



Nivel de exigencia y confiabilidad de pruebas pedagógicas de ejecución máxima

Edgar Aníbal Cárdenas Ayala

Ingeniero Químico, Doctor en Ciencias de la Educación.

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

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

Resumen

En el presente trabajo de investigación de tipo tecnológica-aplicado se estudiaron las variables nivel de exigencia y de confiabilidad 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 exigencia y de confiabilidad, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los estudiantes de las instituciones educativas de la Región Junín. Y, como hipótesis: existe una relación entre los estadígrafos psicométricos de nivel de exigencia y de confiabilidad, 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 baja, inversa y con significatividad promedio de 0,479 así también, el coeficiente de determinación pone de relieve que sólo el 5% del nivel de exigencia está determinada por el valor de la confiabilidad y el 95% se debe a otros factores; y, que se tienen mayores grados de dependencia del nivel de exigencia de la confiabilidad y menores grados de influencia de la confiabilidad en el nivel de exigencia 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 aplicada a la investigación educativa. Psicometría

Exigency level and reliability of pedagogical testing

Abstract

In this technological-applied research work, the variables of exigency level and reliability 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, of exigency level and reliability, 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 relationship between the psychometric stadigraphs of exigency level and reliability, 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 0,479 as well, the coefficient of determination highlights that only 5% of the value of exigency level is determined by the value of reliability and 95% is due to other factors; and, that there are higher degrees of influence of exigency in the reliability and lower degrees of dependence of the reliability in the exigency level of the pedagogical tests of maximum execution that are applied to students of educational institutions in the Junín Region.

Key Words:

Edumetric, Statistics and educational investigation, Psychometric

Introducción

Desde el año 2008, se realizaron investigaciones centradas en averiguar algunas de las relaciones existentes entre los estadígrafos psicométricos de confiabilidad, poder de discriminación, nivel de dificultad, sensibilidad y validez de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín; quedando por hacer, el estudio de muchas otras relaciones que pudieran existir entre los estadígrafos psicométricos ya mencionados.

Por otro lado, las Técnicas del Procesamiento de Datos e Información, implican, entre otros temas, el cálculo de los estadígrafos psicométricos de los instrumentos de recolección de datos: nivel de exigencia (dificultad) y confiabilidad; quedando en el tintero la relación que existe entre estos dos; es decir, existe primacía de una de ellos sobre el otro, uno es causa o efecto del otro, por qué hay que calcular siempre los dos o es necesario solo uno de ellos.

Entonces, para poder subsanar, en parte, el inconveniente citado líneas arriba, en el presente trabajo de investigación se determinó que, existe una correlación baja, inversa y con poca significatividad; así también, el coeficiente de determinación pone de manifiesto que sólo el 5% del nivel de exigencia está determinado por el valor de la confiabilidad y el 95% se debe a otros factores; y, que el grado de dependencia del nivel de exigencia con respecto de la confiabilidad es mayor que el grado de influencia de la confiabilidad en el nivel de exigencia; de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, pone de manifiesto

Método

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.

Nivel de exigencia: dificultad-facilidad (estadígrafos de deformación)

La deformación consiste en analizar la simetría o asimetría (ausencia de simetría) de las distribuciones Nunnally / Berstein (1995:128). La asimetría 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.

Para efectos de análisis estadístico de los instrumentos de recolección de datos a través de la deformación es pertinente introducir en concepto de dificultad o facilidad Shiefelbein (1999:286), es decir, un ítem componente de un instrumento de recolección de datos o todo el instrumento será fácil (legible, entendible, inteligible, etc.) si presenta una asimetría negativa; en cambio será todo lo contrario si presenta una

asimetría positiva.

