

ESCALAS O ÍNDICES PARA LA MEDICIÓN DE CONSTRUCTOS: EL DILEMA DEL ANALISTA DE DATOS

Eiliana Montero Rojas*
Universidad de Costa Rica, Costa Rica

Resumen

Se intenta disipar algunas de las mayores confusiones existentes entre los analistas de datos en relación con la naturaleza de los índices y las escalas y también en relación con la metodología para su construcción. Se definen conceptos clave y se listan siete diferencias fundamentales entre índices y escalas. Estas son: El número de componentes o dimensiones representados en el indicador, el para qué (propósito) del indicador, la utilización de métodos psicométricos, el peso o ponderación de los ítems o reactivos, el tratamiento de los valores faltantes (missing), la necesidad de estandarizar estadísticamente la medida y el grado de interpretabilidad. A lo largo de todo el documento se dan ilustraciones utilizando ejemplos reales. Se espera que esta discusión contribuya a una toma de decisiones informada por parte de los profesionales que trabajan analizando y/o construyendo este tipo de medidas y por parte de aquellos que deben usarlas o analizarlas ocasionalmente.

Palabras clave: Escalas, índices, constructos

Abstract

An attempt is made to dissipate some of the major confusions among data analysts regarding the nature of scale and indexes, and also the methodology for their construction. Key concepts are defined and seven basic differences between indexes and scales are listed. They are: the number of components or dimensions represented in the indicator, the purpose of the indicator, the use of psychometric methods, the weights assigned to the items or reactivities, the treatment of the missing values, the need for statistical standardization and the degree of interpretability. Throughout the document illustrations are provided with real examples. It is hoped that this discussion contributes to an informed decision making by the professionals that work analyzing and/or constructing this type of measures and by those who have to use them or analyze them occasionally.

Key words: Scales, indexes, constructs.

Introducción

El objetivo básico de este documento es contribuir a disipar algunas de las mayores confusiones existen entre los investigadores sociales y analistas de datos en relación con la naturaleza de los índices y escalas y también en torno a la metodología para su construcción y validación.

Esta temática posee relevancia dado que se presentan frecuentemente instancias reales que ilustran problemas metodológicos en la elaboración e interpretación de índices y escalas. En ciertas ocasiones, al conocer sólo de manera superficial el enfoque psicométrico, algunos investigadores novatos intentan obtener, por ejemplo, la medida de confiabilidad Alpha de Cronbach para todas las preguntas de un cuestionario, aun cuando se trate de grupos distintos de reactivos que intentan representar diferentes constructos. Otro caso se da cuando se incluyen en el análisis psicométrico variables directamente observables como el sexo y la edad. Estos ejemplos evidencian la necesidad de capacitar a los investigadores para que puedan distinguir apropiadamente las instancias en que se deben emplear tales técnicas psicométricas y aquellas que requieren otros procedimientos de análisis para su elaboración y validación.

* Instituto de Investigaciones Psicológicas, Universidad de Costa Rica, San José de Costa Rica. Tel: (506) 207-5636. Fax: (506) 207-5493. E-mail: emontero@cariari.ucr.ac.cr

Asimismo, es también frecuente entre investigadores de áreas diferentes a la psicología, o aún entre algunos psicólogos, la aplicación de instrumentos e interpretación sustantiva de sus puntajes, sin haber realizado previamente el necesario proceso empírico para garantizar su validez y confiabilidad. Esto sucede a veces cuando se trata de medidas que fueron desarrolladas y validadas en otro contexto social o cultural.

Además, esta autora ha conocido casos donde investigadores inexpertos de áreas como psicología, salud y educación han realizado interpretaciones sustantivas de los puntajes de cada uno de los ítems individuales de una escala, ignorando que en este tipo de medidas el puntaje total es el que representa el nivel del constructo en cada individuo, y que, por tanto, no tiene sentido interpretar por sí mismo cada ítem o reactivo que compone la escala.

Anastasi & Urbina (1998), Babbie (2000), Kerlinger (1988), y Nunnally y Bernstein (1995), son algunos de los autores más conocidos que, con ciertas diferencias, han ayudado a estructurar y divulgar los conceptos y técnicas desarrollados en este artículo.

