

Sensibilidad y poder de discriminación de pruebas pedagógicas de ejecución máxima

Recepción: 23 agosto 2023

Aceptación: 25 octubre 2023

<https://doi.org/10.26490/uncp.investigacionyEducacion.2023.4.1.1886>

Sensitivity and power of discrimination of pedagogical testing of execution maximum

Edgar Aníbal Cárdenas Ayala
Universidad Nacional del Centro del Perú
ecardenas@uncp.edu.pe y anibalito.c@gmail.com
<https://orcid.org/0000-0002-8888-2401>

Nacionalidad: peruana. Investigador y docente de Ciencias Naturales, Métodos Cuantitativos para la Investigación Científica e Idiomas (inglés y portugués). Doctor en Ciencias de la Educación y Magister en Didáctica Universitaria; ambos por la Universidad Nacional del Centro de Perú.

Resumen

En el presente trabajo de investigación de tipo descriptivo correlacional, se estudiaron las variables sensibilidad y poder de discriminación de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín. Se planteó como objetivo general, analizar estadísticamente e interpretar la relación existente entre las citadas variables. Y, como hipótesis: existe una relación de causalidad entre los estadígrafos psicométricos de sensibilidad y de poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín. Arribando a la conclusión, entre otras, que existe una correlación $r = 0,951$, alta y directa con significatividad promedio de 0,01 así también, el coeficiente de determinación pone de relieve que hasta el 91 % del valor del nivel de exigencia está determinado por el valor del poder de discriminación, y el 9 % se debe a otros factores; y, que existe igualdad entre el grado de dependencia y el grado de influencia de la sensibilidad en el poder discriminativo y viceversa de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín

Palabras clave: Edumetría, Estadística e investigación educativa, Psicometría

Abstract

In the present research work of a correlational descriptive type, the variables sensitivity and power of discrimination of the items of the pedagogical tests of execution maximum that are applied to the students of the educational institutions of the Junín Region were studied. The general objective was to statistically analyze

and interpret the relationship between the aforementioned variables. And, as a hypothesis: there is a causal relationship between the psychometric statistics of sensitivity and the power of discrimination, of the items of the pedagogical tests of execution maximum that are applied to the students of the educational institutions of the Junín Region. Arriving at the conclusion, among others, that there is a correlation $r = 0,951$, high, direct with average significance of 0,01 as well, the determination coefficient highlights that up to 91% of the value of sensitivity is determined by the value of the power of discrimination, and 9% is due to other factors; and, that there is equality between the degree of dependence and the degree of influence of sensitivity on discriminative power and vice versa, of the pedagogical tests of execution maximum that are applied to the students of the educational institutions of the Junín Region.

Keywords: Edumetric, Statistics and educational investigation, Psychometric

Introducción

Desde hace más de una década y de manera casi ininterrumpida se vienen realizando de manera sistemática investigaciones centradas en averiguar las relaciones que existen entre los diferentes estadígrafos psicométricos de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín; pero aún no se ha agotado el estudio de todas las relaciones existentes, es por eso que en la presente investigación se realiza el estudio puntual de la relación entre los estadígrafos psicométricos de sensibilidad y de poder de discriminación de las mencionadas pruebas.

Por otro lado, el análisis cuantitativo y cualitativo de datos provenientes de la aplicación de las pruebas pedagógicas, implica, entre otros temas, el cálculo de los estadígrafos psicométricos sensibilidad y de poder de discriminación; quedando por determinar la relación que existe entre estos dos; es decir, existe primacía de uno 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, en el presente trabajo se calculó la correlación r de Pearson entre la sensibilidad y el poder de discriminación, la que tuvo un valor $r = 0,951$ alto, directo y con significatividad promedio de 0,01; así también, el coeficiente de determinación pone de manifiesto que hasta el 91 % del valor del nivel de exigencia está determinado por el valor del poder de discriminación, y el 9 % se debe a otros factores; y, que existe igualdad entre el grado de dependencia y el grado de influencia de la sensibilidad en el poder discriminativo y viceversa, de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín

