

Nivel de dificultad (inteligibilidad) y validez de pruebas pedagógicas de ejecución máxima

Manalip tupuyninpi (yačhakyaypi) yačhachina tapukuykuna allipkayninwan anchap lulanapi

Okarapë kengagantsi komëtaropage(ogotagantsi) aike monkaratagantsi ogotagantsipagekë omanapagesonori

Okantayeta opomerentsiyeti aisati okametsayeti omonkarayetatari ayosanoyetiri

Recibido: 21 Junio 2019 Corregido: 06 Octubre 2019 Aprobado: 06 Abril 2020

Edgar Aníbal Cárdenas Ayala

Nacionalidad: Peruana, Filiación: Universidad Nacional del Centro del Perú

Correos: ecardenas@uncp.edu.pe y anibalito.c@gmail.com;

ORCID: <https://orcid.org/0000-0002-8888-2401>

Resumen

En el presente trabajo de investigación de tipo tecnológica–aplicada se estudiaron las variables nivel de dificultad (inteligibilidad) 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 nivel de dificultad (inteligibilidad) 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 nivel de dificultad (inteligibilidad) 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 muy baja, directa y con significatividad promedio de 0,543 así también, el coeficiente de determinación pone de relieve que sólo el 8,30% del nivel de dificultad (inteligibilidad) está determinada por el valor de la validez y el 91,70% se debe a otros factores; y, que el nivel de dependencia del nivel de dificultad (inteligibilidad) con respecto de la validez es mayor que el nivel de influencia de la validez en el nivel de dificultad (inteligibilidad); de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín.

Palabras clave:

Psicometría, inteligibilidad, validez, pruebas pedagógicas.

Lisichiku limaykuna:

Allikaytupuy, yačhakyay, allikaynin, yačhachina tapuy.

Nibarintsipage Katingaro:

Ogotokotagantsi, ogotaneri, kaninaro, monkaratagantsi ogotagantsipagekë

Nantsipe ayoyeteri:

monkaramentotsi, Ayenkishiriane, Okametsatiri, monkaratirori ayoyetakeri

Level of Difficulty (Intelligibility) and Validity of Pedagogical Testing

Abstract

It was statistically analyzed and interpreted the relationship between psychometric statisticians of difficulty level (intelligibility) and validity of the items of pedagogical testing applied students in educational institutions of Junín Region. It had specific objectives, calculate psychometric statisticians of difficulty level (intelligibility) and validity; correlation and regression coefficients, between the difficulty level (intelligibility) and validity of pedagogical testing applied students in educational institutions of Junín Region. It had like hypothesis: there is a relationship between psychometric statisticians of difficulty level (intelligibility) and validity of pedagogical testing applied students in educational institutions of Junín Region. It came to the conclusion, among others, that there is a very low, direct correlation with an average significance of 0,543, as well, the determination coefficient highlights that only 8,30% of the level of difficulty (intelligibility) is determined by the value of the validity and 91,70% is due to other factors; and, that the level of dependence of the level of difficulty (intelligibility) with respect to validity is greater than the level of influence of validity on the level of difficulty (intelligibility); of the maximum execution pedagogical tests that are applied to the students of the educational institutions of the Junín Region.

Keywords

Psychometric, intelligibility, validity, pedagogical testing.

Nível de dificuldade (inteligibilidade) e validade de provas pedagógicas de execução máxima

Resumo

Na presente pesquisa de tipo tecnológica-aplicada se estudaram as variáveis de nível de dificuldade (inteligibilidade) e validade dos itens das provas pedagógicas de execução máxima aplicadas aos estudantes das instituições educativas da região de *Junín*. O objetivo geral foi analisar estatisticamente e interpretar a relação existente entre os estatísticos psicométricos de nível de dificuldade (inteligibilidade) e validade dos itens das provas pedagógicas de execução máxima aplicados aos estudantes em instituições educativas da região de *Junín*. E como hipótese: existe uma relação entre estatísticos psicométricos de nível de dificuldade (inteligibilidade) e validade dos itens das provas pedagógicas de execução máxima aos estudantes de instituições educativas da região de *Junín*. Concluiu-se que existe uma correlação muito baixa, direta e com significativa média de 0,543, do mesmo modo, o coeficiente de determinação ressalta que apenas 8,30% do nível de dificuldade (inteligibilidade) é determinado pelo valor da validade e 91,70% é devido a outros fatores; e que o nível de dependência em relação à validade é maior que o nível de influência da validade no nível de dificuldade (inteligibilidade) das provas pedagógicas de execução máxima aplicadas aos estudantes das instituições educativas da região de *Junín*.

