



EFFECTO DE LA FORMA DE LA DISTRIBUCIÓN Y DE LA MEDIA EN EL INDICE p_0 DE HUYNH

[Effect of the shape of distribution and mean in Huynh's p_0 index]

por

[Article record](#)

[About authors](#)

[HTML format](#)

Gonzalo Almerich (Gonzalo.Almerich@uv.es)

Rosa M. Bo Bonet (Rosa.Bo@uv.es)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

In the field of Educational Evaluation, the Criterion-Referenced Evaluation is a very relevant, though incipient approach. In this paper, we present a simulation study oriented to the analysis of a reliability index: the index p_0 calculated by means of the Huynh's method for criterion-referenced tests. The simulation study has been carried out with the aid of computer software developed *ad hoc*. The aim of this paper is to describe the influence of two variables – the distribution shape and the mean on that index. The conclusion is that both variables consistently affect the reliability index. Furthermore, the obtained evidence brings relevant information about desirable metrical characteristics for applied use of the index in the criterion-referenced tests.

Keywords

Evaluation, Assessment, criterion-referenced evaluation, reliability, criterion-referenced reliability, classification consistency indexes, p_0 index, Huynh's method, mean, shape of distribution.

Resumen

En el ámbito de la Evaluación Educativa, la Evaluación Criterial es una corriente muy importante, aunque incipiente. En este artículo, presentamos un estudio de simulación cuyo propósito es el análisis de un índice de fiabilidad: el índice p_0 de Huynh para pruebas de referencia criterial. El estudio de simulación se ha realizado con la ayuda de un programa de ordenador creado para tal fin. El objetivo es describir la influencia de dos variables –la forma de la distribución y la media– en este índice. La conclusión es que ambas variables están afectando consistentemente el índice de fiabilidad. Además, las evidencias obtenidas aportan información relevante sobre características métricas deseables para la posterior utilización del índice en las pruebas criterioles.

Descriptores

Evaluación, Evaluación Criterial, Fiabilidad, Fiabilidad en pruebas criterioles, índices de consistencia de la clasificación, índice p_0 , método de Huynh, media, forma de la distribución.

1.- Introducción

En la actualidad, la Evaluación Educativa es un área de especial importancia en la Educación, ya que nos permite mejorar la calidad

en la Educación. Ello significa la necesidad de realizar evaluaciones en todos los ámbitos del Sistema Educativo, y también extender la realización de evaluaciones a aquellos ámbitos y acciones formativas fuera del Sistema

Educativo, como por ejemplo los cursos formativos de Formación Profesional, tanto Ocupacional como Continua.

Como forma de respuesta a los planteamientos de la actual concepción de la Evaluación Educativa una escuela de gran implantación es la Evaluación de Referencia Criterial. El nacimiento del término Referido al Criterio (Jornet y Suárez, 1994) aparece por primera vez en el artículo de Glaser y Klaus (1962), aclarando posteriormente este concepto Glaser (1963). El término se consolida a partir del artículo de Popham y Huseck (1969), produciéndose un desarrollo importante de esta escuela entre los años 70 y 80 del siglo XX.

Dentro de la Evaluación Criterial uno de los primeros problemas que surgieron fue la fiabilidad y su concepción. Popham y Huseck (1969), y posteriormente Hambleton y Novick (1973), indicaron que no eran adecuados los enfoques clásicos de la fiabilidad y, por consiguiente, los indicadores de fiabilidad clásicos no deben ser utilizados para las pruebas de referencia criterial dadas las características distintivas de estas pruebas, de manera que la fiabilidad en pruebas de referencia criterial debía de formularse de modo distinto a la fiabilidad existente para las pruebas normativas. Ello supuso que en la década de los años 70 diversos autores (Hambleton y Novick, 1973; Swaminathan, Hambleton y Algina, 1974; Livingston, 1972; Brennan y Kane, 1977) propusieran diversos indicadores para estimar la fiabilidad en pruebas de referencia criterial (Berk, 1984; Subkoviak, 1984).

1.1.- El índice p_o . El método de cálculo de Huynh

Hambleton y Novick (1973) propusieron un indicador para el cálculo de la fiabilidad en pruebas criterioales, el coeficiente de acuerdo p_o , que representa la consistencia de la toma de la decisión. Este índice nos indica la proporción de sujetos consistentemente clasificados como aptos/no-aptos, domi-

na/no-domina, etc., a través de dos formas de tests paralelas o dos administraciones del mismo test. La ventaja de dicho indicador es su fácil cálculo e interpretación. Sin embargo, una desventaja del mismo es que el método de cálculo que propusieron estos dos autores requería la administración de dos pruebas. Por consiguiente, surgió la necesidad de definir métodos de cálculo de este índice que requiriesen una sola administración de la prueba. Así, surgieron tres métodos de cálculo de este índice que requieren la administración de una única prueba. Estos tres métodos son los propuestos por Marshall y Haertel (1976), Huynh (1976) y Subkoviak (1976).

En 1976, Huynh propone un método para estimar el coeficiente p_o y también el coeficiente kappa. Este es un método matemáticamente elegante, que se basa en el modelo beta-binomial de Keats y Lord (1962), que parte de dos asunciones básicas: las distribuciones verdaderas de los estudiantes en el test se distribuyen como una distribución beta, y que en un test de n-ítems, si se administra repetidamente la distribución de sus puntuaciones, tomará la forma de una distribución binomial.

En relación con el cálculo de este índice mediante el método de Huynh, debido a su complejidad, se han propuesto métodos más sencillos que permiten obtener el valor del índice, como son los propuestos por el mismo Huynh (1976), Peng y Subkoviak (1980) y Subkoviak (1988), todos ellos basados en tablas.

Este método de cálculo del índice p_o , entre sus ventajas presenta que requiere una sola administración para la estimación del índice, produce pequeños errores estándar para muestras de tamaño del aula (Subkoviak, 1978) y la violación de la asunción de dificultad del ítem parece tener un efecto insignificante sobre las estimaciones (Huynh & Saunders, 1980; Subkoviak, 1978). Entre sus desventajas se puede apuntar que su aproxi-

mación conceptual y matemática es compleja, así como su cálculo es demasiado complicado para realizarlo a mano, y produce estimaciones conservadoras del índice para tests cortos (Huynh & Saunders, 1980; Subkoviak, 1978).