Aproximaciones teóricas y conceptuales

Pruebas de ejecución máxima

Son aquellas que miden las características de la persona de acuerdo a una calificación de correcto o errado y pueden ser de rendimiento, aptitud y habilidad.

a. De rendimiento

Miden el aprendizaje que se ha producido como resultado de las experiencias en un programa de estudios o entrenamiento. No tienen un tiempo fijo establecido, puede referirse a lo aprendido en una semana, un mes, un año, etc. Se utilizan para tomar decisiones en relación al avance de una estudiante.

b. De aptitud

Señalan capacidades para adquirir ciertas conductas o habilidades, dadas las oportunidades adecuadas. Miden los resultados de experiencias de aprendizaje generales e incidentales y su marco de referencia se enfoca hacia el futuro. Tienen por finalidad predecir lo que puede aprender en el futuro.

c. De habilidad

Miden las experiencias de aprendizaje más amplias y tienen su marco de referencia en el presente e indican el poder para realizar una tarea en el aquí y el ahora.

Estadígrafos psicométricos de las pruebas de ejecución máxima

Todo instrumento de recolección de datos debe reunir los siguientes requisitos esenciales: confiabilidad, validez, poder de discriminación, nivel de exigencia y sensibilidad; el presente trabajo de investigación halló la relación que existe entre los estadígrafos psicométricos de sensibilidad y poder de discriminación de las pruebas de ejecución máxima de rendimiento, por lo tanto, las aproximaciones teóricas están orientadas a estos dos estadígrafos psicométricos.

Sensibilidad (Kurtosis – estadígrafo de apuntamiento)

La kurtosis es el indicador de lo plana o “picuda” que es una curva. Si una distribución tiene relativamente un elevado pico, se llama distribución leptokúrtica, mientras que si es achatada se denomina distribución platikúrtica, la distribución normal constituye una distribución mesokúrtica. Para una curva normal la kurtosis es igual a 0,263. Una kurtosis mayor a 0,263 indica que la curva es platikúrtica y, por tanto, no hay grandes diferencias entre los datos. Una kurtosis menor de 0,263 indica que la curva está nutrida de casos medios, presenta leptokurtosis y existen diferencias significativas entre los datos.

Para efectos de análisis estadístico de los instrumentos de recolección de datos (prueba pedagógica) a través del apuntamiento es pertinente introducir en concepto

de sensibilidad Shiefelbein (1999:287), es decir, un ítem componente de la prueba pedagógica o toda la prueba será sensible en la medida pueda medir las diferencias mínimas del nivel aprendizaje alcanzado por los estudiantes sobre la materia evaluada. Por ejemplo, una balanza que mide en kilogramos no es sensible al peso de cartas aéreas en el correo. De manera similar, una prueba pedagógica que mida conocimientos en forma muy general, será no muy sensible, de sensibilidad intermedia o muy sensible.

Si la distribución de los resultados de una prueba pedagógica presenta leptokurtosis será no muy sensible, este tipo de distribución se presenta cuando la mitad de los ítems tiene un poder de discriminación aceptable, y la otra mitad tiene un poder de discriminación muy alto; todos los sujetos pertenecientes del espacio muestral de la investigación contestan correctamente, incorrectamente o no contestan las preguntas de la prueba pedagógica. Cuando hay mucha variación en los resultados de una prueba, la distribución de los mismos presenta platikurtosis, indicando que existe mucha sensibilidad a lo que mide la prueba. Esto quiere decir que, los ítems han sido mejor formulados para determinar diferencias de conocimientos o comprensión. También es posible interpretar como que la distribución platikúrtica implica más heterogeneidad que la distribución leptokúrtica.

Si aplicamos la misma prueba pedagógica a dos grupos distintos de estudiantes, puede ser sensible en uno de ellos y poco sensible en el otro, lo cual depende de las características de los estudiantes o de la variación del conocimiento que éstos posean en cada grupo.