Palavras-chave:

Psicometria, inteligibilidade, validade, provas pedagógicas.

Datos del autor

Edgar Anibal 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 y Magíster en Didáctica Universitaria; ambos por la Universidad Nacional del Centro de Perú.

Introducción

Entre los años 2008 al 2018, 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 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 nivel de dificultad (inteligibilidad) 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.

Para dicho efecto, en el presente trabajo de; así también, el coeficiente de determinación pone de relieve que sólo el 8,30% del nivel de dificultad (inteligibilidad) está determinada por el valor de la validez y el 91,70% se debe a otros factores; y, que el nivel de dependencia del nivel de dificultad (inteligibilidad) con respecto de la validez es mayor que el nivel de influencia de la validez en el nivel de dificultad (inteligibilidad); de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín.

Nivel de dificultad (inteligibilidad) — estadígrafos de Deformación

La deformación consiste en analizar la simetría o asimetría (ausencia de simetría) de las distribuciones. La simetría puede ser positiva o negativa. La deformación estadística se deduce comparando la distribución con la forma de la curva normal que corresponde a una distribución simétrica. Si el valor calculado de la asimetría es +/- 0,37 o menor, la curva es —a efectos prácticos— simétrica. Entre +/- 0,37 y +/-1,00 la curva tiene una asimetría aceptable. Cuando la asimetría es mayor de +/-1,00 la curva es claramente asimétrica. (Nunnally y Berstein 1995, p. 128).

Para efectos de análisis estadístico de los instrumentos de recolección de datos a través de la deformación es pertinente introducir el concepto de dificultad o facilidad, es decir, un ítem componente de un instrumento de recolección de datos o todo el instrumento será fácil (legible, inteligible, etc.) si presenta una asimetría negativa; en cambio será todo lo contrario si presenta una asimetría positiva.

Por ejemplo, si el instrumento de recolección de datos es una prueba pedagógica, la asimetría positiva indica que el nivel de conocimiento o comprensión que está midiendo el instrumento es demasiado alto, en cambio si la asimetría es negativa indica que el nivel de conocimiento o comprensión que está midiendo el instrumento es demasiado bajo, para los niveles reales del grupo. También estará midiendo otros factores que no son parte del objetivo de la prueba pedagógica pero que influyen en los resultados. (Shiefelbein 1999, p. 286).

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

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 2018, 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 2018.

Tipo de investigación

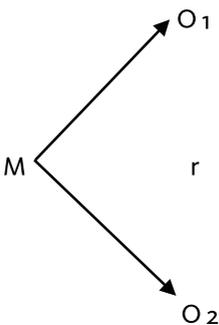
Dentro del campo de la investigación educativa y de acuerdo a las características de la hipótesis y los objetivos se enmarcó dentro del tipo de investigación tecnológica-aplicada.

Diseño de la investigación

El diseño que se empleó en esta investigación fue el correlacional (Sánchez y Reyes 2006, p.113), 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. Además, este diseño se adecua perfectamente a la presente investigación,

El diseño tiene el siguiente esquema:



Dónde: M: Ítems de las pruebas pedagógicas que se aplican a los estudiantes de las instituciones educativas de la Región Junín.
 O_1, O_2 : Estadígrafos psicométricos de, nivel de dificultad (inteligibilidad) y validez, respectivamente.
 r: Coeficiente de correlación.

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, 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 Región Junín.

Instrumentos de recopilación de datos

Las diez pruebas pedagógicas que se seleccionaron como muestra fueron las que se aplicaron.

Se consideraron a estas diez pruebas pedagógicas, debido a que los resultados de los estadígrafos psicométricos calculados --nivel de dificultad (inteligibilidad) y validez-- fueron óptimos para los objetivos del presente estudio.