Una de las características de los estudios acerca del funcionamiento de éste y otros tipos de indicadores de fiabilidad en pruebas de referencia criterial es que la investigación se ha dirigido principalmente al estudio de aspectos teóricos, como el ajuste del modelo teórico a los datos observados, con la consiguiente adecuación o no del índice a partir del modelo estudiado, planteamiento de soluciones cuando el modelo no se ajusta, etc. (Almerich, 2000). Por lo tanto, es necesario abordar la influencia que diferentes factores (la media, la puntuación de corte, la forma de la distribución, el número de ítems, etc.) y su combinación ejercen en los distintos indicadores de fiabilidad, como por ejemplo el p_0 obtenido con los métodos de Huynh (1976) y Subkoviak (1976), el índice $F(1)$ de Brennan (1980) y el índice K^2 de Livingston (1972).

En este artículo se presenta un estudio del índice p_0 para pruebas de referencia criterial obtenido mediante el método de Huynh (1976), cuyo objetivo es comprobar el efecto que sobre dicho índice ejercen dos factores, la media y la forma de la distribución, definida por la asimetría y la curtosis.

2.- Metodología

A partir de las características propias de esta investigación, se ha utilizado un método reciente en el ámbito educativo: la simulación. La simulación se constituye en una herramienta útil y sencilla que permite poder observar el funcionamiento de las técnicas o herramientas objeto de estudio a partir de todas y cada una de las condiciones experimentales que establece el investigador, con la finalidad de poder ofrecer conclusiones acerca del mismo. Su objetivo sería la generación de conjuntos de datos, según el modelo predeterminado por el investigador, lo cual

implica la utilización de la informática para generar esos conjuntos de datos, a partir de los cuales se analiza el funcionamiento de las técnicas o herramientas.

Este tipo de herramienta, la simulación, ha sido empleada anteriormente en algunos estudios relacionados con la fiabilidad en pruebas de referencia criterial, como los estudios de Algina y Noe (1978), Subkoviak (1978) y Spray & Welch (1990), indicando que a partir de éstos las principales condiciones de la investigación que se han tenido en cuenta en estos estudios sobre fiabilidad en pruebas de referencia criterial son la puntuación de corte, el número de sujetos y la longitud del test.

En este caso, para nuestra investigación necesitábamos, por una parte, generar matrices de respuestas de los sujetos a partir de las condiciones experimentales que cumpliesen la característica de producir respuestas dicotómicas, es decir, matrices de respuestas de sujetos formadas por unos y ceros, y por otra parte, el cálculo de los índices de fiabilidad en pruebas de referencia criterial. Para tal fin se hizo imprescindible un programa de ordenador para llevar a cabo nuestra investigación, y en vista de la no disponibilidad de un programa semejante, se diseñó mediante el lenguaje de programación Visual Basic dicho programa de ordenador, COMAFI, que, por un lado, nos permite generar matrices de respuestas de los sujetos, de ceros y unos, y, por otro, calcular éste y otros índices de fiabilidad, junto con otros estadísticos (media, varianza, etc.).

Las condiciones experimentales, junto con sus niveles, que se han tenido en cuenta en la investigación son: la forma de la distribución, definida por la asimetría, cuyos niveles son: -1,5, -0,75, 0, 0,75, 1,5; y la curtosis, con los siguientes niveles: -1,5, -0,75, 0, 0,75, 1,5; la puntuación de corte, con los niveles: 40%, 50%, 60%, 70% y 80%; la media, con los siguientes niveles: 40%, 50%, 60%, 70% y 80%; el número de sujetos, con

tres niveles: 30, 120, 1000; el número de ítems, también con tres niveles: 10, 20, 40; y finalmente el índice clásico de fiabilidad KR-20 y el KR-21.

Las distintas condiciones de experimentación han supuesto los distintos casos, mediante el cruce de las distintas condiciones de experimentación y sus niveles (5x5x5x5x3x3). Para cada uno de los casos la muestra debía tener un tamaño de 500 matrices con la condición de que no estuviesen repetidas, es decir, diferentes entre sí, si bien se ha de indicar que en algunos casos no ha sido posible obtener ninguna matriz y en algunos otros no ha sido posible obtener las 500 matrices, por la complejidad de la situación. Para generar tales matrices se estableció que para cada combinación se produjeran como mínimo 30.000 replicaciones para cada caso.

Una vez obtenidos los resultados de la simulación se realizaron los distintos análisis estadísticos mediante el programa SPSS 12.0.

Como hemos indicado anteriormente, éste artículo se centra en la revisión de los resultados más relevantes obtenidos con el índice p_o , calculado mediante el método de Huynh

(1976), respecto del efecto producido por las variables media y la forma de la distribución, teniendo en cuenta los niveles respectivamente indicados.

3.- Resultados

Esta sección la dividiremos en tres apartados. En el primero, indicaremos el funcionamiento del índice respecto de la media, tratado de una forma global. En el segundo apartado, realizaremos la comparación de las puntuaciones totales obtenidas en cada una de las formas de la distribución, para en el tercer apartado realizar un análisis conjunto del funcionamiento del índice en relación con la posición de la media y la forma de la distribución.

3.1.- Efecto de la media

En este apartado describiremos el efecto de la media, que corresponde a la media de las puntuaciones de los sujetos en una prueba determinada. En este caso expresaremos la media en forma de porcentaje, y que representa el tanto por ciento de la puntuación total de la prueba. En esta descripción tendremos en cuenta los datos globales del indicador p_o , obtenido con el método de Huynh.

Tabla 1.- Resultados globales de los porcentajes de la media de la prueba

	Media	Des. Típica
40%	0,8376	0,0873
50%	0,8150	0,0753
60%	0,8104	0,0845
70%	0,8406	0,1027
80%	0,8865	0,1226
Total	0,8317	0,0941

La media total del índice es 0,8317, presentando poca variabilidad, lo que indica la estabilidad del índice.