Tabla N°1

Clasificador de la sensibilidad

KURTOSIS	VALOR	SENSIBILIDAD
Mesokurtosis	Ku = 0,263	Sensibilidad intermedia
Leptokurtosis	Ku < 0,263	No muy sensible
Platikurtosis	Ku > 0,263	Muy sensible

Poder de discriminación

Mide el grado con el cual el ítem es capaz de establecer diferencias entre las personas con niveles altos y bajos de una habilidad, aptitud o conocimiento que está siendo evaluado. Para calcular, se separan los resultados de las pruebas considerando el grupo superior (27%) y el grupo inferior (27%), luego se obtienen separadamente para cada ítem el porcentaje de participantes que responden correctamente, ambos datos se restan y el resultado final es la discriminación que tiene el ítem. La fórmula es la siguiente:

$$Dis = GS - GI$$

Dónde:

GS: % del grupo superior que contestó correctamente el ítem.

GI: % del grupo inferior que contestó correctamente el ítem.

El valor obtenido debe ser positivo y para aceptar el ítem debe ser igual o mayor a 0,30.

El poder discriminativo de las pruebas pedagógicas representa el grado de diferenciación, separación, distanciamiento o dispersión de los resultados observados mediante las respuestas que el estudiante manifieste ante cada pregunta planteada. Desde la perspectiva psicométrica, se espera que, mayoritariamente, las respuestas acertadas, correctas o buenas provengan de aquellos estudiantes que también muestran un rendimiento o desempeño general adecuado en el total de la prueba, y que las respuestas erradas u omitidas provengan del grupo de estudiantes que presente un rendimiento general también deficiente en la prueba. Es decir, se espera que cada ítem sea capaz de permitir una diferenciación lo más nítida posible entre los postulantes de alto y bajo rendimiento, lo que, cuantitativa e idealmente, equivaldría a 100%, o en proporciones a 1,00. La magnitud de esta capacidad diferenciadora de los ítems se expresa por índices de discriminación, cuyo cálculo se realiza mediante el método correlacional, coeficiente de correlación punto biserial ($r_{p.bis}$).

$$P. D. = r_{p.bis} = \frac{(p_i - X_i)}{s_t} \left(\frac{p}{q} \right)^{1/4}$$

Dónde:

P.D.: Poder de discriminación del ítem.

p: promedio de las puntuaciones totales de los estudiantes que respondieron bien el ítem analizado.

X_i: promedio de las puntuaciones totales de todos los estudiantes considerados en el análisis.

s_t: desvío de las puntuaciones totales de todos los estudiantes considerados en el análisis.

p y q: proporciones de los estudiantes que, respondieron bien, mal u omitieron el ítem analizado.

Tabla N°2

Clasificador del poder de discriminación del ítem

TIPO DE DISCRIMINACIÓN	VALOR
Discriminación muy buena	de 0,40 a 0,99

Discriminación aceptable	de 0,30 a 0,39
Discriminación intermedia	de 0,20 a 0,29
Discriminación inaceptable	de 0,05 a 0,19

Fuente: (Santibáñez, 2001, p. 187)

Metodología 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 primario, secundario y tecnológico de la Región Junín, los estudiantes de la Maestría y Doctorado en Educación, en los años académicos 2021 al 2022, en la asignatura de Análisis de datos cuantitativos y cualitativos; las mismas que se consignan en los anexos del presente trabajo de investigación.

Muestra

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 Escolares: 2021 al 2022.

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 descriptivo correlacional.

Método básico de la investigación

El método utilizado en la investigación fue el descriptivo

Diseño de la investigación

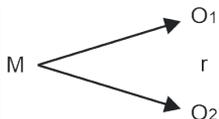
El diseño utilizado en esta investigación fue el descriptivo, porque resulta imposible el control experimental riguroso de las variables. Además, este diseño se adecua perfectamente a la presente investigación.