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 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.

Las técnicas y procedimientos utilizados fueron estadísticos ya que la investigación es de carácter cuantitativa, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics 20, con los menús correspondientes: estadísticos descriptivos --asimetría--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. El cálculo del nivel de dificultad (inteligibilidad), se realizó con el computador con el paquete mencionado líneas arriba, mediante el estadígrafo asimetría. Para el análisis de los valores de los coeficientes de dificultad de cada uno de los ítems de las pruebas, se tuvo en cuenta el rango de valores: Iguales a 0,00 o muy cercanos --indican simetría; cualitativamente, dificultad intermedia--. Positivos, indican --asimetría positiva, dificultad--. Negativos, indican --asimetría negativa, facilidad--. (Shiefelbein 1999, p. 286).

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, p. 296).

Resultados

Tabla n° 1
Resumen de correlaciones de Pearson entre los estadígrafos psicométricos de nivel de dificultad (Inteligibilidad) y validez

PRUEBA PEDAGÓGICA	Correlación de Pearson r	Coefficiente de determinación r^2	Variabilidad explicada r^2 (%)
N°01	-0,144	0,021	2,10
N°02	-0,106	0,011	1,10
N°03	-0,047	0,002	0,20
N°04	-0,097	0,009	0,90
N°05	0,384	0,148	14,80
N°06	0,286	0,082	8,20
N°07	0,740	0,548	54,80
N°08	-0,066	0,004	0,40
N°09	0,085	0,007	0,70
N°10	0,037	0,001	0,10
PROMEDIO	0,107	0,083	8,30

Fuente: Elaboración propia

En la tabla N°2, para las diez pruebas pedagógicas (dicotómicas), se tienen mayores grados de dependencia, del nivel de dificultad (inteligibilidad) con respecto de la validez de las pruebas; y menores grados de influencia, de la validez en el nivel de dificultad (inteligibilidad) de las respectivas pruebas.

Docimasia de hipótesis de la correlación promedio

Hipótesis nula (H_0): La correlación entre los estadígrafos psicométricos de nivel de dificultad (inteligibilidad) 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 nivel de dificultad (inteligibilidad) y validez de las pruebas pedagógicas de ejecución máxima es significativa.

$$H_1: p \neq q$$

Distribución de la muestra: t de student en función de r 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%.

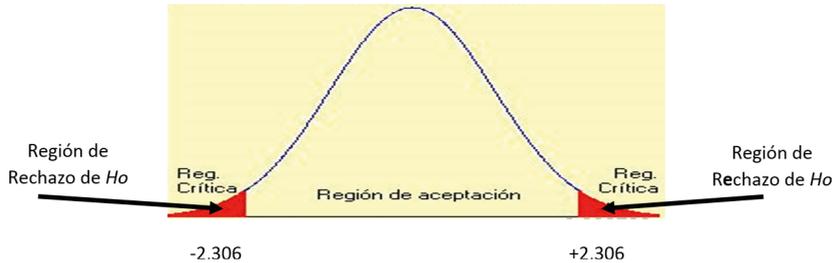
Prueba estadística: Se elige la prueba t de student bilateral, en función de la r de Pearson promedio calculada, $r = 0,107$ para las diez pruebas pedagógicas de ejecución máxima.

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

$$t = \frac{0,107}{\sqrt{\frac{1-0,083}{8}}} = \frac{0,107}{0,3386} = 0,3160$$

$$t \text{ calculada} = -0,3160$$

Región crítica: $t_{\text{tabulada}} = \pm 2,306$. 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 $t_{\text{calculada}} < t_{\text{tabulada}}$ se acepta la H_0 .

Conclusión estadística: La correlación entre los estadígrafos psicométricos de nivel de dificultad (inteligibilidad) y validez de las pruebas pedagógicas de ejecución máxima no es significativa.

