

**LOS CONSTRUCTOS EN LAS INVESTIGACIONES PEDAGÓGICAS:
CUANTIFICACIÓN Y TRATAMIENTO ESTADÍSTICO**

**CONSTRUCTS IN PEDAGOGICAL RESEARCHES: QUANTIFICATION AND
STATISTICAL TREATMENT**

Silvio Faustino Soler Cárdenas¹

silviosc.mtz@infomed.sld.cu

Recibido: 5 de julio de 2013

Aceptado para su publicación: 4 de octubre de 2013

RESUMEN

El proceso de medición en las ciencias sociales es esencialmente diferente a aquel que se ejecuta en las ciencias naturales y exactas. En este trabajo se precisan los aspectos característicos de la cuantificación de constructos que se presentan en investigaciones educativas, haciendo especial énfasis en el proceso de construcción y validación de Tests que se usan para estos fines. Se discuten ejemplos reales, donde se pone de manifiesto la necesidad de combinar los métodos clásicos de la Estadística, con procedimientos psicométricos para un

ABSTRACT

The measurement process in social sciences is very different from that of the natural and exact ones. Through this work, characteristic issues of quantification of constructs, which take place in educational researches, are analyzed, with special attention on the construction and validation of tests used for these ends. Actual examples are discussed where appear the necessity of combining classical methods of statistics with psychometric tools, in order to make a correct processing and quantitative analysis of psychological constructs.

¹ Licenciado en Matemática. Profesor Auxiliar. Máster en Educación Médica. Metodólogo de Formación Académica en el Post – Grado. Matanzas. Cuba..

adecuado procesamiento y análisis cuantitativo de constructos psicológicos.

Palabras claves: Constructo, confiabilidad, alfa de Cronbach, propiedades psicométricas, métodos estadísticos.
Keywords: Construct reliability, Cronbach's alpha, psychometric properties and statistical methods.

INTRODUCCIÓN

La metodología estadística juega un papel muy importante en las ciencias de la educación. Sus procedimientos y métodos típicos de razonamiento constituyen elementos muy valiosos para el análisis e interpretación de datos empíricos de las investigaciones pedagógicas y psicológicas.

En términos generales, la estadística contribuye a la investigación de diferentes maneras: a) viabiliza la planificación óptima de experimentos de forma tal que con un mínimo de esfuerzo se logre el máximo de información; b) aporta procedimientos para la caracterización cuantitativa de diagnósticos experimentales; c) permite la comprobación de hipótesis estadísticas.

La Estadística es una disciplina científica cuya misión principal consiste en orientar a los investigadores en la búsqueda de asociaciones, regularidades y explicaciones causales de fenómenos observables. El razonamiento estadístico es, en fin de cuentas, un complemento importante del proceso del conocimiento inductivo.

Debido a que el desarrollo temprano del aparato conceptual y metodológico de la Estadística, se puso en función de problemas de las ciencias naturales y exactas, la aplicación de esta disciplina a la Psicología y la Educación ha originado una serie de problemas específicos, siendo uno de los principales el de la medición psicológica: principios, métodos y procedimientos idóneos para la medición de longitudes, áreas, volúmenes y peso, resultan de poca o ninguna utilidad cuando se aplican de manera irreflexiva a la cuantificación de atributos psicológicos.

En este trabajo, el autor se propone, tomando como base la teoría psicométrica de los test psicológicos, precisar las características fundamentales de la medición psicológica e ilustrar con situaciones reales la aplicación de ciertos procedimientos estadísticos al análisis de mediciones de naturaleza psicológica.

DESARROLLO

La noción de constructo

A diferencia de las variables físicas, las variables psicológicas no se pueden observar de manera directa. No quiere esto decir, que en psicología no se midan conductas directamente observables, cuya cuantificación se suele obtener a través de magnitudes bien establecidas: duración, frecuencia, intensidad, entre otras, sino que, incluso en estos casos, las conductas observables se interpretan como indicios o resultados de variables inobservables más complejas. Atributos como "autoestima", "habilidad lectora", "razonamiento analógico", "competencia profesional", "rendimiento académico", son variables que se cuantifican mediante indicadores medidos en formas de comportamiento.