Dentro de los diseños descriptivos se encuentra el diseño descriptivo correlacional, y es el que se aplicó para determinar la relación entre los estadígrafos psicométricos, sensibilidad y poder de discriminación, de los ítems de las

pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes 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, sensibilidad y poder de discriminación, los que fueron analizados. (Sánchez, 2006, p. 105).

El diseño puede diagramarse de la siguiente manera:



Dónde:

M: Ítems de la prueba pedagógica que se aplican a los estudiantes de las instituciones educativas de la Región Junín.

O₁, O₂: Estadígrafos psicométricos de, sensibilidad 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 estudiantes de las instituciones educativas de Región Junín.

Variable 02: (V-02)

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

Técnicas e instrumentos de recopilación de datos

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

Estadígrafos psicométricos, sensibilidad y poder de discriminación, 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 —mencionados en el anterior acápite— 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 sensibilidad y poder de discriminación. Los diez 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 tipo cuantitativa, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics 25, con los menús correspondientes: estadísticos descriptivos —media, suma, desvío y kurtosis—. El cálculo de los estadígrafos se realizó para cada uno de los ítems, en cada una de las diez pruebas. Los datos, pueden observarse en los Anexos respectivos.
- La sensibilidad —teniendo en cuenta que todas las pruebas fueron dicotómicas—; se determinó para cada ítem, mediante la aplicación del software IBM SPSS Statistics 25, con el menú kurtosis. Los resultados de estas sensibilidades, para cada uno de los ítems de las diez pruebas, estuvieron comprendidos entre 0,259 y 0,817; dando lugar al comentario apreciativo singular en términos de no muy sensible, sensibilidad intermedia y muy sensible, (Shiefelbein, 1999, p. 286).

El poder de discriminación, se hizo mediante el cálculo de medias, desvíos y la aplicación de la fórmula respectiva. Para la interpretación de los valores de los coeficientes de poder de discriminación de cada uno de los ítems de las pruebas, se tuvo en cuenta el rango de valores: superiores a 0,30 —de 0,30 a 0,39 cualitativamente indican poder de discriminación aceptable y de 0,40 a 0,99 cualitativamente indican poder de discriminación muy bueno—; (Santibáñez, 2001, p. 187).

Resultados

Tabla N°3

Resumen de correlaciones de pearson entre los estadígrafos psicométricos de sensibilidad y poder de discriminación

Prueba Pedagógica	Correlación de Pearson r	Coefficiente de determinación r ²	Variabilidad explicada r ² (%)
Nº 01	0,816	0,666	0,666
Nº02	0,981	0,962	0,962
Nº 03	0,990	0,980	0,980
Nº 04	0,962	0,925	0,925
Nº 05	0,987	0,974	0,974

Nº 06	0,988	0,976	0,976
Nº 07	0,995	0,990	0,990
Nº 08	0,830	0,689	0,689
Nº 09	0,983	0,966	0,966
Nº 10	0,973	0,947	0,947
PROMEDIO	0,951	0,908	0,908

Para las diez pruebas pedagógicas (dicotómicas), se tiene una correlación de Pearson promedio $r = 0,951$; correlación alta y fuerte, con significatividad bilateral promedio de 0,01; por ser positiva, indica que existe una correlación directa entre los estadígrafos psicométricos de sensibilidad y poder de discriminación; es decir, a mayor valor de sensibilidad mayor valor del poder de discriminación de las pruebas. Por otro lado, el coeficiente de determinación promedio es $r^2 = 0,908$; lo que se interpreta —mediante el promedio de la variabilidad explicada— que hasta el 90,80% del valor del nivel de exigencia de las pruebas está determinado por el valor de poder de discriminación, y el 9,20% se debe a otros factores.

