

# Poder de discriminación y validez de pruebas de ejecución máxima

Presentación: 10 de julio de 2020

Aprobación: 10 de diciembre de 2020

Edgar Aníbal Cárdenas Ayala

Universidad Nacional del Centro del Perú

anibalito.c@gmail.com y ecardenas@uncp.edu.pe

Ingeniero Químico, Doctor en Ciencias de la Educación

## Resumen

En el presente trabajo de investigación de tipo tecnológica-aplicado se estudiaron las variables poder de discriminación y de validez 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 los estadígrafos psicométricos, de poder de discriminación y de validez, 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 positiva entre los estadígrafos psicométricos de poder de discriminación y de validez, 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 alta, directa y con significatividad promedio de 0,01 así también, el coeficiente de determinación pone de relieve que sólo el 86,7% del valor del poder de discriminación está determinada por el valor de la validez y el 13,3% se debe a otros factores; y, que se tienen mayores grados de influencia de la validez en el poder de discriminación y menores grados de dependencia del poder de discriminación de la validez 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

## *Power of discrimination and validity of pedagogical testing*

### Abstract

In this technological-applied research work, the variables of discrimination power and validity of the items of the pedagogical tests of maximum execution 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 psychometric stadigraphs, discrimination power and validity, of the items of the pedagogical tests of maximum execution that are applied to students of the educational institutions of the Junín Region. And, as a hypothesis: there is a positive relationship between the psychometric stadigraphs of discrimination power and validity, of the items of the pedagogical tests of maximum execution that are applied to the students of the educational institutions of the Junín Region. Coming to the conclusion, among others, that there is a high, direct correlation with an average significance of 0,01 as well, the coefficient of determination highlights that only 86,7% of the value of the discrimination power is determined by the value of validity and 13,3% is due to other factors; and, that there are higher degrees of influence of validity in the power of discrimination and lower degrees of dependence of the power of discrimination in the validity of the pedagogical tests of maximum execution that are applied to students of educational institutions in the Junín Region.

**Keywords:** Edumetric, Statistics and educational investigation, Psychometric

## Introducción

Desde el año de 2008 al año de 2019, se realizaron investigaciones centradas en averiguar algunas de las relaciones que existen entre los estadígrafos psicométricos de confiabilidad, poder de discriminación, nivel de dificultad (inteligibilidad), sensibilidad y validez de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes 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 análisis cuantitativo y cualitativo de datos, implica, entre otros temas, el cálculo de los estadígrafos psicométricos de los instrumentos de recolección de datos: de poder de discriminación y de validez; 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.

Por lo que, en el presente trabajo se calculó la correlación  $r$  de Pearson entre el poder de discriminación y la validez, la que tuvo un valor alto, directo y con significatividad promedio de 0,01; así también, el coeficiente de determinación pone de manifiesto que sólo el 86,7% del poder de discriminación está determinado por el valor de la validez; y, el 13,3% es debido a otros factores; finalmente, que existen mayores grados de influencia de la validez en el poder de discriminación y menores grados de dependencia del poder de discriminación de la validez 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.

### 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:

$$\text{Dis} = \text{GS} - \text{GI}$$

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 ( $r_{p,bis}$ )

$$P. D. = r_{p,bis} = (p_i - X_t) / s_t \cdot (p/q)^{1/2}$$

Dónde:

P.D.: Poder de discriminación del ítem

$p_i$ : promedio de las puntuaciones totales de los estudiantes que respondieron bien el ítem analizado

$X_t$ : 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°1

#### *Clasificador del poder de discriminación del ítem*

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 (2001:187)

## Validez

La validez de un instrumento está dada por grado en que efectivamente mide lo que procura medir, y nada más que ello. Si un instrumento carece de validez no es recomendable basarse en él para tomar decisiones que pretendan ser justificadas, ya que el instrumento mismo no tendrá sentido, tal como la manifiesta (Aiken 1995, p. 94).