Si consideramos los valores del índice p_o , estos se sitúan en un rango que tiene como

valor mínimo 0,8104, en la media 60%, y valor máximo 0,8865, en la media 80%. Esto nos describe una curva en forma de U, como podemos observar en el gráfico 1, y que esta curva se ajusta a una función cuadrática.

Tabla 2.- Estimación de la curva descrita por el valor del índice p_0 en cada una de las medias

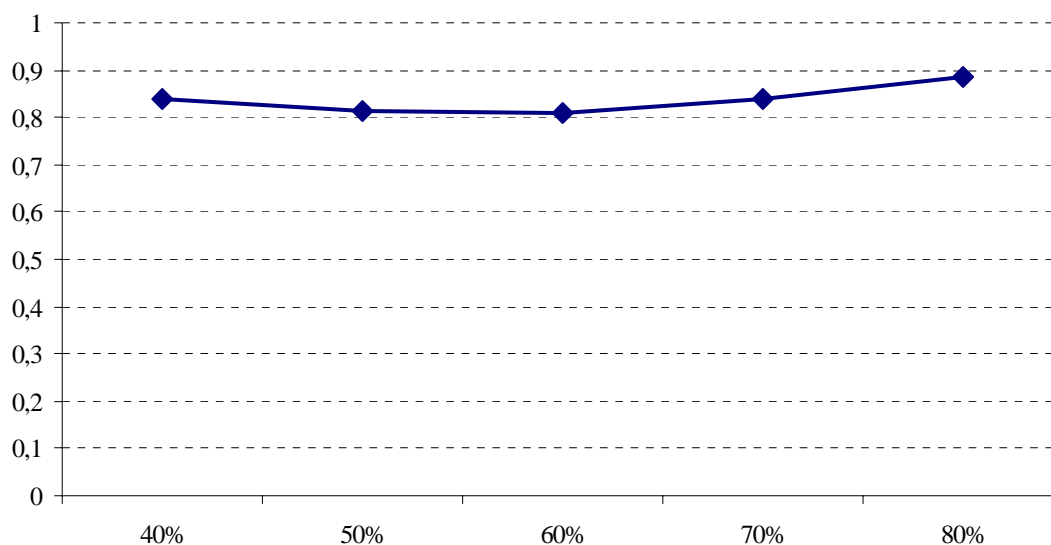
	Curva	R ²	Significación
Media total	Cuadrática	0,99	,005

Si comparamos las medias totales en cada porcentaje con la media total del índice, podemos indicar que las medias son mayores a la media total en el porcentaje 40%, 70% y 80%, y es menor en los otros porcentajes, 50% y 60%.

En relación con la variabilidad apuntaremos que ésta es menor que la desviación típica total en los porcentajes 40%, 50% y 60%, y mayor en los porcentajes 70% y 80%.

Si consideramos las medias en los distintos porcentajes señalaremos que el valor menor se produce en el porcentaje 60%, aumentando a medida que nos acercamos a los extremos, siendo mayor en los porcentajes 70% y 80% que en los porcentajes 40% y 50%. En relación con la variabilidad, podemos observar que la menor variabilidad corresponde al porcentaje 50%, siendo un poco mayor en el 40% y 60%, aumentando a medida que aumentamos el porcentaje de la puntuación.

Gráfico 1.- Medias globales del índice p_0 , calculado mediante el método de Huynh, teniendo en cuenta el porcentaje de la media



Por lo tanto, se puede observar que el valor del índice p_0 se incrementa al distanciarse la media de los porcentajes centrales de la puntuación, situándose los valores menores en los porcentajes medios, así como indicar que los porcentajes centrales implican una mayor estabilidad del índice que los porcentajes más extremos.

3.2.- Efecto de la forma de la distribución

En este apartado indicaremos el funcionamiento general que se produce en el índice dependiendo de la forma de la distribución. Para ello, en una primera parte indicaremos el funcionamiento del valor p_0 teniendo en cuenta como funcionan la asimetría y la curtosis de una forma global. En la segunda

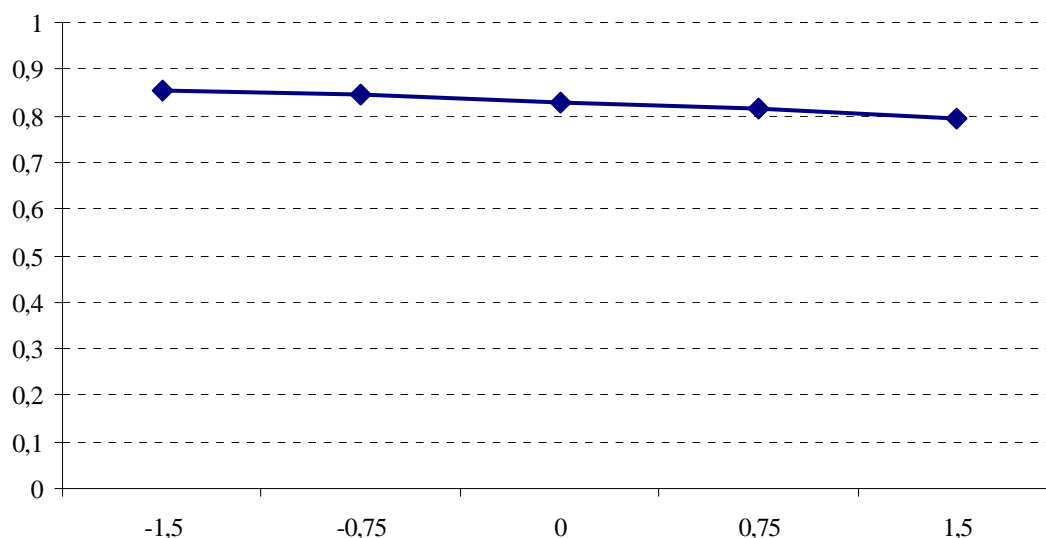
parte, haremos referencia al funcionamiento del índice p_0 en las distintas distribuciones, pero sólo teniendo en cuenta el valor total del índice p_0 en cada distribución.