Definiciones básicas

Constructo: Es una conceptualización que requiere de un marco teórico para ser definido. En general, las conceptualizaciones que estudian los investigadores científicos presentan la característica de que no existe un claro consenso a nivel social en cuanto a cómo definirlos o medirlos, sino que para lograrlo se debe contar con una teoría que los sustente. Otra característica fundamental de los constructos es que no son directamente observables y su “captura” a nivel empírico requiere generalmente de rigurosos procedimientos. (Babbie, 2000; Cortada de Kohan, 2001; Kerlinger, 1988). Y es que es evidente que el estudio de constructos, rasgos o variables latentes es el “pan de cada día” en la investigación social y psicológica. Se intentan estudiar y medir rasgos de personalidad, atributos intelectuales, actitudinales, etc. Ejemplos de constructos son “desarrollo humano”, “violencia de género”, “calidad de vida”, “actitud hacia la estadística”, “aptitud académica”, etc.

Indicador: El indicador es la operacionalización de una variable o constructo. Debe describir las operaciones de medición necesarias para su cálculo empírico. En general, es más fácil definir indicadores para medir variables, comparados con la medición de constructos. Por ejemplo, la edad se puede medir con el indicador “número de años cumplidos de vida”, mientras que la medición de un constructo como ansiedad requiere de procedimientos más sofisticados, menos obvios y que se reporten de manera más rigurosa sus evidencias de validez. (Babbie, 2000; Kerlinger, 1988).

Índice: El índice es un tipo especial de indicador. Es una medida compuesta que relaciona variables o constructos de diferente naturaleza. (Gómez, 1997). Quizá algunos de los índices más conocidos son los índices de precios, en los cuales se relacionan precios y cantidades de artículos o bienes. Otro de los índices más conocidos es el Índice de Desarrollo Humano de las Naciones Unidas, que relaciona, en una sola medida, tres constructos diferentes: Educación, medida por dos indicadores: tasa de alfabetismo y tasa de matriculación combinada; longevidad, medida por la esperanza de vida; y poder adquisitivo, medido por el indicador ingreso anual per cápita en dólares. (United Nations Development Programme, 2007).

Escala: La escala es también una medida compuesta, pero sólo relaciona variables de la misma naturaleza y, en general, se usa para medir constructos complejos como actitudes, valores, habilidades intelectuales y rasgos de personalidad. En psicología se han diseñado muchas escalas para medir rasgos de personalidad como estabilidad emocional (neuroticismo), introversión, empatía, etc. y también habilidades intelectuales. (Nunnally & Bernstein, 1995; Anastasi & Urbina, 1998; Martínez, 2005).

En general, tendemos a usar más los índices que las escalas en nuestro trabajo estadístico, quizá porque para construir apropiadamente una escala es necesario un manejo exhaustivo del marco de referencia teórico que permite definir el constructo a medir. Ejemplos de índices son:

a. *El Promedio de Admisión a la Universidad de Costa Rica*. Es un índice porque relaciona dos variables de claramente diferente naturaleza: el promedio de notas de los dos últimos años de la educación secundaria y el puntaje en una prueba de Aptitud Académica o Prueba de Admisión de dicha Universidad. Actualmente ambos componentes tienen la misma ponderación en el Promedio de Admisión, o sea cada uno representa el 50% del Promedio de Admisión.

b. *Índice de Desarrollo Humano*. Como ya se mencionó antes, el Índice de Desarrollo Humano relaciona tres constructos diferentes: Educación, longevidad y poder adquisitivo. Hay dos indicadores para Educación y un indicador respectivamente para cada una de las otras dos dimensiones. Los tres componentes tienen el mismo peso en el índice, es decir cada uno contribuye con un tercio al valor del índice.

c. *Nota final en un curso de Estadística*. Es un índice cuando para su cálculo se usan indicadores de diferentes dimensiones como por ejemplo, notas de exámenes, notas de presentaciones orales, notas de trabajos en grupo. Por el contrario, podría ser considerado una escala si sólo relacionara, por ejemplo, notas de exámenes y se pudiera asegurar que lo que se pretende medir con esas pruebas es fundamentalmente un único constructo. De todas formas, en este último caso habría que definir claramente cuál es el rasgo que se pretende medir.