Tabla N°4

Resumen de regresiones lineales entre los estadígrafos psicométricos de sensibilidad y poder de discriminación

Prueba Pedagógica	Grado de dependencia del poder de discriminación con respecto de la sensibilidad b	Grado de influencia de la sensibilidad en el poder de discriminación k
Nº 01	0,764	0,871
Nº 02	0,946	1,018
Nº 03	1,068	0,918
Nº 04	0,978	0,947
Nº 05	1,016	0,959
Nº 06	1,081	0,903
Nº 07	0,975	1,015
Nº 08	1,034	0,665
Nº 09	0,997	0,979
Nº 10	1,028	0,921
PROMEDIO	0,989	0,920

Los promedios de los valores de los grados de dependencia del poder de discriminación con respecto de la sensibilidad y de los grados de influencia de la sensibilidad en el poder de discriminación de las diez pruebas pedagógicas son bajos y directos con una diferencia de 0,069 (diferencia no significativa). Por lo que

se asumen como iguales, entonces, se pone de manifiesto que existe igualdad entre el grado de dependencia y el grado de influencia de la sensibilidad en el poder discriminativo y viceversa.

Docimasia de hipótesis de la correlación promedio

- Hipótesis nula (H_0): La correlación entre los estadígrafos psicométricos de sensibilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima no es significativa.

$$H_0: p = q$$

- Hipótesis alterna (H_1): La correlación entre los estadígrafos psicométricos de sensibilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima es significativa.

$$H_1: p \neq q$$

- 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 diez 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 eligió la prueba t de student bilateral, en función de la $r = 0,876$ de Pearson promedio, para las diez pruebas pedagógicas de ejecución máxima.

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

$$t = \frac{0,951}{\sqrt{\frac{1-0,904}{8}}} = \frac{0,951}{0,096} = 9,906$$

$$t_{calculada} = +9,906$$

Región crítica: $t_{tabulada} = +/- 2,306$. 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 rechazo del H_0 ; es decir, la $t_{\text{calculada}} > t_{\text{tabulada}}$, se acepta la H_1 .
- Conclusión estadística: La correlación entre los estadígrafos psicométricos de sensibilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima es significativa.

Discusión final de resultados

Para tratar de dar solución al problema de la presente investigación: ¿Qué relación existe entre los estadígrafos psicométricos de sensibilidad y de poder de discriminación, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín?

En primer lugar, la sensibilidad de cada uno de los ítems que constituyen una prueba pedagógica, es el estadígrafo de apuntamiento —la kurtosis— de los calificativos obtenidos después de su administración; ésta mide las diferencias mínimas del nivel aprendizaje alcanzado por los estudiantes sobre la materia evaluada (Shiefelbein, 1999, p. 287); es decir, si un ítem tiene una sensibilidad de 0,263 —distribución normal o mesokúrtica— indica una sensibilidad intermedia, el ítem pone de manifiesto que las diferencias entre los aprendizajes no son significativas. En cambio, si el valor de la kurtosis es mayor que 0,263 —distribución platikúrtica— prescribe que no hay grandes diferencias entre los datos. Finalmente, si el valor de la kurtosis es menor que 0,263 —distribución leptokúrtica— indica que existen diferencias significativas entre los datos. Los estadígrafos de apuntamiento analizan, a cada uno de los ítems de la prueba pedagógica, que en términos cualitativos están caracterizados como: sensibilidad intermedia, muy sensible o no muy sensible (Nunnally y Berstein, 1995, p. 129).

En segundo lugar, el poder de discriminación, está explicado mediante el estadígrafo de relación, correlación punto biserial; para cada uno de los ítems de las pruebas, en los que se tuvo en cuenta el rango de valores: superiores a 0,30 —de 0,30 a 0,39 cualitativamente indican poder de discriminación aceptable y de 0,40 a 0,99 cualitativamente indican poder de discriminación muy bueno—; tal como lo prescribe (Santibáñez, 2001, p. 187).