En primer lugar, indicaremos el comportamiento general de la asimetría y la curtosis, teniendo en cuenta todos los casos de la investigación.

Tabla 3.- Valores totales asimetría

Asimetría	Media	Des. Típica
1,5	0,7919	0,1273
0,75	0,8164	0,1070
0,00	0,8288	0,0912
-0,75	0,8453	0,0769
-1,5	0,8557	0,0730

Gráfico 2.- Medias globales del índice p_0 , calculado mediante el método de Huynh, teniendo en cuenta la asimetría



A partir de la tabla 2 y el gráfico 2 podemos concluir, por una parte, que los valores totales del índice con asimetría negativa son mayores que los que tienen asimetría positiva. Por otra parte, podemos indicar que a

medida que los valores de la asimetría pasan de valores positivos a valores negativos el índice p_0 aumenta su valor, describiendo una curva que se ajusta mejor a una función lineal.

Tabla 4.- Estimación de la curva descrita por el valor del índice p_0 en cada uno de los niveles de asimetría

	Curva	R ²	Significación
Asimetría	Lineal	0,98	,001

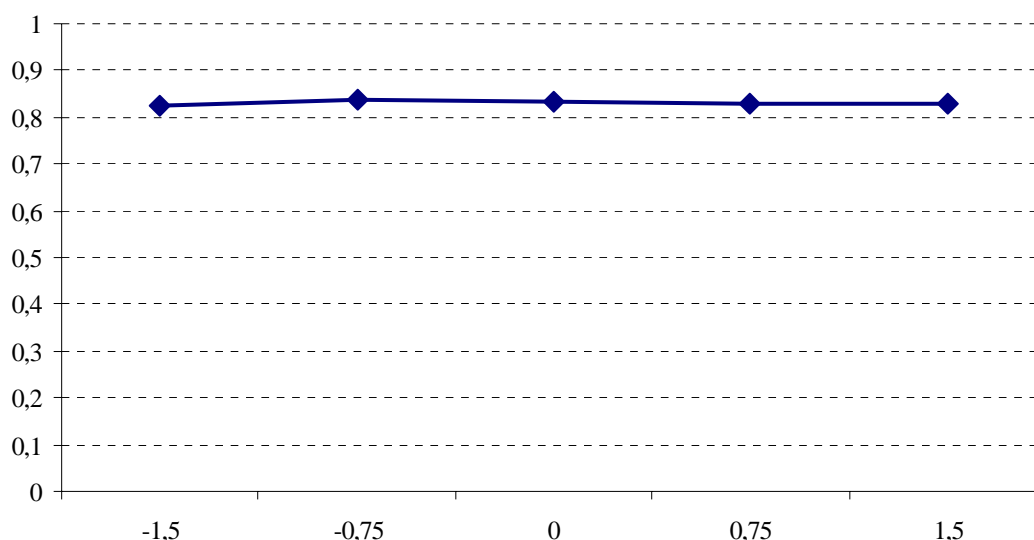
En relación con la curtosis, de forma general, apuntaremos que existe una cierta tendencia a aumentar el valor del índice p_0 , al pasar la curtosis de un valor positivo a un

valor negativo. Sin embargo, en este caso no se ajusta a ninguna curva en concreto, por ejemplo ni lineal ($p > 0,05$) ni cuadrática ($p > 0,05$).

Tabla 5.- Valores totales de la curtosis

Curtosis	Media	Des. Típica
1,5	0,8279	0,1033
0,75	0,8299	0,0975
0,00	0,8349	0,0893
-0,75	0,8388	0,0833
-1,5	0,8261	0,0719

Gráfico 3.- Medias globales del índice p_0 , calculado mediante el método de Huynh, teniendo en cuenta la curtosis



En segundo lugar, como indicábamos anteriormente, compararemos las puntuaciones totales a lo largo de las distintas formas de la distribución (tabla 6).

Los valores del índice p_0 se sitúan en un rango que se sitúan entre un valor máximo igual a 0,8589, en la distribución con asimetría = -1,5 y curtosis = 0,75, y un valor mínimo igual a 0,7578, en la distribución con asimetría 1,5 y curtosis 0.

En la comparación de las puntuaciones totales a lo largo de la forma de la distribución con la puntuación global, cuyo valor es 0,8317, observaremos que las puntuaciones totales en las diferentes distribuciones son mayores a la puntuación total media en las distribuciones definidas por asimetría negativa, sin tener en cuenta la curtosis, salvo en la distribución definida por la asimetría = -

0,75 y curtosis = -1,5, donde el valor del índice $p_0= 0,8061$. Si la asimetría es positiva el valor de p_0 en cada distribución es menor que el valor medio total del índice, y no dependiendo de la curtosis, a excepción de la distribución con asimetría= 0 y curtosis = -0,75, en cuyo caso el valor del índice $p_0=0,8428$.

Si consideramos la asimetría el valor del índice p_0 tiene una tendencia a aumentar a medida que el valor de la asimetría va pasando de valores positivos a valores negativos. Por consiguiente, los valores totales son superiores cuando la asimetría es negativa y, por tanto, son inferiores en la asimetría positiva.

En referencia con la curtosis, el valor del índice p_0 aumenta a medida que el valor de la curtosis disminuye, a excepción de las distribuciones con asimetría igual a 1,5, ya que

cuando decrece la curtosis disminuye el índice p_0 . Por lo tanto, en las distribuciones leptocúrticas el índice p_0 es menor que en las distribuciones platicúrticas.