En la literatura se usan diferentes términos y expresiones para nombrar a este tipo de variable: "variable no observable"; "variable latente"; "intangible" y "constructo". Esta última denominación, que resulta de la españolización del vocablo inglés "construct", es de uso generalizado en el campo de la medición psicológica desde que fue definido por Cronbach y Meehl: "A construct is some postulated attribute of people assumed to be reflected in test performance" (Cronbach L., Meehl .1954, p. 281-302)

Los constructos, como modelos de variables no observables, se caracterizan por cuatro aspectos:

- 1) El sujeto que mide, diseña, valida, implementa y aplica el instrumento de medida.
- 2) Las condiciones bajo las cuales se efectúa el proceso de medición
- 3) El instrumento de medición, que se desea sea estable con respecto a los resultados que proporciona.

4) El sujeto medido

Este último factor, el sujeto medido, es el rasgo distintivo de la medición psicológica en cuanto al hecho de que influye de manera especial en las mediciones. Ni siquiera es suficiente decir que los sujetos “responden” a los Tests. Los sujetos hacen mucho más que eso: piensan, deciden, actúan, interpretan lo que el evaluador explora; los sujetos producen, crean, generan activamente la respuesta a los ítems de los Tests. La medición psicológica es un proceso donde el evaluado es plenamente activo y el evaluador es “pasivo” (en el sentido de que está sujeto a un determinado algoritmo de medición establecido de antemano).

La medición psicológica cambia a los sujetos debido a que estos aprenden, se fatigan; es un proceso que implica ante todo un acto de interacción social entre un sujeto y un contexto demandante y otro que se comporta para satisfacer esas demandas.

Cuando se aplican tests psicológicos, en rigor no es cierto que se observe al sujeto. Tampoco se observa ninguna dimensión o faceta tangible del sujeto. Generalmente ni siquiera se observa la conducta del sujeto, ni la cognitiva ni la afectiva ni la motora. Por lo general solamente se observan ciertos productos de esa conducta. Productos restringidos en los que no pueden reflejarse la mayor parte de las conductas cognitivas (¿Cómo lo ha hecho? ¿Qué decisiones ha tomado? ¿Por qué razones?), ni afectivas (¿Qué deseaba?, ¿Qué sentía y por qué?), ni volitivas (¿A qué aspira el sujeto al actuar así? ¿Qué clase de resultados desea obtener en realidad?).

Aunque en la práctica psicométrica se dice que los sujetos “son medidos” por los tests, sería más preciso decir que los sujetos “se comportan” ante los tests. El problema del sujeto que se mide proviene del hecho de que son personas y por tanto siempre cambiantes en el tiempo y los instrumentos no “miden” una cualidad del sujeto que se observa, más bien activan al sujeto para que genere conductas de ciertos tipos de los que se considera alguna faceta del producto.

La medición de constructos

Debido a que los constructos psicológicos constituyen abstracciones de la realidad que

sólo pueden ser evaluadas indirectamente, el desarrollo del proceso de medición para este tipo de variables presenta algunos problemas específicos:

- No hay una única aproximación a la medida de cualquier constructo que sea universalmente aceptada, ya que las medidas de un constructo psicológico son siempre indirectas, basadas en conductas que se perciben como relevantes al constructo, siempre existe la posibilidad de que dos teorías diferentes seleccionen conductas diferentes para definir operacionalmente el constructo.
- Las medidas psicológicas generalmente se basan en muestras limitadas de conducta, lo que plantea un importante problema a la hora de decidir qué muestras serían adecuadas y cuáles insuficientes.
- La medida obtenida siempre está sujeta a error: consecuencia de tomar muestras limitadas de conducta y de trabajar sobre personas; cambian por efecto de factores muy diversos como la fatiga, el olvido, el aburrimiento, entre otros.
- Escasez de unidades bien definidas en las escalas de medida a las que se ajustan las variables psicológicas.
- Los constructos psicológicos no pueden ser expresados exclusivamente en términos de definiciones operacionales sino que también deben haber demostrado relaciones con otros constructos o fenómenos observables. Se hace necesario un segundo nivel en el que los constructos sean definidos en términos de sus relaciones, ya sean lógicas o matemáticas, con otros constructos dentro de un sistema teórico amplio.