Tabla 1. Clasificador de los ítems por índice de facilidad--dificultad (inteligibilidad)

VALORES DE LA ASI-METRÍA	Facilidad--dificultad	Inteligibilidad
De: -2,001 a menos	Muy fácil	Es muy inteligible
De: -1,001 a -2,000	Fácil	Es inteligible
De: -0,201 a -1,000	Medianamente fácil	Medianamente inteligible
De: -0,001 a -0,200	Tendencia a la facilidad	Tendencia a la inteligibilidad
0,000	Dificultad intermedia	Inteligibilidad intermedia
De: 0,001 a 0,200	Tendencia a la dificultad	Tendencia a la ininteligibilidad
De: 0,201 a 1,00	Medianamente difícil	Medianamente ininteligible
De: 1,001 a 2,000	Difícil	Es ininteligible
De: 2,001 a más	Muy difícil	Es muy ininteligible

Fuente: Elaboración propia

Confiabilidad

En muchas ocasiones, a la hora de recabar información sobre aspectos diversos de la vida o actividad de los individuos, se utilizan instrumentos Aiken (1996:90) tales como: pruebas pedagógicas, fichas de observación, cuestionarios, escalas, test, listas de cotejo, etc., constituidos por una serie de ítems o enunciados que deben estar relacionados entre sí y cuyas puntuaciones individuales previamente sumadas o promediadas nos dan el valor global de los individuos en el citado instrumento, sea éste una escala, un test una lista de cotejo, etc. Ejemplos de este tipo de instrumentos se utilizan a diario en la práctica investigadora: escalas de diferentes tipos para medir actitudes de los individuos hacia personas, objetos o situaciones; encuestas con varios enunciados para diagnosticar situaciones concretas del factor humano, la cultura, la excelencia, etc.; mallas para evaluar condiciones ergonómicas o psico-sociológicas del trabajo de los individuos, test para determinar o diagnosticar conocimientos, aptitudes, rasgos de personalidad, etc., de las personas; y un sinnúmero de aplicaciones en cualquier área de la actividad humana.

Se dice que un instrumento de recolección de datos es confiable Hernández y otros (2003), cuando con el mismo se obtienen resultados similares al aplicarlo dos o más veces al mismo grupo de individuos o cuando lo que se aplica o administra son formas alternativas del instrumento.

Coefficientes de confiabilidad de instrumentos de recolección de datos:

Alfa de Cronbach:

$$AC = \left(\frac{K}{K-1} \right) \frac{1 - \sum s_i^2}{s_t^2}$$

Dónde:

AC : Coeficiente de confiabilidad

k : Número de ítems del instrumento.

s_i^2 : Sumatoria de varianzas individuales de cada ítem.

s_t^2 : Varianza total del instrumento.

Tabla 2. Clasificación de la confiabilidad

Confiabilidad nula	menos de 0,53
Confiabilidad baja	de 0,54 a 0,59
Confiable	de 0,60 a 0,65
Muy Confiable	de 0,66 a 0,71
Confiabilidad excelente	de 0,72 a 0,99
Confiabilidad perfecta	1,00

Fuente: Shiefelbein (1999:296)

Población

Se tomó como población de la investigación a todas las pruebas pedagógicas objetivas de ejecución máxima que aplicaron a sus alumnos, en las diferentes instituciones educativas del nivel secundario de la Región Junín, los estudiantes de la Maestría en Educación, hasta el año académico 2015, así como, los egresados del Pregrado de la Facultad de Educación hasta el año 2015; las mismas que están consignadas en los anexos de sus respectivas tesis de grado.

Muestra

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

Nota: Las pruebas pedagógicas objetivas policotómicas de ejecución máxima son algo escasas.

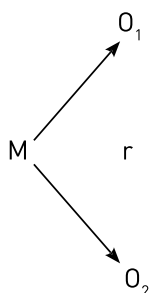
Diseño de la investigación

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

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

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

El diseño puede diagramarse de la siguiente manera:



Dónde:

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

O_1, O_2 : Estadígrafos psicométricos, de nivel de exigencia y confiabilidad, respectivamente.

r: Coeficiente de correlación.