En relación con la variabilidad también podemos observar que si la asimetría es positiva la variabilidad es mayor que cuando la asimetría es menor. Si tomamos como referencia el valor de la variabilidad global, el valor de cada una de las desviaciones típicas es mayor que esta cuando la asimetría es positiva, mientras que cuando la asimetría es negativa en todos los casos la variabilidad es menor que la global. Cuando la asimetría es 0, en este caso cuando la curtosis es positiva el valor de la desviación típica es mayor que la global, mientras que si es negativa el valor es menor que el valor global. Así podemos indicar que a medida que la asimetría pasa de valores positivos a valores negativos la variabilidad tiende a disminuir, asimismo podemos indicar que a medida que la curtosis disminuye se produce un decrecimiento de la variabilidad.

Como conclusión podemos indicar que el valor de las puntuaciones globales aumenta a medida que el valor de la asimetría pasa de valores positivos a valores negativos, así cuando la curtosis pasa de valores positivos a valores negativos. Asimismo podemos indicar que también es más estable, ya que la variabilidad tiende a ser menor.

3.3.- Efecto de la media y de la forma de la distribución

En este apartado en primer lugar describiremos el funcionamiento del índice teniendo en cuenta la media conjuntamente con la forma de la distribución, que recordemos que estaba definida por la asimetría y la curtosis, y en segundo lugar compararemos cada caso con el caso total.

3.3.1.- Descripción del funcionamiento del índice en función de la media y la distribución de la forma

En primer lugar, se puede indicar que si la asimetría es positiva o igual a cero y la curtosis positiva, y en alguna situación cuando es igual a cero, en este caso el valor del índice p_0 sigue el patrón general descrito anteriormente, es decir, el valor menor del índice p_0 se sitúa cuando el porcentaje de la media es el 60% aumentando a medida que nos alejamos hacia los extremos, siendo mayor el valor del índice p_0 en los porcentajes mayores del 60%. Si la asimetría es negativa y la curtosis positiva, el patrón descrito anteriormente no se produce exactamente, pues el valor menor del índice p_0 se sitúa cuando el porcentaje es el 50%, aumentando a medida que nos alejamos de ese porcentaje, y al igual que anteriormente es mayor el valor del índice p_0 en los porcentajes altos que en el 40%. En los otros casos, no se puede indicar la existencia de un patrón concreto para todos los casos.

Por lo tanto, cuando la curtosis es positiva y la asimetría positiva el valor menor se encuentra en el porcentaje del 60%, mientras que en la curtosis positiva y asimetría negativa el valor menor se encuentra en el 50%. En la curtosis negativa no existe un patrón bien definido, si bien podemos indicar que en algunas situaciones el valor menor se encuentra en el porcentaje del 40%. En todas las situaciones, según nos alejamos hacia los extremos el porcentaje se produce un incremento en el valor del índice, mayor cuando el porcentaje es el 70% y 80%, que no en el porcentaje del 40% y 50%.

En segundo lugar, se produce un incremento en el valor del índice p_0 a medida que la asimetría toma valores menores y la curtosis también. Así, los valores más altos del índice se producen cuando la asimetría es igual a -1,5, mientras que los valores más bajos se producen cuando la asimetría es 1,5 y la curtosis positiva.

El rango de los valores cuando la media tiene un porcentaje del 40% se sitúa entre $p_0 = 0,7134$, cuando la asimetría es 0,75 y la

curtosis -1,5 y $p_0 = 0,8560$ en la asimetría 0 y la curtosis 1,5. En la mayoría de los casos el valor del índice se sitúa en torno al valor de $p_0 = 0,84$. En el porcentaje de la media del 50% indicar que el rango se sitúa entre $p_0 = 0,7414$, en la asimetría 1,5 y curtosis 0, y $p_0 = 0,8539$, en la asimetría 0 y la curtosis -0,75. En este caso, los valores en las distintas formas de distribución son mayores, respecto del valor total, cuando la asimetría es 0 y negativa, y en todos los casos de la curtosis. Si el porcentaje es igual al 60%, el rango de los valores se encuentra entre un valor mínimo de $p_0 = 0,7085$, en la asimetría 1,5 y curtosis 0, un valor máximo de $p_0 = 0,8691$, en la asimetría -0,75 y curtosis -0,75. Los valores, en este porcentaje, son menores, respecto del valor de referencia total, cuando la asimetría es 1,5 y 0,75, siendo mayores cuando la asimetría es 0 y negativa. En el porcentaje del 70% indicar que la amplitud del rango de las medias se sitúa entre un valor de $p_0 = 0,7573$, en la asimetría 1,5 y curtosis 0, y un

valor de $p_0 = 0,8966$. Respecto del valor de referencia, indicaremos que éste es menor en la asimetría positiva, mientras que éste es mayor en la asimetría negativa. En el porcentaje del 80%, indicar que el valor mínimo del índice $p_0 = 0,8561$, en la asimetría 0,75 y la curtosis -1,5, y el valor mayor del índice $p_0 = 0,9074$, en la asimetría -1,5 y la curtosis 0. En relación con el valor de referencia, indicar que en casi todos los casos se sitúa entre 0,87 y 0,89, siendo superior a 0,90 en la asimetría negativa y la curtosis positiva.

En tercer lugar, y en relación con la variabilidad se puede apuntar que si la asimetría es positiva en todos los casos de la curtosis se produce un aumento de la misma según el porcentaje se incrementa. Sin embargo, cuando la asimetría es negativa la variabilidad es menor en los porcentajes del 50%, 60% y 70%, aumentando en los otros dos porcentajes.

Tabla 6 (múltiple).- Resultados del índice p_0 en las distintas medias en las distintas formas de distribución, definidas por la asimetría y la curtosis.

