-5,009	-0,021
Nº03	-1,115	-0,018
Nº04	-2,244	-0,002
Nº05	-4,101	-0,029
Nº06	1,605	0,004
Nº07	-0,792	-0,041
Nº08	6,103	0,003
Nº09	-0,843	-0,006
Nº10	-15,323	-0,002

Para las diez pruebas pedagógicas (dicotómicas), se tienen mayores grados de dependencia, de la sensibilidad con respecto de la validez de las pruebas; y menores grados de influencia, de la validez en la sensibilidad de las respectivas pruebas.

Docimasia de hipótesis de la correlación promedio

Hipótesis nula (H<sub>0</sub>): La correlación entre los estadígrafos psicométricos de validez y sensibilidad de las pruebas pedagógicas de ejecución máxima no es significativa.

$$H_0: \rho = 0$$

Hipótesis alterna (H<sub>1</sub>): La correlación entre los estadígrafos psicométricos de validez y sensibilidad de las pruebas pedagógicas de ejecución máxima es significativa.

$$H_1: \rho \neq 0$$

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 pruebas pedagógicas de ejecución máxima consideradas en el espacio muestral, será:

$$gl = 10 - 2 = 8$$

Nivel de confianza: 95%; nivel de significación  $\alpha = 0,05$ ; y un error de 5%.

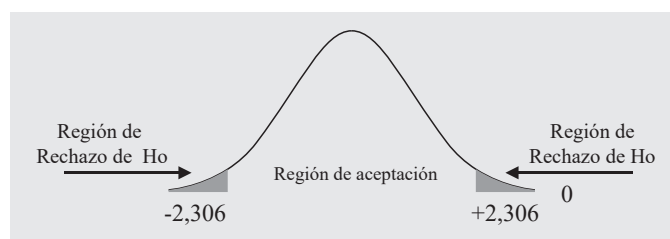
Prueba estadística: Se elige la prueba t de student bilateral, en función de la  $r = -0,111$  de Pearson promedio, para las diez pruebas pedagógicas de ejecución máxima

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

$$t = \frac{-0,111}{\sqrt{\frac{1-0,012}{8}}} = \frac{-0,111}{0,3514} = -0,3159$$

$$t \text{ calculada} = -0,3159$$

Región crítica: t tabulada =  $\pm 2,306$ . Puesto que  $H_1$  es bilateral, la región crítica abarca todos los valores de  $t \geq +2,306$  o  $t \leq -2,306$ .



- Decisión: Puesto que la t calculada está ubicada en la región de aceptación de la  $H_0$ ; es decir, la t calculada  $< t$  tabulada, se acepta la  $H_0$ .
- Conclusión Estadística

La correlación entre los estadígrafos psicométricos de sensibilidad y validez de las pruebas pedagógicas de ejecución máxima no es significativa.

## Discusión

Para dar solución en parte al problema de la investigación: ¿Qué relación existe entre los estadígrafos psicométricos de sensibilidad y 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?

En primer lugar hay que poner de manifiesto que la sensibilidad está explicada mediante el estadígrafo kurtosis, para cada uno de los ítems de las pruebas así como para el puntaje total, en los que se tuvo en cuenta el rango de valores: mayores que 0,263 ---indican mucha sensibilidad en diferentes grados---, el valor de 0,263 ---indica sensibilidad intermedia---, y los menores que 0,263 ---indican no mucha sensibilidad en diferentes grados---; tal como lo prescribe Shiefelbein (1999:286).

En segundo lugar, el cálculo de la confiabilidad como estadígrafo psicométrico debe tener un valor superior o igual a 0,60, Shiefelbein (1999:296), tanto para el ítem como para todo el instrumento de

recolección de datos.

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

El valor del coeficiente de correlación de Pearson promedio  $r = 0,179$ , entre los estadígrafos psicométricos de sensibilidad y confiabilidad es bajo, con una significatividad bilateral de 0,532, que por ser positivo se dice que existe una relación directa entre los estadígrafos; es decir, a mayor valor sensibilidad mayor valor de la confiabilidad, 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 sensibilidad y la confiabilidad, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es  $r^2 = 0,03441$ ; lo que quiere decir es que, en promedio el 3,4 % del valor de la sensibilidad de las pruebas está determinado por el valor de la confiabilidad, y el 96,6% se debe a otros factores. Por cierto, que, para emitir cualquier juicio valorativo sobre lo expuesto, habría que analizar no solo diez 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 --para todas la pruebas-- son bajas; se puede aseverar que la causalidad buscada entre la sensibilidad y la confiabilidad es mínima; pero si esta aseveración se generalizara, 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 --para las diez pruebas pedagógicas-- de los coeficientes de regresión lineal, se tienen mayores grados de dependencia, directos positivos, de la sensibilidad con respecto de la confiabilidad de las pruebas; y menores grados de influencia, directos positivos, de la confiabilidad en la sensibilidad de las respectivas pruebas.