El grado en el que un individuo posee uno de estos atributos, solamente puede inferirse a partir de la observación de su conducta. Este carácter no observable de la mayor parte de las variables psicológicas, hace especialmente difícil el diseño de tests o instrumentos para su medida.

En el proceso de medición de un constructo es necesario distinguir tres aspectos: 1) definición del constructo; 2) construcción de un test cuyos ítems sean variables

directamente observables; 3) establecimiento de la escala de medición del constructo.

La definición del constructo. Los constructos se definen a partir de una serie de comportamientos manifiestos, es decir, se piensa que al constructo corresponde una determinada conducta, y se afirma que después de determinar una serie de ítems con respecto a esa conducta, el constructo está formulado y/o se corresponde con la realidad observada. De aquí que se tengan en cuenta los siguientes aspectos:

Análisis de contenidos. Se plantean una serie de cuestiones abiertas a los participantes sobre el constructo que se quiere evaluar, y sus respuestas se clasifican en tópicos o temáticas. Los tópicos predominantes se toman entonces como componentes mayores del constructo y a su vez serán los que produzcan mayor proporción de ítems.

Revisión de investigaciones publicadas. Se trata de realizar búsquedas bibliográficas relacionadas con el tipo de variables que se quieren medir y establecer aproximaciones con los ítems deseables en función de lo más representativo en la literatura.

Incidentes críticos. Se analizan patrones característicos de los extremos del continuo que se desea medir y en función de esto se redactan los ítems que permitan graduar a los sujetos en un atributo determinado.

Observación natural. Observar las conductas relacionadas con el constructo a medir en situaciones naturales y usarlas para definir el constructo.

Juicio de expertos. Recurrir a especialistas en el tema para que nos proporcionen otras definiciones previamente elaboradas.

Construcción del test. Es preciso plantearse el constructo en toda su complejidad y adoptar decisiones en torno a la representatividad de los ítems para muestrearlos adecuadamente. Para esto es conveniente:

Conceptualización rigurosa de los componentes relevantes del constructo.

Diseñar ítems en las áreas asociadas a los componentes relevantes.

Selección de ítems que permitan discriminar a las personas en cuanto al grado de posesión de los componentes relevantes del constructo.

La escala de medición se define a partir de la determinación de los puntajes del test. Existen varias posibilidades:

Suma (ponderada o no) de los puntajes de los ítems

Promedio (ponderado o no) de los puntajes de los ítems

Método de comparación por pares

Escala de Likert

Análisis por escalograma

Diferencial semántico

La escala de Guttman

Un ejemplo real: diseño de un test de habilidades múltiples (THM)

El ingreso a la educación superior en gran parte de las universidades del mundo exige un proceso de selección previo que se caracteriza por la evaluación de conocimientos y habilidades básicas que deben poseer los aspirantes para lograr un nivel de entrada, que les permita durante el proceso formativo la adquisición y asimilación de contenidos y valores humanistas propios de la profesión.

En Cuba, durante el quinquenio 2000 – 2005, el proceso selectivo de estudiantes para el ingreso a carreras de ciencias médicas incluía la aplicación de un test de habilidades múltiples, diseñado por un grupo de expertos en el campo de las ciencias médicas, atendiendo a requerimientos metodológicos de la evaluación, para dar respuesta a la necesidad de ofrecer un instrumento evaluador de los conocimientos y habilidades fundamentales que deben poseer los aspirantes que ingresen en la educación médica superior y su destacamento especial “Carlos J. Finlay”. Con este test se pretende explorar el constructo “habilidades múltiples”.

Definición del constructo

Las habilidades múltiples se estructuraron atendiendo a cuatro dimensiones:

Información general. Esta dimensión se sustenta en la necesidad de evaluar los conocimientos básicos de anatomía y biología vinculados con las ciencias médicas que el aspirante debe haber adquirido en su formación precedente y fundamentalmente, en el desarrollo de la tendencia orientadora de la personalidad hacia las carreras de la salud, a partir de un enfoque humanista de la Psicología y la Pedagogía que abogan por el reconocimiento del carácter activo del estudiante y su participación como sujeto en el proceso de enseñanza aprendizaje, que destaca el papel protagónico del sujeto en la elección de la profesión. La orientación profesional, por ser un campo complejo, se aborda de forma estructurada sobre la concepción teórica de la personalidad y el papel del proceso docente – educativo, lo que implica además, la educación y desarrollo de la esfera motivacional y cognitiva vinculadas con la elección de la profesión y el desempeño posterior