	Asimetría = 1,5 Curtosis = 1,5		Asimetría = 1,5 Curtosis = 0,75		Asimetría = 1,5 Curtosis = 0	
	Media	Des. Típica	Media	Des. Típica	Media	Des. Típica
40%	0,8325	0,0936	0,8035	0,0829	0,7826	0,0714
50%	0,7706	0,1049	0,7435	0,0916	0,7414	0,0865
60%	0,7509	0,1216	0,7359	0,1119	0,7085	0,0966
70%	0,8178	0,1524	0,8030	0,1539	0,7573	0,1659
80%	0,8917	0,1673	0,8894	0,1591	0,8741	0,1698
Total	0,8017	0,1265	0,7820	0,1239	0,7578	0,1444

	Asimetría = 0,75 Curtosis = 1,5		Asimetría = 0,75 Curtosis = 0,75		Asimetría = 0,75 Curtosis = 0	
	Media	Des. Típica	Media	Des. Típica	Media	Des. Típica
40%	0,8456	0,0905	0,8443	0,0795	0,8475	0,0659
50%	0,7975	0,0958	0,8016	0,0864	0,8062	0,0741
60%	0,7900	0,0982	0,7704	0,1056	0,7675	0,0988
70%	0,8260	0,1302	0,8296	0,1215	0,8068	0,1368
80%	0,8945	0,1533	0,8909	0,1529	0,8884	0,1496
Total	0,8197	0,1080	0,8179	0,1065	0,8191	0,1048

	Asimetría = 0,75 Curtosis = -0,75		Asimetría = 0,75 Curtosis = -1,5	
	Media	Des. Típica	Media	Des. Típica
40%	0,8222	0,0573	0,7134	0,0272
50%	0,7968	0,0611	0,7745	0,0331
60%	0,7614	0,0807	0,7540	0,0554

70%	0,7839	0,1406	0,7791	0,1148
80%	0,8850	0,1441	0,8561	0,1535
Total	0,8061	0,1059	0,8289	0,1455

	Asimetría = 0 Curtosis = 1,5		Asimetría = 0 Curtosis = 0,75		Asimetría = 0 Curtosis = 0	
	Media	Des. Típica	Media	Des. Típica	Media	Des. Típica
40%	0,8560	0,0943	0,8442	0,0902	0,8462	0,0742
50%	0,8047	0,0837	0,8145	0,0735	0,8209	0,0606
60%	0,7857	0,0935	0,7954	0,0800	0,8070	0,0704
70%	0,8232	0,1190	0,8051	0,1284	0,8050	0,1188
80%	0,8884	0,1491	0,8884	0,1445	0,8829	0,1385
Total	0,8192	0,1058	0,8215	0,1012	0,8289	0,0911

	Asimetría = 0 Curtosis = -0,75		Asimetría = 0 Curtosis = -1,5	
	Media	Des. Típica	Media	Des. Típica
40%	0,8441	0,0590	0,8128	0,0586
50%	0,8539	0,0475	0,8384	0,0464
60%	0,8274	0,0540	0,8229	0,0409
70%	0,8146	0,0949	0,8204	0,0804
80%	0,8766	0,1296	0,8706	0,1142
Total	0,8428	0,0793	0,8301	0,0697

	Asimetría = -0,75 Curtosis = 1,5		Asimetría = -0,75 Curtosis = 0,75		Asimetría = -0,75 Curtosis = 0	
	Media	Des. Típica	Media	Des. Típica	Media	Des. Típica
40%	0,8405	0,1163	0,8401	0,1033	0,8442	0,0876
50%	0,8173	0,0752	0,8218	0,0674	0,8359	0,0570
60%	0,8174	0,0652	0,8332	0,0559	0,8555	0,0455
70%	0,8376	0,0881	0,8496	0,0693	0,8641	0,0524
80%	0,8777	0,1191	0,8832	0,1022	0,8853	0,0892
Total	0,8335	0,0934	0,8426	0,0821	0,8531	0,0681

	Asimetría = -0,75 Curtosis = -0,75		Asimetría = -0,75 Curtosis = -1,5	
	Media	Des. Típica	Media	Des. Típica
40%	0,8387	0,0727	0,7814	0,0388
50%	0,8510	0,0479	0,8075	0,0353
60%	0,8691	0,0381	0,8246	0,0311
70%	0,8653	0,0450	0,8583	0,0327
80%	0,8847	0,0758	0,8782	0,0617
Total	0,8586	0,0576	0,8061	0,0402

	Asimetría = -1,5 Curtosis = 1,5		Asimetría = -1,5 Curtosis = 0,75		Asimetría = -1,5 Curtosis = 0	
	Media	Des. Típica	Media	Des. Típica	Media	Des. Típica
40%	0,8354	0,1230	0,8357	0,1079	0,8477	0,0892
50%	0,8112	0,0746	0,8228	0,0638	0,8157	0,0489
60%	0,8389	0,0518	0,8528	0,0442	0,8552	0,0371
70%	0,8829	0,0424	0,8966	0,0329	0,8884	0,0256

80%	0,9071	0,0421	0,9064	0,0341	0,9074	0,0249
Total	0,8539	0,0791	0,8589	0,0707	0,8532	0,0591

3.3.2.- Comparación entre las media totales del índice p_0 y las medias obtenidas en cada forma de la distribución.

En este punto compararemos la curva descrita por las medias en cada uno de los casos producidos por la asimetría y la curtosis con el caso general en el que se han tenido en cuenta todos los casos.

Como indicábamos anteriormente, la curva descrita por las medias totales describe una función cuadrática, y si la comparamos con las curvas descritas (tabla 7) en todas las situaciones producidas por la forma de la distribución señalaremos, en general, que en los casos en los cuales la asimetría es positi-

va o igual a cero y con curtosis positiva la curva descrita se ajusta a una función cuadrática; si la asimetría es negativa y la curtosis positiva la curva descrita se ajusta a una función cúbica, a excepción del caso con asimetría -0,75 y curtosis 0,75 que en este caso es cuadrática; y en los casos en que la curtosis es negativa no existe un patrón concreto, pues las curvas descritas se ajustan a funciones lineales, cuadráticas y cúbicas. Por lo tanto, todo ello corrobora lo que anteriormente se había apuntado, en relación con el patrón descrito por cada uno de los casos y su comparación con el patrón general.

Tabla 7.- Estimación de la curva en función de los valores del índice p_0 como consecuencia de la media según la forma de la distribución.