Validez de un instrumento de medición, tradicionalmente, la validez de un instrumento de evaluación se define como el grado con el cual mide aquello para lo cual fue diseñado. Una desventaja de esta definición es la implicancia de que un instrumento tiene sólo una validez, que se supone se establece mediante en sólo estudio para determinar si el instrumento mide lo que debe medir. En realidad, un instrumento puede tener muchos tipos de validez, dependiendo de los propósitos específicos con los que se diseñó, como a la población a la que se dirige y al método para determinar su validez. (Aiken 1995, p. 95).

El término validez denota la utilidad científica de un instrumento de medida, en el que puede establecerse ampliamente que tan bien mide lo que pretende medir, señalando que ninguna medida es

útil a la larga sin evidencias de su validez, la cual se refiere a que tan bien mide lo que se supone que mide en el contexto en el que será aplicado. (Nunnally y Berstein 1995, p.92).

La validez se refiere al grado en que un instrumento proporciona información que es apropiada a la decisión que se toma; y siendo la validez la primera y más importante cuestión que debe uno plantearse con respecto a cualquier procedimiento de prueba, de debe contestar para lograr ello, las siguientes preguntas: ¿mide el instrumento lo que deseamos medir?, ¿en qué grado lo mide?, ¿mide sólo lo que se desea medir? (Thorndike y Hagen 1991, p. 61)

El coeficiente de la validez de un instrumento fluctúa entre cero y uno, donde un coeficiente de cero significa una validez nula y de uno representa una validez total. Cuanto más se acerque el coeficiente a cero hay más error aleatorio y sistemático en la medición realizada. Finalmente, la validez de un instrumento se refiere al grado en que mide lo que pretende medir y es determinada por los errores sistemáticos que se presentan en el proceso de medición.

Analíticamente se calcula el coeficiente de validez de criterio predictiva para cada ítem; mediante la correlación ítem-test o también llamada ítem-total.

Todos los coeficientes de validez de los ítems deben ser mayores o iguales que 0,2000; para que sean considerados válidos, tal como lo estipula (Shiefelbein 1999, p. 296)

## **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 en Educación, en los años académicos 2016 y 2019, 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: 2017 y 2019.

Tener en cuenta que las pruebas pedagógicas objetivas policotómicas de ejecución máxima para el desarrollo de la presente investigación fueron algo escasas.

### **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 utilizado en la investigación fue el descriptivo.

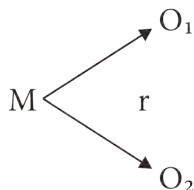
### **Diseño de la investigación**

El diseño que se empleó 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, nivel de dificultad (inteligibilidad) y validez, 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 nivel de dificultad (inteligibilidad) y validez, los que fueron analizados. Sánchez (2006:105).

El diseño puede diagramarse de la siguiente manera:



Dónde: M: Ítems de las pruebas pedagógicas que se aplican a los estudiantes de las instituciones educativas de la Región Junín.  
O<sub>1</sub>, O<sub>2</sub>: Estadígrafos psicométricos de poder de discriminación y validez, 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, poder de discriminación y validez, 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 a sus estudiantes los docentes.

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

## Procedimientos de recopilación 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 nivel de dificultad (inteligibilidad) y validez. 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 carácter cuantitativo, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics 24, con los menús correspondientes: estadísticos descriptivos —media y desvío— y escala —análisis de la fiabilidad—. El cálculo de los estadígrafos se realizó para cada uno de los ítems y para el puntaje total obtenido en cada una de las diez pruebas. Los datos, pueden observarse en los anexos respectivos.

- El cálculo del poder de discriminación, se realizó con el computador y con el paquete mencionado líneas arriba, mediante el cálculo de medias y desvíos para luego aplicar la fórmula respectiva. Para el análisis 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 (2001:187).
- El cálculo de la validez (también lo realizó el computador con el paquete mencionado líneas arriba) el que está basado en la fórmula para el cálculo del coeficiente de validez de criterio predictiva para cada ítem; mediante la correlación ítem–test o también llamada ítem–total. 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 cada uno de los ítems de las diez pruebas son válidos; por ende, las diez pruebas también son válidas, Shiefelbein (1999:296).