Razonamiento verbal. El razonamiento verbal evidencia el desarrollo del lenguaje, precisión en el empleo del vocabulario, capacidad para discernir ideas con claridad y conocimiento del significado de variados vocablos. La aptitud verbal denota al lector inteligente y presupone efectividad en la comunicación en sus formas oral y escrita. El razonamiento verbal desempeña un papel fundamental en el aprendizaje, además de ser un buen predictor del éxito académico, por tanto, el obtener altos puntajes en los correspondientes ítems sugiere el afrontamiento de forma óptima de las futuras tareas universitarias.

Razonamiento lógico – matemático. Los aspirantes que muestran habilidades de este tipo de razonamiento, evidencian desarrollo en la solución de problemas cotidianos y solución de ejercicios que requieren generalización y aplicación, mediante algoritmos lógicos en función de las capacidades de razonamiento y abstracción, propias del proceso de pensamiento y de los conocimientos adquiridos en su formación previa.

Ingeniosidad y creatividad. Los aspirantes deben mostrar posibilidades de dar solución

a situaciones problemáticas complejas, transitando de lo desconocido a lo conocido como un resultado que debe alcanzar y utilizando las capacidades de generalización, identificación de elementos esenciales y otras del pensamiento abstracto

Construcción del test y la escala de medición

Se confeccionó un test de 50 ítems dicotómicos: en cada una de las dimensiones 1; 2 y 3 se incluyeron 15 ítems y cinco en la cuarta. A cada ítem se le asignó un puntaje máximo de 2 y el puntaje del test se definió como la suma de los puntajes de todos los ítems.

Nótese que no todas las dimensiones tienen la misma importancia en el sentido de la proporción del puntaje total que a cada una le corresponde.

Dos propiedades psicométricas importantes: validez y confiabilidad

Una vez concluida la confección de un test se presenta el problema de determinar en qué medida la información obtenida por conducto de su aplicación refleja el atributo que se está explorando. Surgen, muy naturalmente, las siguientes preguntas: ¿Mide el test lo que realmente debe medir?; ¿Con qué precisión y estabilidad se mide lo que se está midiendo?

La primera pregunta está relacionada con la "validez" de un test; la segunda concierne a lo que se conoce con el término "confiabilidad".

Validez

Un aspecto importante de la calidad de un test es el que se refiere a la validez, es decir, a la correspondencia que existe entre lo que el test pretende medir y lo que mide realmente.

Ello abarca tanto las exigencias propias del test, como la determinación de los índices valorativos que permiten su calificación.

La validez nos proporciona un control directo de la forma en que un test cumple sus

propósitos.

La determinación de la validez requiere criterios de expertos generalmente externos y nos capacita para determinar el grado de acierto del diseño del test a partir de los resultados obtenidos, sobre la base de una muestra representativa de los mismos.

Tipos de estudios de validez

La validez usualmente se clasifica en tres grupos: a) validez de contenido; b) validez de criterio; c) validez de construcción.

Validez de contenido. La validez de contenido consiste en el examen sistemático del contenido del test para determinar si constituye una muestra representativa del dominio de conductas que se pretende evaluar (dominio cognitivo).

Los estudios de validez de contenido se basan fundamentalmente en técnicas de consenso basadas en opiniones de expertos.

Validez de criterio. Cuando se realiza un estudio de validez de criterio, se comparan los resultados de la aplicación del test con otros criterios externos a lo que se le llama "variable – criterio".

Se dice que un test tiene validez de criterio cuando, a partir de los resultados obtenidos, se pueden hacer inferencias válidas sobre una variable que no se puede medir directamente por el test.

Un estudio de validez de criterio implica la ejecución de las siguientes tareas:

Determinación de un criterio adecuado y de un procedimiento para medirlo.

Determinación de una muestra representativa de personas para aplicarle el test.

Aplicar el test y anotar el puntaje obtenido de cada examinado.

Obtener una medida del criterio para cada examinado en el momento en que sea posible.

Determinar la fortaleza de la relación entre los resultados del test y los del criterio.