Así también, contrastando los resultados del presente trabajo con los obtenidos por Cárdenas (2008, 2010, 2014, 2015 y 2017), se observa que hay que continuar con mayor cantidad de observaciones respecto a las relaciones existentes entre los diferentes estadígrafos psicométricos, porque en estos trabajos demuestra que, existe una relación de causalidad entre la facilidad-dificultad y la sensibilidad; así como, entre la validez, el poder discriminativo y nivel de dificultad; igualmente existe una correlación alta, fuerte, significativa e inversa entre la confiabilidad y la validez; así como, una correlación moderada, significativa e inversa, entre la confiabilidad y el poder de discriminación; en cambio en los resultados de Cárdenas (2016), éste obtiene una correlación baja, inversa y con poca significatividad entre el nivel de exigencia y la confiabilidad; de los ítems de las pruebas pedagógicas que se aplican a los alumnos de las instituciones educativas de la Región Junín.

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 (2000), quién pone de manifiesto que también se deben tener en cuenta a los estadígrafos edumétricos.

## Conclusiones

- ✓ La sensibilidad de la mayoría 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 términos cualitativos, están en el rango de no muy sensibles.
- ✓ El estadígrafo psicométrico de 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, fluctúa entre los valores 0,200 y 0,763; en términos cualitativos son válidos todos.

- ✓ La correlación es débil e inversa, con una significatividad bilateral promedio de 0,533; el coeficiente de determinación pone de manifiesto que sólo el 3,47% de la sensibilidad 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 valor de la validez y el 9653% se debe a otros factores
- ✓ 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 se tienen mayores grados--inversos-- de dependencia de la sensibilidad con respecto de la validez de las pruebas; y menores grados --también inversos-- de influencia de la validez en la sensibilidad de las respectivas pruebas.

## Sugerencias

- ✓ Realizar un estudio de la relación que existe entre los coeficientes de validez y nivel de dificultad, entre los coeficientes de validez y el poder de discriminación, entre el nivel de dificultad y el poder de discriminación, entre el poder de discriminación y la sensibilidad; de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín.
- ✓ Se sugiere difundir el presente trabajo de investigación, para su discusión y enriquecimiento a través de la crítica especializada.

## Referencias

- Aiken, L. (1995). Test psicológicos y evaluación. Prentice Hall Hispanoamericana.
- Ary; Jacobs y Razavieh. (1992). Introducción a la investigación pedagógica. McGraw- Hill.
- Cárdenas, A. (2007). Pertinencia de los Instrumentos de recolección de datos con los métodos estadísticos de confiabilidad. (I.I. FPH-UNCP, trabajo de investigación inédita). Huancayo, Perú.
- Cárdenas, A. (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.
- Cárdenas, A. (2009). Poder discriminativo, nivel de exigencia y capacidad predictiva de los exámenes de selección, I.I.FPH-UNCP, trabajo de investigación inédito, Huancayo, Perú.
- Cárdenas, A. (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.
- Cárdenas, A. (2011). Índice de discriminación de los ítems politómicos en las pruebas pedagógicas. Prospectiva universitaria, V- 8, 01, 119-126. Huancayo, Perú.
- Cárdenas, A. (2012). Análisis estadístico del índice de dificultad de los ítems politómicos en las pruebas pedagógicas. Horizonte de la ciencia, 02, 74-81. Huancayo, Perú.
- Cárdenas, A. (2013). Índice de dificultad y asimetría de los ítems en las pruebas pedagógicas. I.I.FPH-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas, A. (2014). Confiabilidad y validez de las pruebas pedagógicas de ejecución máxima. I.I.FPH-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas, A. (2015). Confiabilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima. I.I.FPH-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas, A. (2016). Nivel de exigencia y confiabilidad de pruebas pedagógicas de ejecución máxima. I.I, FE-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas, A. (2017). Sensibilidad y confiabilidad de pruebas pedagógicas de ejecución máxima. I.E.I.,

FE-UNCP, trabajo de investigación inédito. Huancayo, Perú

- Davis, F. (1989). *Analyze des items*. Nauwelaerts.
- Delgado; Escurra y Torres. (2006). *La medición en psicología y educación*. Hozlo S.R.L.
- Fermín, M. (1995). *La evaluación, los exámenes y las calificaciones*. Kapeluz.
- Gonzalvo, G. (1988). *Diccionario de metodología estadística*. Morata S.A.
- Hambleton y Novik. (1989). *Toward an integration of theory and method for criterion referenced test*. Measurmet.
- Hernández; Fernández y Baptista. (2003). *Metodología de la investigación*. McGraw- Hill.
- Nunnally y Bernstein. (1995). *Teoría psicométrica*. Mc Graw-Hill.
- Quezada, N. (2010). *Metodología de la investigación-Estadística aplicada en la investigación*. MACRO.
- Sánchez y Reyes. (2006). *Metodología y Diseños en la Investigación Científica*. Visión Universitaria.
- Schiefelbein, E. (1999). *Teoría, técnicas, procesos y casos en el planeamiento de la educación*. Ateneo.
- Santibañez, J. (2000). *Manual para la evaluación del aprendizaje estudiantil (conceptos, procedimientos, análisis e interpretación para el proceso evaluativo)*. Trillas.
- Santisteban, C. (2009). *Principios de psicometría*. Síntesis.
- Tavella, N. (1991). *Análisis de los ítems en la construcción de instrumentos Psicométricos*. Trillas.
- Tembrick, T. (1981). *Evaluación*. Editorial Narcea.