d. *Indicadores indirectos de nivel socioeconómico*. Sucede con estos algo similar que con el anterior. Si podemos válidamente argumentar que el conjunto de preguntas, ítems o reactivos que componen el indicador de nivel socioeconómico efectivamente están midiendo un sólo constructo teórico con ese nombre, entonces diríamos que es una escala, de otra manera sería un índice. Un ejemplo de escala socioeconómica fue la que se usó hace algunos años en una investigación sobre factores asociados al rendimiento académico en la Universidad de Costa Rica. Se preguntó a una muestra aleatoria de estudiantes acerca de la tenencia o no en su núcleo familiar directo, de los siguientes artículos o servicios: Televisión por cable o satélite, agua caliente en toda la casa, vehículo (no de trabajo), computadora, acceso a Internet (desde la casa), teléfono celular y pertenencia a un club de recreación privado. Luego, con los datos recolectados se realizó un análisis factorial que mostró evidencia clara de un solo factor o dimensión subyacente, por lo podemos llamarla escala socioeconómica. (Montero, Villalobos & Valverde, 2007).

e. *Índice de Calidad de Vida en niños con enfermedades terminales*. Este es un ejemplo que se deriva de una consultoría en que participó esta autora. Es un índice, ya que relaciona tres dimensiones de diferente naturaleza: (a) Aspectos médicos, (b) Aspectos de la cuidadora del niño(a), y (c) Aspectos del medio ambiente del niño(a). Los aspectos médicos y los aspectos de la cuidadora a su vez representan escalas. (Zúñiga & Montero, 2007).

Diferencias entre índices y escalas

Pasemos ahora a comentar las principales diferencias entre índices y escalas, habiendo ya definido cada uno de estos.

1. En la escala está más claramente definido el constructo, y se supone una sola dimensión subyacente. Por el contrario, en el caso del índice se puede trabajar con constructos y variables definidos menos claramente desde el punto de vista teórico, tales son los casos descritos anteriormente de las notas finales de un curso y el indicador de nivel socioeconómico. Usamos índices, en general, para medir estas

dimensiones y nos preocupamos menos por la definición teórica de los constructos. En el primer caso el constructo podría ser “rendimiento académico” y en el segundo “nivel socioeconómico”. En ambas situaciones deberíamos contar con un marco teórico que permitiera definir si se trata efectivamente de un constructo, pero en la práctica sucede que muchas veces no contamos con ese marco de referencia o no lo construimos. Pero si se quisiera llamar “escala” a cualquiera de esas medidas compuestas, deberíamos poder argumentar que los ítems así llamados están midiendo efectivamente un constructo teórico llamado “nivel socioeconómico” o “rendimiento académico”.

En este contexto, además de la teoría, una herramienta muy útil para ayudarnos a dilucidar la estructura subyacente en un grupo de indicadores o ítems es el análisis de factores. Esta técnica de la estadística multivariada puede utilizarse una vez que se han recolectado datos empíricos, para arrojar luz sobre si lo que se está midiendo con el instrumento es un índice o una escala. Si claramente la estructura factorial apunta hacia una dimensión, componente o factor, entonces podemos concluir que hay evidencia de que los ítems componen una escala, si la medida se construyó con ese fin.

2. Debido a que la escala trata de medir un sólo constructo teórico es más útil para ser usada con propósitos de diagnóstico. Precisamente en el caso de la escala vamos a tener mayor claridad en términos de lo que está midiendo y de cómo interpretar valores altos o bajos en los puntajes. Si el constructo es “Actitud hacia la estadística” y Juan y María tienen ambos 20 puntos en la escala podemos concluir que ambos presentan la misma intensidad en términos del constructo. Por el contrario, en el caso de un índice, como el de Desarrollo Humano, dos países pueden exhibir el mismo valor en el índice pero tener muy diferentes combinaciones de los tres componentes. Por ejemplo, el país A podría tener valores altos en el Producto Interno Bruto (PIB) per Cápita y bajos valores en los indicadores de Educación, y el país B podría exhibir valores altos de los indicadores de educación y valores bajos en el PIB per Cápita. Aún así los dos podrían tener el mismo valor en el Índice de Desarrollo Humano. De esta manera decimos que los índices, en general, tienen menos valor diagnóstico, debido a que estamos combinando variables y constructos de diferente naturaleza y por ello su *interpretabilidad* es menor que la de las escalas. En este contexto entendemos interpretabilidad como la capacidad que posee un indicador para poder inferir, a partir de sus valores, un sentido o explicación teórica, en términos de sus constructos subyacentes.