## Resultados

**Tabla N°2**

*Resumen de correlaciones de Pearson entre los estadígrafos psicométricos de poder de discriminación y validez*

<b>PRUEBA PEDAGÓGICA</b>	<b>Correlación de Pearson R</b>	<b>Coefficiente de determinación <math>r^2</math></b>	<b>Variabilidad explicada <math>r^2</math> (%)</b>
Nº 01	0,928	0,861	86,1
Nº 02	0,973	0,947	94,6
Nº 03	0,955	0,955	95,5
Nº 04	0,916	0,916	91,6
Nº 05	0,918	0,918	91,8
Nº 06	0,945	0,945	94,5
Nº 07	0,922	0,922	92,2
Nº 08	0,352	0,352	35,2
Nº 09	0,914	0,914	91,4
Nº 10	0,935	0,935	93,5
<b>PROMEDIO</b>	0,876	0,867	86,7

Fuente: Elaboración propia

Para las diez pruebas pedagógicas (dicotómicas), se tiene una correlación de Pearson promedio  $r = 0,876$ ; correlación alta y fuerte, con significatividad bilateral promedio de 0,01, por ser positiva, existe una correlación directa entre los estadígrafos psicométricos de poder de discriminación y validez; es decir, a mayor valor del poder de discriminación mayor valor de la validez de las pruebas. Por otro lado, el coeficiente de determinación promedio es  $r^2 = 0,867$ ; lo que se interpreta —mediante el promedio de la variabilidad explicada— que hasta el 86,70% del valor del poder de discriminación de las pruebas está determinado por el valor de la validez, y el 13,30% se debe a otros factores.

**Tabla N°3**

*Resumen de regresiones lineales entre los estadígrafos psicométricos de poder de discriminación y validez*

<b>PRUEBA PEDAGÓGICA</b>	<b>Grado de dependencia del poder de discriminación con respecto de la validez b</b>	<b>Grado de influencia de la validez en el poder de discriminación k</b>
N° 01	0,809	1,064
N° 02	0,874	1,084
N° 03	0,838	1,117
N° 04	0,765	1,198
N° 05	1,017	0,902
N° 06	0,822	1,149
N° 07	1,010	0,913
N° 08	1,007	0,349
N° 09	0,857	1,065
N° 10	0,801	1,168
<b>PROMEDIO</b>	<b>0,880</b>	<b>1,005</b>

Fuente: Elaboración propia

Para las diez pruebas pedagógicas, en promedio, existe menor grado de dependencia, del poder de discriminación con respecto de la validez; y mayor grado de influencia de la validez en el poder de discriminación de las respectivas pruebas.

### **Docimasia de hipótesis de la correlación promedio**

- Hipótesis nula ( $H_0$ ): La correlación entre los estadígrafos psicométricos de poder de discriminación y validez 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 poder de discriminación y validez 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

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

$$t = \frac{0,876}{\sqrt{\frac{1-0,867}{8}}} = \frac{0,876r}{0,1289} = 6,796$$

$$t_{calculada} = -6.796$$

- 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  $t_{calculada} > t_{tabulada}$ , se acepta la  $H_1$ .

### Conclusión estadística:

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

### Discusión final de resultados

Para dar solución en parte al problema de la investigación: ¿Qué relación existe entre los estadígrafos psicométricos de poder de discriminación y de validez, 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, hay que poner de manifiesto que el poder de discriminación está explicado mediante el estadígrafo de relación correlación punto biserial 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 (2001:187).