Discusión

En primer lugar hay que poner de manifiesto que el nivel de dificultad (inteligibilidad) está explicado mediante el estadígrafo de asimetría, para cada uno de los ítems de las pruebas, en los que se tuvo en cuenta el rango de valores: iguales o muy próximos a 0,00 --indican dificultad (inteligibilidad) intermedia--, valores positivos mayores que 0,00 --indican mucha dificultad (inteligibilidad)--, y los valores negativos menores que 0,00 --indican mucha facilidad (inteligibilidad)--; tal como lo prescribe. (Shiefelbein 1999, p. 286).

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, p. 296)

En la presente investigación, se han observado estas consideraciones teóricas y los valores calculados de nivel de dificultad (inteligibilidad) 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 $r = 0,107$, entre los estadígrafos psicométricos de nivel de dificultad (inteligibilidad) y de validez es bajo o débil, con una significatividad bilateral de 0,543, que por ser positivo se dice que existe una relación directa entre los estadígrafos; es decir, a mayor dificultad 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 nivel de dificultad (inteligibilidad) y el de validez, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es $r^2 = 0,083$; lo que quiere decir es que, en promedio el 8,3 % del valor del nivel de dificultad (inteligibilidad) de las pruebas está determinado por el valor de la validez, y el 91,7% 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 el nivel de dificultad (inteligibilidad) y la validez 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, del nivel de dificultad (inteligibilidad) con respecto de la validez de las pruebas; y menores grados de influencia, de la validez en el nivel de dificultad (inteligibilidad) de las respectivas pruebas.

Así también, contrastando los resultados de la presente investigación con los obtenidos por el autor en sus investigaciones anteriores, se observa que hay que continuar con mayor cantidad de observaciones respecto a las relaciones existentes entre los diferentes estadígrafos psicométricos, ya que en sus investigaciones demuestra que, existe una relación de causalidad entre la facilidad-dificultad y la sensibilidad (estadígrafos de deformación y apuntamiento) (Cárdenas 2008); así como, entre la validez, el poder discriminativo y nivel de dificultad (Cárdenas 2010); igualmente existe una correlación alta, fuerte, significativa e inversa entre la confiabilidad y la validez (Cárdenas 2014); así como, una correlación moderada, significativa e inversa, entre la confiabilidad y el poder de discriminación (Cárdenas 2015); en cambio en los resultados de, éste obtiene una correlación baja, inversa y con poca significatividad entre el nivel de exigencia y la confiabilidad (Cárdenas 2017), así también, concluye que la correlación entre la validez y la sensibilidad es débil e inversa (Cárdenas 2018) en las pruebas pedagógicas de ejecución máxima que se aplican 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 2000).

Conclusiones

El nivel de dificultad (inteligibilidad) de la mayoría de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín.

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.

El coeficiente de determinación pone de manifiesto que sólo el 8,30% del valor del nivel de dificultad (inteligibilidad) 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 91,70% se debe a otros factores, esto se comprueba con el coeficiente de correlación.

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 hay mayor dependencia del nivel de dificultad (inteligibilidad) con respecto de la validez de las pruebas; y menor influencia de la validez en el nivel de dificultad (inteligibilidad) de las respectivas pruebas.

Recomendaciones

Realizar un estudio de la relación que existe entre los coeficientes de; validez y poder de discriminación, nivel de dificultad y poder de discriminación, y, poder de discriminación y sensibilidad; de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes 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 bibliográficas

- Aiken, L. (1995). *Test psicológicos y evaluación*. México D.F., México: Prentice Hall Hispanoamericana.
- Cárdenas Ayala, 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 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. (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. (2016). Nivel de exigencia y confiabilidad de pruebas pedagógicas de ejecución máxima. I.I, FE-UNCP, trabajo de investigación inédito. Huancayo, Perú.
- Cárdenas Ayala, A. (2017). Sensibilidad y confiabilidad de pruebas pedagógicas de ejecución máxima. I.E.I., FE-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
- Nunnally y Bernstein. (1995). *Teoría psicométrica*. México: Editorial Mc Graw-Hill.
- 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.
- Santibañez, J. (2000). *Manual para la evaluación del aprendizaje estudiantil (conceptos, procedimientos, análisis e interpretación para el proceso evaluativo)*. México: Trillas.
- Thorndike y Hagen. (1991). *Medición y evaluación en psicología y educación*. México: Trillas.