3. En general, en las escalas es recomendable y necesario el uso de métodos psicométricos para establecer evidencias de su grado de validez y confiabilidad. Una de las medidas más importantes que podemos usar para establecer el grado de confiabilidad de una escala es el Alpha de Cronbach, que mide la confiabilidad desde el punto de vista de la consistencia interna del instrumento. Pero para su correcta aplicación se requiere que el instrumento sea unidimensional, es decir que esté midiendo una sola cosa, tal como en las escalas. Los índices, por definición, están mezclando en una sola medida compuesta variables de diferente naturaleza, y por tanto, sería errado tratar de usar el Alpha de Cronbach para establecer su grado de confiabilidad. En el caso de las escalas, por el contrario, estos métodos son los más apropiados. Un ejemplo de este tipo de análisis se presenta en el cuadro 1, en donde se muestra la salida del procedimiento RELIABILITY del SPSS en la depuración de la escala mencionada anteriormente para medir la dimensión de Aspectos Médicos como componente del Índice de Calidad de Vida en niños con enfermedades terminales. A partir de la Tabla 1 y observando los valores del índice de discriminación y la columna *Alpha al eliminar el ítem*, podemos concluir que los ítems M2, M7 y M9 deberían eliminarse de la escala para lograr mayores niveles de confiabilidad.

De esta forma, los métodos psicométricos nos proveen de poderosas herramientas para dotar de calidad técnica a las escalas que construimos. En este caso específico nos permiten construir escalas de alta confiabilidad. Desafortunadamente el uso de la psicometría es menos apropiado en el caso de los índices, precisamente porque no se cumple con el supuesto fundamental de unidimensionalidad. Por el contrario, la construcción de índices y el análisis de su calidad técnica requiere más del concurso de expertos para definir y analizar sus dos propiedades fundamentales: validez y confiabilidad.

Tabla 1.
Resultados iniciales del análisis de ítems para la escala de Aspectos Médicos¹

Nombre de la variable	Promedio	Desviación estándar	Índice de discriminación	Alpha al eliminar el ítem
M1 Tiempo de diagnóstico	,7131	1,024	,389	,692
M2 Dolor	2,2623	,960	,018	,749
M3 Presencia de síntomas estresantes	1,2295	,869	,442	,685
M4 Estado de conciencia	2,4918	,730	,413	,692
M5 Integridad física del paciente	1,2951	1,026	,508	,670
M6 Actividad motora	1,2377	1,005	,556	,662
M7 Respiración	2,6148	,776	,136	,726
M8 Alimentación	1,7541	1,215	,652	,635
M9 Medicamentos	2,7623	,515	,132	,722
M10 Vías de aplicación de medicamentos	2,2377	1,260	,465	,678

Alpha de Cronbach= 0.7163. Fuente: Montero, E. (consultoría)

4. En las escalas generalmente todos los ítems o reactivos tienen el mismo peso y están respondidos con el mismo tipo de categorías. Precisamente porque todos los reactivos, ítems o preguntas están dirigidos a medir el mismo constructo, en general se construyen usando el mismo formato y las mismas categorías de respuesta, que pueden ser dicotómicas (SI-NO), ordinales tipo Likert de frecuencia (Desde Siempre o Casi siempre hasta Nunca o Casi nunca), ordinales tipo Likert de grado de acuerdo (Desde Muy de acuerdo hasta Muy en desacuerdo) o, correcto-incorrecto, en el caso de pruebas para medir habilidades intelectuales. En los índices no sucede lo anterior, precisamente porque las variables que se combinan son de diferente naturaleza. En el caso del Índice de Desarrollo Humano, por ejemplo, algunos indicadores son porcentajes (la tasa de alfabetismo y la tasa de matriculación combinada), otros están dados en años (Esperanza de Vida) y otros en dólares (PIB Per Cápita anual). Conviene aclarar aquí que las medidas llamadas “tasa de alfabetismo” y “tasa de matriculación combinada” son en realidad porcentajes pues relacionan una parte con su totalidad; sin embargo, por convención, tradicionalmente se les ha llamado “tasas”.

En las escalas es inusual asignar diferentes ponderaciones a los ítems. En general, todos los reactivos van a tener la misma ponderación en la medida compuesta que representa el puntaje en el constructo y en este caso esa medida compuesta no es otra cosa que la suma simple de los puntajes de los ítems, una vez que se han recodificado aquellos sea necesario para garantizar que todos queden con sus puntajes alineados en la misma dirección (es decir, representando, por ejemplo bajos niveles del constructo con puntajes bajos y altos niveles del constructo con puntajes altos). No hay gran necesidad de ponderar para que todos tengan el mismo peso en la escala, pues, como se mencionó antes, en general van a tener el mismo formato y las mismas categorías de respuesta.