En la presente investigación, se han observado estas consideraciones teóricas y los valores calculados de sensibilidad y de poder de discriminación 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 $r = 0,951$; entre los estadígrafos psicométricos de sensibilidad y de poder de discriminación es alta y fuerte, con una significatividad bilateral de 0,01; que por ser positivo pone de manifiesto que existe una relación directa entre los estadígrafos; es decir, a mayor sensibilidad mayor 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 sensibilidad y el poder de discriminación, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es $r^2 = 0,908$; lo que quiere decir es que, en promedio el 91 % del valor de la sensibilidad de las pruebas está determinado por el valor del poder de discriminación, y el 9% 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 también podría considerarse como 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 —entre las pruebas de ejecución máxima utilizadas como espacio muestral del presente estudio— son altos; entonces, se puede aseverar que la causalidad buscada entre la sensibilidad y el poder de discriminación; por lo tanto se podría generalizar la mencionada causalidad, para todas las pruebas pedagógicas de ejecución máxima que pudieran haber; es más, esta generalización no sería riesgosa, aun considerando que la muestra para la presente investigación no fue tan vasta; tampoco aleatoria.

Por otro lado, observando los resultados —para las diez pruebas pedagógicas de ejecución máxima— de los coeficientes de regresión lineal, se tienen iguales grados de dependencia y de influencia; es decir no existe subordinación, de la sensibilidad con respecto al poder de discriminación de las pruebas, ni viceversa.

Así también, contrastando los resultados del presente trabajo, con los obtenidos por Cárdenas en trabajos de investigación realizados los años 2008, 2009, 2010, 2011, 2014, 2015 y 2017, se observa que hay que continuar realizando una mayor cantidad de observaciones respecto a las relaciones existentes entre los diferentes estadígrafos psicométricos, porque en esos 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 (2018) concluye que la correlación entre la validez y la sensibilidad es débil e inversa en las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín. Cárdenas (2020) concluye que existe una correlación alta y directa entre el poder de discriminación y la validez en las pruebas de ejecución máxima aplicadas a los estudiantes de las instituciones educativas de la Región Junín. Finalmente, Cárdenas (2022) concluye que existe una correlación alta y directa entre el nivel de exigencia y el poder de discriminación en las pruebas de ejecución máxima aplicadas a los estudiantes 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; tales como los estadígrafos edumétricos. Santibáñez (2001).

Conclusiones

El valor cuantitativo de la 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, fluctúa entre 0,259 y 0,817; dando lugar al comentario apreciativo singular en términos de no muy sensible, sensibilidad intermedia y muy sensible,

El poder de discriminación de gran parte de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín, en términos cualitativos, está en el rango de aceptable y muy bueno.

La correlación $r = 0,951$ es alta y directa, con una significatividad bilateral promedio de 0,01; el coeficiente de determinación pone de manifiesto que hasta el 91% del valor del nivel de exigencia, —de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín— está determinado por el valor del poder de discriminación, y el 9% se debe a otros factores.

El análisis de regresión lineal, de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín, pone de manifiesto que, existe igualdad entre el grado de dependencia y el grado de influencia de la sensibilidad en el poder discriminativo y viceversa, de las respectivas pruebas.