En segundo lugar, el cálculo de la validez como estadígrafo psicométrico debe tener un valor superior o igual a 0,2000, para cada uno de los ítems. Shiefelbein (1999:296)

En la presente investigación, se han observado estas consideraciones teóricas y los valores calculados de poder de discriminación y de validez 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  $r = 0,876$ ; entre los estadígrafos psicométricos de poder de discriminación y de validez es alta y fuerte, con una significatividad bilateral de 0,01; que por ser positivo se dice que existe una relación directa entre los estadígrafos; es decir, a mayor poder de discriminación mayor validez, 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 el poder de discriminación y el de validez, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es  $r^2 = 0,867$ ; lo que quiere decir es que, en promedio el 86,7 % del valor del poder de discriminación de las pruebas está determinado por el valor de la validez, y el 13,3% 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 —entre las pruebas de ejecución máxima utilizadas como espacio muestral del presente estudio— son altos; entonces, se puede aseverar que la causalidad buscada entre el poder de discriminación y la validez es fuerte; 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 menores grados de dependencia, es decir existe menor subordinación, del poder de discriminación con respecto de la validez de las pruebas; y mayores grados de influencia, de la validez en el poder de discriminación de las respectivas pruebas.

Así también, contrastando los resultados del presente trabajo con los obtenidos por Cárdenas (2008, 2010, 2014, 2015, 2017, 2018 y 2019), 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, así también, 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. Finalmente Cárdenas (2019) arriba a la conclusión que la correlación entre el nivel de dificultad (inteligibilidad) y la validez es débil y directo en las pruebas de ejecución máxima aplicadas a los estudiantes de la institución educativa 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 (2000).

## **Conclusiones**

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.

El valor cuantitativo de la 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 0,200 y 0,837; lo que en términos cualitativos significa validez total.

La correlación es alta y directa, con una significatividad bilateral promedio de 0,01; el coeficiente de determinación pone de manifiesto que sólo el 86,7% del valor del poder de discriminación 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 de la validez y el 13,3% 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 se tienen mayores grados —directos— de influencia de la validez en el poder de discriminación de las pruebas; y menores grados —también directos— de dependencia del poder de discriminación de la validez de las respectivas pruebas.

## Referencias bibliográficas

- Aiken, L. (1995). *Test psicológicos y evaluación*. México D.F., México: Prentice Hall Hispanoamericana.
- Ary; Jacobs y Razavieh. (1992). *Introducción a la investigación pedagógica*. México D.F., México: McGraw-Hill.
- Cárdenas Ayala, 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 Ayala, 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 Ayala, A. (2011). *Índice de discriminación de los ítems politómicos en las pruebas pedagógicas*. Prospectiva universitaria, V- 8, 01, 119-126. Huancayo, Perú.
- Cárdenas Ayala, 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 Ayala, A. (2018). *Validez y sensibilidad de pruebas pedagógicas de ejecución máxima*. I.E.I., FE-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas Ayala, A. (2019). *Nivel de dificultad (inteligibilidad) y validez de pruebas pedagógicas de ejecución máxima*. Horizonte de la ciencia, 19, 209-219. Huancayo, Perú.
- Davis, F. (1989). *Analyze des items*. Lovaina, París: Nauwelaerts.
- Delgado; Escurra y Torres. (2006). *La medición en psicología y educación*. Lima, Perú: Hozlo S.R.L.
- Fermín, M. (1995). *La evaluación, los exámenes y las calificaciones*. Buenos Aires, Argentina: Kapeluz.
- Gonzalvo, G. (1988). *Diccionario de metodología estadística*. Madrid, España: Morata S.A.
- Hambleton y Novik. (1989). *Toward an integration of theory and method for criterion referenced test*. New York, USA: Measurmet.
- Hernández; Fernández y Baptista. (2003). *Metodología de la investigación*. México: McGraw-Hill.
- Nunnally y Bernstein. (1995). *Teoría psicométrica*. México: Editorial 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.
- Schieffelbein, E. (1999). *Teoría, técnicas, procesos y casos en el planeamiento de la educación*. Buenos Aires, Argentina: Ateneo.
- Santibañez, J. (2000). *Manual para la evaluación del aprendizaje estudiantil (conceptos, procedimientos, análisis e interpretación para el proceso evaluativo)*. México: Trillas.
- Santisteban, C. (2009). *Principios de psicometría*. Madrid, España: Editorial Síntesis.
- Tavella, N. (1991). *Análisis de los ítems en la construcción de instrumentos Psicométricos*. México: Trillas.
- Tembrick, T. (1981). *Evaluación*. Madrid, España: Editorial Narcea.
- Thorndike 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.