Por el contrario, la decisión de cómo ponderar cada uno de los componentes de un índice es quizá una de las más importantes decisiones que se deben tomar durante su construcción. Aunque existen diferentes criterios para asignar las ponderaciones o pesos a las variables del índice, debemos decir aquí que el marco teórico y el juicio experto son fundamentales en este proceso de asignación de ponderaciones.

5. El tratamiento de los valores faltantes es más fácil de resolver en las escalas, puesto que podemos usar los datos de los otros componentes o ítems de la escala para estimar o imputar lo que el sujeto habría respondido, dado que todos los ítems están midiendo el mismo constructo y es razonable pensar que el sujeto seguirá en ese ítem la tendencia que mostró en los valores que sí respondió. Por el contrario, si en el

índice se presentan valores faltantes, se deben seguir procedimientos clásicos estadísticos tales como la sustitución por la media del conjunto de datos, porque no se puede suponer que el sujeto va a seguir, en el valor faltante, la misma tendencia que exhibe en los otros componentes del índice (son constructos o variables de diferente naturaleza). Resulta entonces más difícil la imputación. Desde este punto de vista son más válidos los procedimientos para realizar dicha imputación, si se puede asegurar que efectivamente todos los reactivos son expresión del mismo constructo subyacente y deben seguir, por tanto, la misma tendencia. Sin embargo, cabe comentar aquí que muchos paquetes estadísticos tradicionales como el SPSS no dan la opción de sustituir los valores faltantes siguiendo la tendencia del individuo en vez de la tendencia del grupo, por lo que la sustitución de valores faltantes para una escala puede volverse una tediosa tarea para el analista.

6. La estandarización estadística (transformar linealmente las medidas para que tengan un promedio de cero y desviación estándar igual a uno) es un paso indispensable en la creación de los índices, precisamente porque es necesario eliminar el efecto de las magnitudes y de las unidades de medida en cada uno de los indicadores, previo a su combinación en una sola medida compuesta. Al estandarizar cada uno de los componentes del índice nos aseguramos que todos ellos estarán dados en las mismas unidades (desviaciones estándar por arriba o por debajo del promedio), para luego combinarlos apropiadamente y darles las ponderaciones o pesos convenientes. El proceso de estandarización estadística es menos necesario en las escalas, porque como ya fue explicado antes, usualmente todos los ítems de una escala están en el mismo formato y utilizan las mismas categorías de respuesta. Valga aclarar en este punto que cuando hablamos aquí de estandarización estadística no nos estamos refiriendo al llamado proceso de “estandarización de un instrumento”, término muy utilizado en psicología y que involucra estudios de validez y confiabilidad para poblaciones específicas, el establecimiento de baremos y estándares de interpretación para sus puntajes.

7. Interpretabilidad. Como ya fue comentado antes, las escalas resultan más claras de interpretar pues están representando una sola dimensión o constructo. De hecho, el propósito fundamental de una escala es lograr una aproximación válida al nivel que exhibe el examinado en el constructo de interés.

¿Cómo decidir entre un índice o una escala?

Enfrentado con la disyuntiva de construir un índice o una escala, el analista de datos debe considerar, primeramente, el propósito sustantivo de la medida. Esperamos que analizando las siete diferencias que se enunciaron en la sección anterior se puedan tomar decisiones más acertadas tanto desde el punto de vista teórico como metodológico. Debe enfatizarse aquí que los índices son muy buenos como medidas resumen y medidas sumativas, pero que sus problemas inherentes de interpretabilidad los hacen menos atractivos para propósitos de diagnóstico. Por ejemplo el Promedio de Admisión no se puede usar para hacer diagnósticos en cuanto a la capacidad de aprendizaje de un estudiante ni tampoco para decir qué carrera debería estudiar. Igualmente observando solamente el valor del Índice de Desarrollo Humano no podríamos identificar si un país determinado necesita invertir más en educación, infraestructura de salud o generación de producción o fuentes de empleo. La clave es usar la herramienta correcta para cada propósito.

Interpretaciones de normas o de criterios

Otro tema asociado que además es poco conocido en Estadística es el tema de los modelos de interpretación de puntajes. En este caso hablamos de dos esquemas básicos: el modelo de interpretación con referencia a normas y el modelo de interpretación con referencia a criterios. En el caso del modelo de normas, la interpretación se basa en la *posición relativa* del puntaje específico en relación con los demás. En este tipo de modelos los puntajes no se interpretan en términos de estándares absolutos, como sí es el caso de medidas creadas bajo un modelo de criterios. Por ejemplo, en el caso del Índice de Desarrollo

Humano, sabemos que un país con un valor de 0.95 en el índice exhibe mayor desarrollo humano que uno con un valor de 0.85, pero decir por ejemplo que un país con un valor de 0.95 está cerca del Desarrollo Humano máximo, dado que el valor más alto que puede tomar el índice, por construcción, es 1, es una interpretación que requiere de más elementos teóricos y sustantivos para poder sustentarse adecuadamente.