Variable independiente: (VI)

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

Variable dependiente: (VD)

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

Instrumentos de recopilación de datos

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

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

Procedimientos de recopilación de datos

Los docentes –mencionados en el párrafo anterior– 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 exigencia y confiabilidad. Los seis instrumentos de recolección de datos, fueron aplicados por los docentes a sus estudiantes, en sus respectivas instituciones educativas

Técnicas de procesamiento y análisis de datos

Siendo la investigación de carácter cuantitativo, para el cálculo se utilizó el paquete estadístico IBM SPSS Statistics v20, 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 seis pruebas. Los datos, pueden observarse en los Anexos respectivos.

- El cálculo del nivel de exigencia (lo realizó el computador con el paquete mencionado líneas arriba) mediante el estadígrafo asimetría. Para el análisis de los valores de los coeficientes de asimetría de cada uno de los ítems de las pruebas así como del puntaje total, se tuvo en cuenta el rango de valores: positivos –indican dificultad en diferentes grados–, cero –indican dificultad intermedia–, y los negativos – indican facilidad en diferentes grados–; Shiefelbein (1999:286).
- El cálculo de la confiabilidad (también lo realizó el computador con el paquete mencionado líneas arriba) el que está basado en la fórmula: alfa de Cronbach. Los resultados de los coeficientes de confiabilidad de cada uno de los ítems las pruebas, fueron mayores o iguales que 0,60; por lo que se concluyó que las seis pruebas son confiables, Shiefelbein (1999:296).

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

Tabla 3. Resumen de correlaciones de Pearson entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad

Prueba pedagógica	Correlación de Pearson r	Coefficiente de determinación r^2	Variabilidad explicada r^2 (%)
policotómica n° 01	-0,113	0,013	1,3%
policotómica n° 02	0,136	0,019	2,0%
dicotómica n° 01	-0,204	0,042	4,2%
dicotómica n° 02	-0,416	0,173	17,0%
dicotómica n° 03	-0,122	0,015	1,5%
dicotómica n° 04	0,191	0,036	3,6%
promedio	-0,157	0,050	5,0%

Para las seis pruebas pedagógicas (dos policotómicas y cuatro dicotómicas), se tiene una correlación de Pearson promedio $r = -0,157$; correlación baja, con significatividad bilateral de 0,479, por ser negativa, existe una correlación inversa entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad; es decir, a mayor nivel de exigencia menor valor de la confiabilidad de la prueba. Por otro lado, el coeficiente de determinación promedio es $r^2 = 0,050$; lo que quiere decir –mediante el promedio de la variabilidad explicada– que hasta el 5% del nivel de exigencia de las pruebas está determinado por el valor de la confiabilidad, y el 95% se debe a otros factores.

Tabla 4. Resumen de regresiones lineales entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad

Prueba pedagógica	Grado de dependencia del poder de discriminación con respecto de la confiabilidad	Grado de influencia de la confiabilidad en el poder de discriminación
	b	k
policotómica n°01	-3,702	-0,003
policotómica n°02	7,779	0,002
dicotómica n°01	-15,118	-0,003
dicotómica n°02	-56,817	0,003
dicotómica n°03	-11,224	-0,001
dicotómica n°04	3,836	0,009
promedio	-12,541	0,001167

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

Docimasia de hipótesis de la correlación media

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

$$H_0: \rho = 0$$

- Hipótesis alterna (H_1): La correlación entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad de las pruebas pedagógicas de ejecución máxima es significativa.