Referencias

- Aiken, L. (1995). *Test psicológicos y evaluación*. Prentice Hall Hispanoamericana.
- Ary; Jacobs y Razaviev. (1992). *Introducción a la investigación pedagógica*. McGraw-Hill.
- Calvo, E. (2015). *Estadística: software estadísticos versus fórmulas estadísticas*. [Trabajo de investigación]. Centro de Producción de Documentos-UPSMP.
- Cárdenas, A. (2008). Análisis estadístico de los instrumentos de recolección de datos a través de los estadígrafos de deformación y apuntamiento. *Revista Horizonte de la ciencia*, 04, 79-88. <http://revistas.uncp.edu.pe/index.php/horizontedelaciencia>
- Cárdenas, A. (2009). *Poder discriminativo, nivel de exigencia y capacidad predictiva de los exámenes de selección* [Investigación no publicada] I.I.FPH-UNCP.
- Cárdenas, A. (2010). Relación entre la validez, el poder discriminativo y nivel de dificultad de las pruebas pedagógicas. *Revista Horizonte de la ciencia*, 01, 119-126. Huancayo, Perú. <http://revistas.uncp.edu.pe/index.php/horizontedelaciencia>
- Cárdenas, A. (2011). Índice de discriminación de los ítems politómicos en las pruebas pedagógicas. *Revista Prospectiva universitaria*, V- 8, 01, 119-126. UNCP. <http://www.uncp.edu.pe/universidad/investigación/index.php>
- Cárdenas, A. (2013). *Dificultad y sensibilidad de los instrumentos de recolección de datos*. Editora Imprenta Ríos S.A.C
- Cárdenas, A. (2014). *Confiabilidad y validez de las pruebas pedagógicas de ejecución máxima*. [Investigación no publicada] I.I.FE-UNCP.
- Cárdenas, A. (2015). Confiabilidad y poder de discriminación de las pruebas pedagógicas de ejecución máxima. *Revista Prospectiva universitaria*, V-13, 01 y 02, 85-90. UNCP . <http://www.uncp.edu.pe/universidad/investigación/index.php>
- Cárdenas, A. (2017). Sensibilidad y confiabilidad de pruebas pedagógicas de ejecución máxima. *Revista Educación y ciencia*, V-3, 09, 71-78. UNCP.
- Cárdenas, A. (2018). *Validez y sensibilidad de pruebas pedagógicas de ejecución máxima*. [Investigación no publicada] I.E.I., FE-UNCP.
- Cárdenas, A. (2020). Poder de discriminación y validez de pruebas de ejecución máxima. I.E.I., FE-UNCP, *Revista Investigación y educación*, 1, 8-21. <https://revistas.uncp.edu.pe/index.php/invest>

- Cárdenas, A. (2022). Nivel de exigencia y poder de discriminación de pruebas pedagógicas de ejecución máxima. *Revista Investigación y educación*, 1, 79-90 <https://doi.org/10.26490/uncp.investigaciónyeducación.2022.3.1.1668>
- Davis, F. (1989). *Analyze des items*. Nauwelaerts.
- Delgado; Ecurra y Torres. (2006). *La medición en psicología y educación*. Hozlo.
- Fermín, M. (1995). *La evaluación, los exámenes y las calificaciones*. Kapeluz.
- Gonzalvo, G. (1994). *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.
- Lozada y Sánchez. (2010). *Estadística inductiva*. Universidad de Holguín Oscar Lucero Moya.
- Nunnally y Bernstein. (1995). *Teoría psicométrica*. Mc Graw-Hill.
- Piris y Zacharzewski. (2017). *Sociedad estadística*. Universidad Nacional de Misiones.
- Quezada, N. (2010). *Metodología de la investigación-Estadística aplicada en la investigación*. MACRO.
- Rubio, A. (2018). *Cuidado al calcular las medidas de forma con el software Excel*. [Trabajo de investigación]. UNALM.
- Sánchez y Reyes. (2006). *Metodología y Diseños en la Investigación Científica*. Visión Universitaria.
- Santa, M. (2015). *Glosario Español-Inglés y English-Spanish de terminología estadística*. Universidad Nacional Abierta y a Distancia.
- Schiefelbein, E. (1999). *Teoría, técnicas, procesos y casos en el planeamiento de la educación*. Ateneo.
- Santibañez, J. (2001). *Manual para la evaluación del aprendizaje estudiantil (conceptos, procedimientos, análisis e interpretación para el proceso evaluativo)*. México: 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*. Narcea.

Thorndikee y Hagen. (1991). *Medición y evaluación en psicología y educación*. Trillas.

Visauta, B. y Martori, J. (2003). *Análisis estadístico con SPSS para Windows*, (Volumen II). McGraw-Hill.



© Los autores. Este artículo es publicado por la revista *Investigación y Educación* del Instituto Especializado de Investigación la Facultad de Educación de la Universidad Nacional del Centro del Perú. Es un artículo de acceso abierto, distribuido bajo los términos de la Licencia Atribución-No Comercial 4.0 Internacional. (<https://creativecommons.org/licenses/by-nc/4.0/>), que permite el uso no comercial y distribución en cualquier medio, siempre que la obra original sea debidamente citada.