Los índices y escalas no necesariamente presentan diferencias en términos de ser más susceptibles unos u otros para cierto tipo de interpretación, aunque por su misma naturaleza pareciera que las escalas pueden de forma más razonable pensarse más apropiadas para intentar interpretaciones de criterio, que suelen, evidentemente, ser las más exigentes.

Ejemplo 1: Índice de calidad de vida para niños con enfermedades terminales

Un equipo de profesionales, liderado por el Dr. Juan Irola del Hospital de Niños en San José Costa Rica, construyó una versión del instrumento para ser aplicada en una prueba piloto. El instrumento incluye tres dimensiones: Aspectos médicos, aspectos de la cuidadora y aspectos del medio ambiente. La prueba piloto incluyó una muestra de 63 niños, cada uno de ellos recibió puntuaciones dos veces en cada dimensión, estas puntuaciones fueron asignadas por distintos examinadores calificados profesionalmente, que además habían recibido instrucciones previas en términos de cómo usar el instrumento.

En Aspectos Médicos y Aspectos de la Cuidadora, por tratarse de escalas, cada una originalmente con 10 ítems, se aplicaron procedimientos derivados de la teoría psicométrica para establecer su grado de consistencia interna (índice de discriminación y Alpha de Cronbach). Con base en estas medidas se tomó la decisión de eliminar algunos ítems. El indicador de Aspectos del Medio Ambiente es un índice y por eso no se utilizaron métodos psicométricos. También se calculó la correlación entre los puntajes totales otorgados por los dos examinadores de cada niño como un indicador de la concordancia entre jueces, obteniéndose valores mayores a 0.8.

Se generó la calificación en cada una de las tres dimensiones dividiendo el total de puntos logrados entre el número máximo de puntos posible en el instrumento. El índice de calidad de vida se calculó haciendo un promedio ponderado de las tres dimensiones, dando doble ponderación a los aspectos médicos.

Se debatió en términos de cómo establecer los “baremos” o estándares para la calificación, si utilizando un modelo de normas o de criterios. Finalmente se decidió seguir un modelo de criterios y el equipo de profesionales, basado en sus propios conocimientos y experiencia, estableció los siguientes parámetros de interpretación absoluta para el índice: Menos de 50%: Calidad de Vida baja; 50 a 75%: Calidad de Vida media, y, más de 75%: Calidad de Vida alta. (Zúñiga & Montero, 2007).

Ejemplo 2: Un índice de educación para jóvenes en riesgo social

Como parte de una evaluación de un programa para generar empleabilidad en jóvenes de comunidades en situación de exclusión social, se deseaba construir un indicador de resultado que permitiera monitorear, a lo largo del tiempo, el comportamiento de la dimensión “Educación” en este tipo de población. Educación es uno de los componentes del constructo “Empleabilidad”, que en términos sencillos, puede definirse como todas aquellas capacidades que le dan oportunidades a una persona de insertarse laboralmente.

La población meta estuvo compuesta por jóvenes entre 13 y 18 años de comunidades de bajo nivel socioeconómico de la región central de Costa Rica que habían completado la educación primaria. Los datos fueron recolectados por medio de una encuesta aplicada a una muestra de 977 jóvenes. (Montero, 2006).

A partir de la encuesta, las preguntas que se propusieron inicialmente para incluir en el índice de educación fueron las siguientes:

1. ¿Cuál fue el último año de estudios que aprobaste?
2. ¿Cuál fue el último año de estudios que cursaste?
3. ¿En el ciclo lectivo pasado (2003) estuviste estudiando?
4. ¿En este ciclo lectivo (2004) vas a estar estudiando?

5. En general, ¿En el último período lectivo que cursaste, cuál fue tu promedio de notas en cada una de las siguientes materias? Para propósitos de control de calidad de la información suministrada por el adolescente, el entrevistador(a) preguntaba también si era posible que le mostrara sus reportes de notas, lo cual frecuentemente sucedió.

6. ¿Cuántas veces tuviste que repetir grados en primaria, en la escuela?