$$H_1: \rho \neq 0$$

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

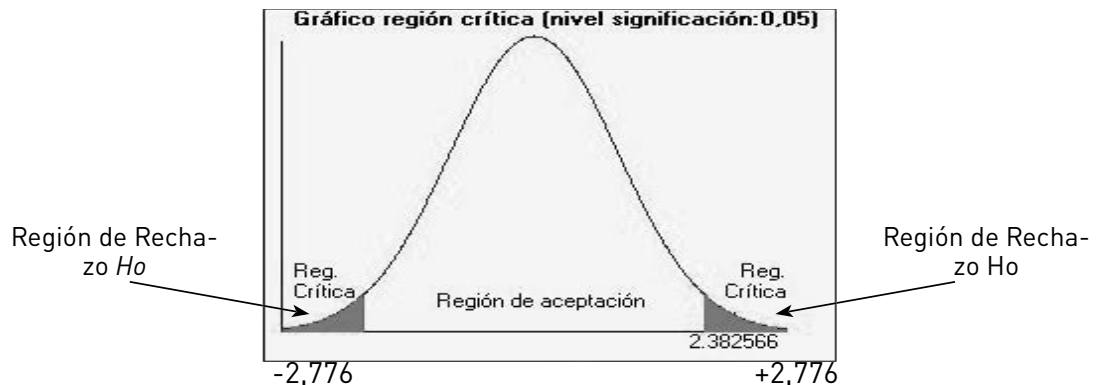
$$gl = 6 - 2 = 4$$

- 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 para los seis instrumentos $r = -0,157$.

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

$$t = \frac{-0,157}{\sqrt{\frac{1-0,050}{4}}} = \frac{-0,157}{0,4873} = -0,3222$$

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



- Decisión: Puesto que la t calculada está ubicada en la región de 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 exigencia y confiabilidad de las pruebas pedagógicas de ejecución máxima no es significativa.

Discusión de resultados

Para poder responder a la interrogante de la investigación: ¿Qué relación existe entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad, de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín?

En primer lugar hay que poner de manifiesto que el nivel de exigencia está explicado mediante el estadígrafo asimetría, para cada uno de los ítems de las pruebas así como para el puntaje total, en los que se tuvo en cuenta el rango de valores: los positivos –indican dificultad en diferentes grados–, cero –indica dificultad intermedia–, y los negativos– indican facilidad en diferentes grados–; tal como lo prescribe Shiefelbein (1999:286).

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

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

El valor del coeficiente de correlación de Pearson promedio $r = -0,157$, entre los estadígrafos psicométricos de nivel de exigencia y confiabilidad es bajo y no significativo a un nivel de 0,01, que por ser negativo se dice que existe una relación inversa entre los estadígrafos; es decir, a mayor valor del nivel de exigencia menor valor de la confiabilidad, de las pruebas y viceversa. Esta forma de interpretar, es estrictamente matemática, y no estaría revelando relaciones de causa–efecto, como demarca el espíritu de la investigación.

Que existe relación entre el nivel de exigencia y la confiabilidad, es un hecho; por eso se hizo necesario calcular el coeficiente de determinación promedio, cuyo valor es $r^2 = 0,050$; lo que quiere decir es que, en promedio el 5 % del valor del nivel de exigencia de las pruebas está determinado por el valor de la confiabilidad, y el 95% se debe a otros factores. Por cierto que para emitir cualquier juicio valorativo sobre lo expuesto, habría que analizar no solo seis pruebas pedagógicas, como es el caso de la presente investigación, sino muchas (muchísimas más); no perder de vista que a la presente investigación es exploratoria y sería

un punto de partida para otras investigaciones que se pudieran realizar.

Observando que, los resultados de los coeficientes de correlación de Pearson--a excepción de la prueba dicotómica N°02-- son bajas; se puede aseverar que la causalidad buscada entre el nivel de exigencia y la confiabilidad es mínima; pero si esta aseveración se generalizara, de manera contundente, para todas las pruebas pedagógicas de ejecución máxima que pudieran haber; y, por cierto que las hay; sería riesgosa y temeraria, porque la muestra para la presente investigación no es tan vasta; es más, no han sido tomada de manera totalmente aleatoria, sino intencional.