Los estudios de validez de criterio se clasifican en dos tipos:

- Validez predictiva

- Validez concurrente

La validez predictiva indica la efectividad del IE en un resultado futuro. Para esto se comparan las calificaciones del IE con una medida directa de la ejecución posterior de los sujetos llamado "criterio".

La validez concurrente se refiere a la interrelación del test y un criterio de medición efectuado en el momento en que se aplica el test.

En algunos casos la validez concurrente es meramente un sustituto de la validez predictiva. No suele ser practicable extender los procedimientos de validación más allá del tiempo requerido para la validez predictiva u obtener una determinada muestra de preselección con fines de aplicación del test. Por tanto, como solución de compromiso se aplica a un grupo sobre el que ya se tienen disponibles los datos del criterio.

La diferencia entre la validez predictiva y la concurrente no sólo se basa en el tiempo sino también en los objetivos de aplicación del test. La validez concurrente es idónea para los instrumentos empleados en el diagnóstico de un estado existente y no para la predicción de resultados futuros.

Validez de construcción

Los procedimientos estadísticos que normalmente se usan en los estudios de validez de construcción se caracterizan por tener cierto nivel matemático de complejidad. Nos limitaremos, pues, a hacer solo mención de algunos que usualmente se aplican en las investigaciones, ellos son los siguientes: diferenciación entre grupos; validez discriminante; método matricial de rasgos múltiples; análisis factorial.

Confiabilidad

Cuando se aplica un test de cualquier tipo (ya sea un test de actitudes, un test de rendimiento en una tarea específica o simplemente un examen escrito para explorar conocimientos), el puntaje obtenido depende de un conjunto de condiciones internas (propias del examinado) y externas (el medio) y por tanto el mismo test aplicado a la misma persona pero en momentos diferentes, puede arrojar puntajes diferentes. En la práctica se presentan frecuentemente numerosos factores difíciles de controlar y que a la larga determinan la inconsistencia de los puntajes de un test.

Cronbach hace referencia a cuatro grupos importantes de tales factores:

Características generales y duraderas del examinado: a) habilidades generales, por ejemplo la lectura; b) habilidades para comprender las instrucciones de un test; c) habilidades para resolver los problemas de tipo general presentados en los tests; d) actitudes, reacciones emotivas y hábitos característicos de las personas sometidas a un test. (Cronbach L. 1968. p. 29)

Características duraderas y específicas del examinado: a) conocimientos y habilidades que requieren los problemas específicos del test; b) actitudes, reacciones emotivas o hábitos que se ponen de manifiesto en las personas al ser evaluadas; c) características generales y momentáneas del examinado; d) la salud, la fatiga y la tensión psíquica; e) la motivación, ya sea positiva o negativa, que produce el hecho de sentirse evaluado; f) los efectos del calor, la luz, la ventilación; g) actitudes actuales, reacciones emotivas o fuerza de los hábitos, en cuanto todos estos representan divergencias de los rasgos más duraderos y estables del examinado.

Características temporales y no generales del examinado: a) cambios debidos a la fatiga o la motivación producidos por el test; b) fluctuaciones en la atención, en la coordinación o en los patrones de juicio; c) fluctuación en la memoria de hechos particulares; d) nivel de práctica en la ejecución de las habilidades o en los conocimientos requeridos para el problema en cuestión; e) buena suerte en la selección de las respuestas al adivinarlas.

Variados han sido los enfoques presentados en la literatura para cuantificar la

inconsistencia de puntajes de tests como consecuencia de la influencia de los factores mencionados anteriormente. En este sentido, Guilbert propone seis propiedades:

Objetividad: grado de concordancia entre los juicios emitidos por examinadores independientes sobre lo que constituye una “buena respuesta” para cada uno de los ítems de un test.

Pertinencia: grado de respeto de los criterios establecidos en la elección de los ítems para que sean conformes con los fines y objetivos del test.

Equilibrio: diferencia entre la proporción de ítems de las áreas que se exploran y la proporción ideal para un test de ese tipo.

Equidad: grado de correspondencia entre los ítems y los rasgos psicológicos que se quieren explorar.

Discriminación: cualidad de cada ítem que permite diferenciar a las personas de puntuaciones altas y bajas.