	Curva	R ²	Significación
Asimetría = 1,5 Curtosis = 1,5	Cuadrática	0,97	,028
Asimetría = 1,5 Curtosis = 0,75	Cuadrática	0,98	,013
Asimetría = 1,5 Curtosis = 0	Cuadrática	0,97	,030
Asimetría = 0,75 Curtosis = 1,5	Cuadrática	0,99	,001
Asimetría = 0,75 Curtosis = 0,75	Cuadrática	0,95	,049
Asimetría = 0,75 Curtosis = 0	Cuadrática	0,96	,039
Asimetría = 0,75 Curtosis = -0,75	Cuadrática	0,91	,090
Asimetría = 0,75 Curtosis = -1,5	Lineal	0,78	,048
Asimetría = 0 Curtosis = 1,5	Cuadrática	0,99	,006
Asimetría = 0 Curtosis = 0,75	Cuadrática	0,93	,072
Asimetría = 0 Curtosis = 0	Cuadrática	0,87	,125
Asimetría = 0 Curtosis = -0,75	Cúbica	0,99	,035
Asimetría = 0 Curtosis = -1,5	Cúbica	0,99	,048
Asimetría = -0,75 Curtosis = 1,5	Cúbico	1,00	,009
Asimetría = -0,75 Curtosis = 0,75	Cuadrática	0,98	,016
Asimetría = -0,75 Curtosis = 0	Lineal	0,83	,030
Asimetría = -0,75 Curtosis = -0,75	Lineal	0,91	,012
Asimetría = -0,75 Curtosis = -1,5	Lineal	0,99	,000
Asimetría = -1,5 Curtosis = 1,5	Cúbica	1,00	,000
Asimetría = -1,5 Curtosis = 0,75	Cúbica	0,99	,038
Asimetría = -1,5 Curtosis = 0	Lineal	0,77	0,07

4.- Conclusiones

Los resultados obtenidos en el estudio indican, en primer lugar, que los distintos valores de la media de la prueba describen una curva en forma de U, en la cual el valor menor del índice p_0 , obtenido con el método de Huynh, se produce cuando el porcentaje de

la media es el 60%, mientras que los valores más altos se producen cuando el porcentaje de la media de la prueba se sitúa en el 70% y 80%. Es decir, aquellos porcentajes de la media que se sitúan en las puntuaciones medias del test tendrán unos valores menores que en las puntuaciones extremas, si bien

estas últimas presentan menor estabilidad en el índice.

Además se ha de indicar que se incrementa el valor medio del índice p_0 si pasamos de una asimetría positiva a una asimetría negativa. En referencia a la curtosis indicaremos que el valor medio del índice aumenta si la curtosis pasa de valores positivos a valores negativos. Si consideramos conjuntamente la asimetría y la curtosis, en general se producen valores más altos a medida que la asimetría y la curtosis son negativas.

Si tenemos en cuenta las formas de la distribución y el porcentaje de la media, se produce una curva en forma de U, si bien no en todos los casos se sigue el patrón general. De esta forma, el valor menor del índice p_0 cambia en relación con el porcentaje, de modo que se sitúa en el porcentaje de la media del 50% cuando la curtosis es positiva y la asimetría negativa, y en el porcentaje del 60% cuando la curtosis es positiva y la asimetría positiva. Finalmente indicaremos que los valores mayores del índice se producen en el porcentaje de la media del 80%. En conjunto, como indicábamos anteriormente, aumenta el valor del índice en todos los porcentajes de la media al disminuir tanto la asimetría como la curtosis, en aquellos casos que se han contemplado.

En síntesis, este tipo de estudios pueden dar soporte para las necesidades apuntadas por Hambleton y Slater (1997) respecto a la fiabilidad y características métricas deseables en los procesos de acreditación cuya extensión y consolidación aún no se han completado en nuestro contexto, lo que implica, en relación con la investigación futura y dada la importancia de la evaluación criterial, la necesidad de realizar más estudios sobre el funcionamiento de este tipo de indicadores de fiabilidad en pruebas de referencia criterial que nos proporcionen conclusiones más potentes y estables, teniendo en cuenta estos y otros factores (puntuación de corte, número de ítems, etc.) que nos orien-

ten en su aplicación, además de otros estudios de tipo más teórico (Lee, Hanson and Brennan, 2002).

5.- Bibliografía

- Algina, J. & Noe, M. J. (1978). A Study of Accuracy of Subkoviak's Single-Administration Estimate of the Coefficient of Agreement Using two True-Score Estimates. *Journal of Educational Measurement*. Vol. 15, nº 2, pp. 101-110.
- Almerich, G. (2000). *Indicadores de fiabilidad en pruebas de referencia criterial*. Tesis doctoral. Universidad de Valencia. No publicada.
- Berk, R.A. (1984). Selecting the Index of Reliability. En Berk, R. A. (Ed.) *A Guide to Criterion-Referenced Test Construction*. The John Hopkins University Press. Baltimore. Pp. 231-266.
- Brennan, R. L. (1980). Applications of Generalizability Theory. En Berk, R. A. (Ed.) *Criterion-Referenced Measurement: The State of the art*. Baltimore. John Hopkins University Press. Pp. 186-232.
- Brennan, R. L. & Kane, M. T. (1977). An Index of Dependability for Mastery Test. *Journal of Educational Measurement*, vol. 14, nº. 3, pp. 277-289.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, pp. 519-521.
- Glaser, R. & Klaus, D.J. (1962). Proficiency measurement: Assessing human performance. En Gagne, R. M. (Ed.) *Psychological Principles in systems development*. New York. Holt, Rinehart and Winston. Pp. 419-474.
- Hambleton, R.K. & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, vol. 10, pp. 159-171.
- Hambleton, R.K. & Slater, S.C. (1997). Reliability of credentialing examinations on the impact of scoring models and standard-