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

Así también, contrastando los resultados del presente trabajo con los obtenidos por Cárdenas (2008, 2010, 2014 y 2015), 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 de los ítems de las pruebas pedagógicas que se aplican a los alumnos de las instituciones educativas de la Región Junín.

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

Conclusiones

- El nivel de exigencia de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, en términos cualitativos, fluctuaron entre la facilidad--dificultad intermedia y la facilidad.
- El estadígrafo psicométrico de confiabilidad de los ítems de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, fluctuaron entre los valores 0,618 y 0,833; en términos cualitativos estuvieron entre confiables y excelentemente confiables.
- La correlación es baja, inversa y con poca significatividad; el coeficiente de determinación pone de manifiesto que sólo el 5% del nivel de exigencia de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, está determinado por el valor de la confiabilidad y el 95% se debe a otros factores
- El análisis de regresión lineal, de las pruebas pedagógicas de ejecución máxima que se aplican a los alumnos de las instituciones educativas de la Región Junín, pone de manifiesto que el grado de dependencia del nivel de exigencia con respecto de la confiabilidad de las pruebas es mayor que el grado de influencia de la confiabilidad en el nivel de exigencia de las respectivas pruebas.

Referencias

- Aiken, L. (1995). *Test psicológicos y evaluación*. México D.F., México: Editorial. Prentice Hall Hispanoamericana.
- Ary; Jacobs y Razavieh. (1992). *Introducción a la investigación pedagógica*. México. D.F., México: Editorial McGraw-Hill.
- Cárdenas Ayala, A. (2007). *Pertinencia de los Instrumentos de recolección de datos con los métodos estadísticos de confiabilidad*. (I.I. FPH-UNCP, trabajo de investigación inédito). Huancayo, Perú.
- 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. (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. (2012). *Análisis estadístico del índice de dificultad de los ítems politómicos en las pruebas pedagógicas*. Horizonte de la ciencia, 02, 74-81. Huancayo, Perú.
- Cárdenas Ayala, A. (2013). *Índice de dificultad y asimetría de los ítems en las pruebas pedagógicas*. I.I.FPH-UNCP, trabajo de investigación inédito,. Huancayo, Perú.
- Cárdenas Ayala, A. (2014). *Confiabilidad y validez de las pruebas pedagógicas de ejecución máxima*. I.I.FPH-UNCP, trabajo de investigación inédito,. 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.
- Davis, F. (1989). *Analyse des items*. Lovaina, París: Nauwelaerts.
- Delgado; Escurra y Torres. (2006). *La medición en psicología y educación*. Lima, Perú: Editorial Hozlo S.R.L.
- Fermín, M. (1995). *La evaluación, los exámenes y las calificaciones*. Buenos Aires, Argentina: Editorial Kapeluz.
- Gonzalvo, G. (1988). *Diccionario de metodología estadística*. Madrid, España: Ediciones Morata S.A.
- Hambleton y Novik. (1989). *Toward an integration of theory and method for criterion referenced test*. New York, USA: Editorial Measurmet.
- Hernández; Fernández y Baptista. (2003). *Metodología de la investigación*. México: Editorial 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: Empresa Editora MACRO.
- Sánchez y Reyes. (2006). *Metodología y Diseños en la Investigación Científica*. Lima, Perú: Editorial Visión Universitaria.

SECCIÓN I: *Artículos Científicos*

Schiefelbein, E. (1999). *Teoría, técnicas, procesos y casos en el planeamiento de la educación*. Buenos Aires, Argentina: Editorial Ateneo.

Santisteban, C. (2009). *Principios de psicometría*. Madrid, España: Editorial Síntesis.

Tavella, N. (1991). *Análisis de los ítemes en la construcción de instrumentos Psicométricos*. México: Editorial Trillas.

Tembrick, T. (1981). *Evaluación*. Madrid, España: Editorial Narcea.

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

Visauta y Martori i Cañas. (2003). *Análisis estadístico con SPSS para windows*, (Volumen II). Madrid, España: Editorial McGraw-Hill.