Eficacia: cantidad promedio de ítems que se pueden responder por unidad de tiempo. (Guilbert. 1977, p. 123)

No obstante, para este fin hay un concepto que, en esencia, desde que fue definido se mantiene vigente en la teoría y aplicaciones de tests: el concepto de confiabilidad presentado por el psicólogo británico Charles Spearman, en sus trabajos de los años 1904 –1913, donde presenta la llamada Teoría Clásica de los Tests (TCT). (Williams R H, Zimmerman D W, Zumbo B D, Ross D. Charles Spearman. 2003, p. 188)

Estudios de confiabilidad

En la actualidad en la literatura aparecen varios enfoques para cuantificar la confiabilidad de un test; no obstante, hay tres teorías dominantes: 1) Teoría Clásica de los Tests, 2) Teoría de Respuesta al Ítem y 3) Teoría de la Generalizabilidad. Estas teorías se complementan en el sentido de que estudian diferentes aristas de la confiabilidad. No obstante, en la literatura se ha impuesto la primera de ellas mediante

el uso del coeficiente alfa de Cronbach. Sin lugar a dudas, este es el recurso numérico más utilizado:

En primer lugar, el trabajo donde aparece por primera vez (publicado en 1951) ha sido citado no menos de 5590 veces. Durante el quinquenio 1995 – 2000 fue citado alrededor de 325 veces anualmente. (Cronbach L. 1951. Consultado en Internet) Para tener una idea de cuán grande es ese número, baste tener en cuenta que en ese período el número promedio anual de citas de un artículo de ciencias sociales fue de 11.

En 2do. lugar, Thomas P. Hogan, Amy Benjamin, and Kristen L. Brezinski reportan la frecuencia de aplicación de varios tipos de coeficientes de confiabilidad que aparecen en una muestra sistemática seleccionada de la APA-published Directory of Unpublished Experimental Mental Measures. Este directorio comprende 37 revistas profesionales de educación, psicología y sociología. (Thomas P. Hogan, Amy Benjamin, and Kristen L. Brezinski. 2000, p. 523) Al coeficiente alfa le correspondió el mayor porcentaje (66,5 %); el segundo lugar lo ocupó el test-retest (19,0%) y todos los demás se aplicaron en menos del 5% de los casos.

En tercer lugar Liu Y y Zumbo B. D. reportan que en una revisión del Social Sciences Citations Index del período 1966 – 1995 (Liu Y, Zumbo BD. 2007, p 56) , el artículo de Cronbach de 1951 había sido citado aproximadamente 60 veces por año en un total de 278 revistas que cubren varias áreas de investigación relacionadas con psicología, educación, sociología, estadística, medicina, enfermería, ciencias políticas y economía.

Una aplicación del coeficiente alfa

El análisis de confiabilidad basado en el coeficiente alfa de Cronbach puede ser muy útil para la selección de los ítems de un test.

Soler Cárdenas S. F. y Soler Pons L. reportan la aplicación de un examen de 30 ítems con un coeficiente alfa igual a $-0,0442$. (Fernández Oliva B., Soler Martínez C., Rodríguez Fontes R., Soler Cárdenas S. F. 2004, p. 8). En este caso, mediante el análisis de confiabilidad basado en coeficientes de discriminación, se logró identificar a los tres ítems que determinaban un valor negativo para el coeficiente alfa. Los

resultados fueron los siguientes:

Coeficiente alfa	Preguntas incluidas
- 0,0442	Todas las preguntas
0,2356	Se elimina la de menor discriminación
0,5853	Se eliminan las dos de menor discriminación
0,6236	Se eliminan las tres de menor discriminación

Nótese cómo la eliminación de tres preguntas determina un aumento ostensible de la confiabilidad del examen.

Análisis estadístico de los resultados de un examen estatal

Vázquez Adán Y, analiza los resultados del examen estatal escrito aplicado en el curso 2009 – 2010 a 264 estudiantes de estomatología pertenecientes a cuatro universidades médicas de Cuba. (Vázquez Adán Y. 2012, p. 42). Este examen explora seis temas que abarcan los contenidos esenciales de toda la carrera.