El procedimiento que se siguió en este ejemplo fue el siguiente:

1. *Corregir por el efecto de la edad.* Primeramente se corrigió el efecto de la edad en aquellas variables del índice que estuvieran afectadas por ésta. Esto se realizó para las dos primeras variables, años aprobados y años cursados. Lo que se hizo en estos dos casos para eliminar el efecto de la edad del adolescente fue dividir el valor de la variable entre la edad en años cumplidos del chico(a), de manera que se reflejara más claramente la dimensión que se desea estudiar: Años aprobados y cursados de educación en concordancia con la edad cronológica.

2. *Imputación de valores faltantes.* Para efectos de esta primera aproximación se trabajó imputando los valores faltantes con el promedio de la serie. Debe mencionarse que, en ningún caso, los valores faltantes representaron más de un 15% de las respuestas.

3. *Estandarizar todas las variables.* Debido a las diferentes unidades de medida y magnitudes de las variables del índice se hizo necesario estandarizarlas a todas, generando variables con promedio 0 y desviación estándar igual a 1. Como se anotó anteriormente, en general, la estandarización previa de las variables o componentes es una condición necesaria para la correcta construcción de un índice.

4. *Obtener el índice bruto como el promedio (simple o ponderado) de las variables estandarizadas.*

a. Primera aproximación, promedio simple: En este esquema se le dio exactamente la misma importancia relativa en el índice a cada una de las seis variables. Dado que se trabajó con las variables estandarizadas, el índice vino dado, sencillamente, por el promedio simple de las seis variables y de esta manera todas tuvieron el mismo peso en el índice.

b. Segunda aproximación, promedio ponderado: Debido a que, según la opinión de esta autora y de acuerdo con el marco teórico, no todas las variables tienen la misma importancia como indicadores de la dimensión “Educación”, se decidió ensayar una segunda aproximación con el siguiente esquema de ponderaciones: Variable 1 (años aprobados): 0.3, Variable 2 (años cursados): 0.15, Variable 3 (estudiando año pasado): 0.1, Variable 4 (estudiando este año): 0.1, Variable 5 (promedio de notas): 0.25 y Variable 6 (veces repitió primaria): 0.1. De tal forma que el índice vino dado por un promedio ponderado de las seis variables, reflejando el peso relativo que se deseaba dar a cada una de acuerdo con consideraciones sustantivas.

5. *Estandarizar el índice bruto.* El último paso en el proceso implicó estandarizar el índice bruto, de manera que el promedio fuera igual a cero y la desviación estándar igual a 1. De esta manera se facilitan

enormemente las interpretaciones, pues se sabe que los valores particulares del índice nos indican a cuántas desviaciones estándar, por encima o por debajo del promedio, se encuentran los individuos específicos. Como es evidente, en este caso se buscaba una interpretación referida a normas. La tabla 2 presenta las estadísticas descriptivas de las variables del índice, aún sin estandarizar.

Tabla 2
Estadísticas descriptivas de las variables del índice (sin estandarizar)

	Frecuencia	Mínimo	Máximo	Media	Desviación estándar
Años aprobados	972	,06	,72	,479	0,083
Años cursados	967	,06	,72	,502	0,078
Qué hizo ciclo lectivo 2003	977	,0	1,0	,845	,362
Qué hizo ciclo lectivo 2004	977	,0	1,0	,832	,374
Promedio de notas	839	30,6	100,0	75,962	8,900
Grados repetidos en primaria	977	,0	4,0	,377	,660
Toral	833				

Desde el punto de vista de la aplicación de muchos de los procedimientos de análisis estadístico, es conveniente que la distribución de los puntajes del índice se aproxime al modelo de la distribución normal. (Tabachnick & Fidell, 2001). Es claro que la distribución observada de los puntajes en la primera aproximación (Figura 1a) se desvía del modelo normal.

Seguidamente se presenta el histograma correspondiente a la segunda aproximación (Figura 1b). En éste se nota un comportamiento de los puntajes más cercano al del modelo normal. Aunque las ponderaciones en este caso se eligieron pensando más bien en criterios teóricos y sustantivos, es claro que también contribuyen a darle la “forma” deseada a la distribución del índice. Por todo lo anterior se prefiere entonces trabajar con el sistema de ponderaciones propuesto en la segunda aproximación.