- setting pòlicies. *Applied Measurement in Education*, vol. 10, N° 1, pp. 19-38.
- Huynh, H (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement*, vol. 13, pp. 253-264.
- Huynh, H. & Saunders, J. C. (1980). Accuracy of two Procedures for Estimating Reliability of Mastery Tests. *Journal of Educational Measurement*. Vol. 17, n°. 4, Pp. 351-358.
- Jornet, J. M. y Suárez, J.M (1994). Evaluación Referida al criterio. Construcción de un test criterial de clase. En Garcia Hoz, V.(Dir.) *Problemas y métodos de investigación en educación personalizada*. Rialp. Madrid. Pp. 419-443.
- Keats, J. A. & Lord, F. M. (1962). A Theoretical Distribution for Mental Test Scores. *Psychometrika*. Vol 21, n°. 1, Pp. 59-72.
- Lee, W., Hanson, B. A. and Brennan, R.L. (2002). Estimating Consistency and Accuracy Indices for Multiples Classifications. *Applied Psychological Measurement*. Vol. 26, n° 4, Pp. 412-432.
- Livingston, S. A. (1972). Criterion-Referenced Applications of Classical Test Theory. *Journal of Educational Measurement*, vol. 9, n°. 1, pp. 13-26.
- Marshall, J. L. & Haertel, E. H. (1976). *The Mean Split-Half Coefficient of Agreement: A Single Administration Index of Reliability for Mastery Tests*. Manuscrito no publicado. University Of Wisconsin.
- Peng, C-Y & Subkoviak, M. J. (1980). A Note on Huynh's Normal Approximation Procedure for Estimating Criterion-Referenced Reliability. *Journal of Educational Measurement*. Vol. 17, n°. 4, pp. 359-368.
- Popham, W.J. & Husek, T.R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, vol. 6, pp. 1-9.
- Spray, J. A. & Welch, C. J. (1990). Estimation of Classification Consistency when the Probability of a Correct Response Varies. *Journal of Educational Measurement*, vol. 27, n° 1, pp. 15-25.
- Subkoviak, M (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, vol. 13, pp. 265-275.
- Subkoviak, M. J. (1978). Empirical Investigation of Procedures for Estimating Reliability for Mastery Tests. *Journal of Educational Measurement*, vol. 15, n°. 1, pp. 111-116.
- Subkoviak, M.J. (1984). Estimating the Reliability of Mastery-Nonmastery Classifications. En Berk, R. A. (Ed.) *A Guide to Criterion-Referenced Test Construction*. The John Hopkins University Press. Baltimore. Pp. 267-291.
- Subkoviak, M. J. (1988). A Practitioner's Guide to Computation and Interpretation of Reliability Indices for Mastery Tests. *Journal of Educational Measurement*, vol. 25, n° 1, pp. 47-55.
- Swaminathan, H., Hambleton, R. K. & Algina, J. (1974). Reliability of Criterion-Referenced Tests: A Decision-Theoretic Formulation. *Journal of Educational Measurement*. Vol. 11, n°. 1, Pp. 263-267.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Gonzalo Almerich (Gonzalo.Almerich@uv.es). Profesor Asociado del Área de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Valencia. Miembro de la [Unidad de Tecnología Educativa](#). Su dirección postal es: Facultad de Filosofía y CC. de la Educación. Avda. Blasco Ibáñez, 30 46010-Valencia (España).

Rosa Bo (Rosa.Bo@uv.es). Profesora Ayudante del Área de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Valencia. Miembro de la [Unidad de Tecnología Educativa](#). Su dirección postal es: Facultad de Filosofía y CC. de la Educación. Avda. Blasco Ibáñez, 30 46010-Valencia (España).

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Almerich, Gonzalo y Bo Bonet, Rosa M. (2006). Efecto de la forma de la distribución y de la media en el índice p_o de Huynh. <i>Revista ELección de Investigación y EValuación Educativa</i> , v. 12, n. 1. http://www.uv.es/RELIEVE/v11n2/RELIEVEv12n1_6.htm Consultado en (poner fecha).
Title / Título	Efecto de la forma de la distribución y de la media en el índice p_o de Huynh. [<i>Effect of the shape of distribution and mean in Huynh's p_o index</i>]
Authors / Autores	Gonzalo Almerich & Rosa M. Bo Bonet,
Review / Revista	Revista ELección de Investigación y EValuación Educativa (RELIEVE), v. 12, n. 1
ISSN	1134-4032
Publication date / Fecha de publicación	2006 (Reception Date : 2005 September 15; Approval Date : 2006 March 7; Publication Date : 2006 March 9)
Abstract / Resumen	<p><i>In the field of Educational Evaluation, the Criterion-Referenced Evaluation is a very relevant, though incipient approach. In this paper, we present a simulation study oriented to the analysis of a reliability index: the index p_o calculated by means of the Huynh's method for criterion-referenced tests. The simulation study has been carried out with the aid of computer software developed ad hoc. The aim of this paper is to describe the influence of two variables – the distribution shape and the mean- on that index. The conclusion is that both variables consistently affect the reliability index. Furthermore, the obtained evidence brings relevant information about desirable metrical characteristics for applied use of the index in the criterion-referenced tests.</i></p> <p>En el ámbito de la Evaluación Educativa, la Evaluación Criterial es una corriente muy importante, aunque incipiente. En este artículo, presentamos un estudio de simulación cuyo propósito es el análisis de un índice de fiabilidad: el índice p_o de Huynh para pruebas de referencia criterial. El estudio de simulación se ha realizado con la ayuda de un programa de ordenador creado para tal fin. El objetivo es describir la influencia de dos variables –la forma de la distribución y la media- en este índice. La conclusión es que ambas variables están afectando consistentemente el índice de fiabilidad. Además, las evidencias obtenidas aportan información relevante sobre características métricas deseables para la posterior utilización del índice en las pruebas criterioles.</p>
Keywords / Descriptores	<p><i>Evaluation, criterion-referenced evaluation, reliability, criterion-referenced reliability, classification consistency indexes, p_o index, Huynh's method, mean, shape of distribution.</i></p> <p>Evaluación, Evaluación Criterial, Fiabilidad, Fiabilidad en pruebas criterioles, índices de consistencia de la clasificación, índice p_o, método de Huynh, media, forma de la distribución</p>
Institution / Institución	Universidad de Valencia (España)
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Spanish (Title, abstract and keywords in english)

Revista ELección de Investigación y EValuación Educativa (RELIEVE)

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).