En la Tabla 1 aparecen las estimaciones por intervalos para las calificaciones promedio de las facultades incluidas en el estudio. Cada intervalo basado en la Estadística Psicométrica (EP) es más preciso (tiene menor longitud) que el correspondiente de la Estadística Clásica (EC). Nótese además que para la facultad 3, según la EP, la calificación promedio es “3” (puntajes entre 70 y 80) y, de acuerdo a la EC, el promedio oscila entre “3” y “4” (puntajes entre 80 y 90). Un comentario similar tiene lugar para la facultad 4. Conforme muestran estos resultados, no siempre las interpretaciones coinciden; aunque lo correcto en este caso es la aplicación de la EP

Tabla 1

Facultad	Promedio	Estimaciones por Intervalos del 99%	
		Estadística Clásica	Estadística Psicométrica
1	77,8	75,91 ----- 79,69	76,93 ----- 78,67
2	86,96	84,61 ----- 89,31	85,58 ----- 88,34
3	77,9	74,17 ----- 81,63	75,88 ----- 79,92
4	82,1	79,93 ----- 84,38	80,98 ----- 83,22

Tabla 2

	T1	T2	T3	T4	T5	T6
T1	-----					
T2	0,44 0,82	-----				
T3	0,33 0,75	0,37 0,85	-----			
T4	0,34 0,82	0,34 0,82	0,34 0,99	-----		
T5	0,33 0,98	0,35 0,99	0,26 0,95	0,24 0,94	-----	
T6	0,31 0,71	0,36 0,84	0,24 0,69	0,19 0,57	0,20 0,77	-----

La Tabla 2 contiene los coeficientes de correlación lineal entre los puntajes de los temas del examen. Dentro de cada celda, el número escrito “en negrita”, que corresponden a la EP es el mayor. En este caso, la Estadística Clásica estima por defecto la correlación entre los temas del examen.

CONCLUSIONES

Los constructos forman parte del conjunto de variables que habitualmente se estudian en las investigaciones educacionales. De aquí la necesidad de capacitar a todos aquellos que se dedican a esta actividad, en el tratamiento adecuado de los constructos.

El enunciado preciso de una variable no observable directamente, así como su caracterización cuantitativa, requieren conocimiento sobre diseño, aplicación y validación de tests psicológicos.

La aplicación de métodos estadísticos de variables psicológicas debe estar precedida de un análisis psicométrico del constructo involucrado.

REFERENCIAS BIBLIOGRÁFICAS

Cronbach L. (1951) Coefficient alpha and the internal structure of tests. Consultado en: <http://www.unc.edu/~rcm/psy330/cronbach.1951.pdf>

Cronbach L. (1968). Fundamentos de la Exploración Psicológica. Edición revolucionaria. Instituto Cubano del libro. La Habana.

Cronbach L., Meehl (1954) P. E. Construct Validity in Psychological Tests. Psychological Bulletin, 52.

Fernández Oliva B., Soler Martínez C., Rodríguez Fontes R., Soler Cárdenas S. F. (2004). Diseño y aplicación del Test de Habilidades Múltiples en el Proceso Selectivo de Ingreso al Destacamento “Carlos J. Finlay”. Educ. Med. Sup. V. 18 No. 1. Ciudad de La Habana.

Guilbert JJ. (1977) Guía Pedagógica. Ginebra. Organización Mundial de la Salud.
Reimpreso por la Organización Panamericana de la Salud.

Liu Y, Zumbo BD. (2007). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: visual analogue scales. 67;620 Educational and psychological measurement

Soler Cárdenas SF y Soler Pons L. (2012). Usos del coeficiente alfa en el análisis de instrumentos escritos. Rev. Med. Electrón. Vol. 34 No. 1. Matanzas.

Thomas P. Hogan, Amy Benjamin, and Kristen L. Brezinski. (2000). Reliability Methods: a Note on the Frequency of Use of Various Types. Educational and Psychological Measurement, Vol. 60 No. 4.

Vázquez Adán Y. (2012). Análisis de fiabilidad del examen estatal escrito de la carrera de Estomatología. Curso 2009 – 2010. Trabajo para optar por el título de Máster en Educación Médica Superior. Centro de desarrollo Académico en Salud. Ministerio de Salud Pública. La Habana.

Williams R H, Zimmerman D W, Zumbo B D, Ross D. Charles Spearman: British Behavioral Scientist. Human Nature Review 2003. Volume 3.