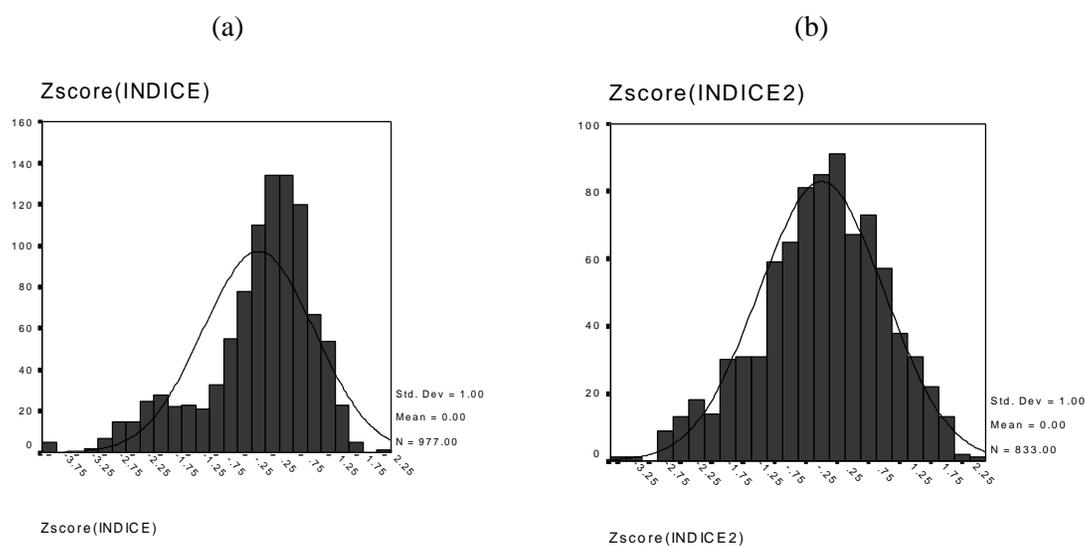


Figura 1. Histogramas del índice estandarizado. (a) Primera aproximación. (b) Segunda aproximación

Conclusión

Finalmente, a modo de conclusión general, debemos resaltar la importancia que reviste para los investigadores el conocimiento a profundidad del marco conceptual de referencia y las diversas opciones metodológicas que existen para la construcción y validación de índices y escalas. Debería quedar claro, a partir de lo mostrado en este artículo, que estas decisiones no pueden tomarse superficialmente. Aunque hay una variedad de metodologías y técnicas disponibles, la escogencia de cualquiera de ellas debe basarse en una decisión informada y fundamentada, que considere igualmente, el propósito o propósitos de la medida que se desea generar o evaluar.

Referencias

- Anastasi, A. & Urbina, S. (1998). *Tests Psicológicos*. México: Prentice Hall Hispanoamericana.
- Babbie, E. (2000). *Fundamentos de la Investigación Social*. México: Thomson.
- Cortada de Kohan, N. (2001, Diciembre). Constructos psicológicos y variables. *Psicología y Psicopedagogía*. Publicación virtual de la Facultad de Psicología y Psicopedagogía de la USAL, 2(2). Obtenido de <http://www.salvador.edu.ar/psic/ua1-9pub02-8-03.htm>
- Gómez, M. (1997). *Elementos de Estadística Descriptiva*. San José, Costa Rica: EUNED.
- Kerlinger, F.N. (1988). *Investigación del Comportamiento*. México: McGraw-Hill.
- Martínez, M. (2005). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis S.A.
- Montero, E. (2006). *Outcome Evaluation of the Project "Stop Exclusion, Create Value: Final Report*. Informe final de investigación. San José Costa Rica: Centro Centroamericano de Población, Universidad de Costa Rica.
- Montero, E.; Villalobos, J. & Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: un análisis multinivel. *RELIEVE: Revista Electrónica de Investigación y Evaluación Educativa*. 13(2), 215-234. Obtenido de http://www.uv.es/RELIEVE/v13n2/RELIEVEv13n2_5.pdf
- Nunnally, J. C. & Bernstein, I. J. (1995). *Teoría Psicométrica* (3ª ed). México, D.F.: McGraw-Hill Latinamericana.
- Tabachnick, B. G. & Fidell, L. S. (2001) (4th ed.). *Using Multivariate Statistics*. Boston MA: Allyn & Bacon.
- United Nations Development Programme (2007). *Human Development Report 2007/2008. Fighting climate change: Human solidarity in a divided world*. New York: Palgrave Macmillan.
- Zúñiga, M. & Montero, E. (2007). *Teoría G: un futuro paradigma para el análisis de pruebas psicométricas*. *Actualidades en Psicología*. 21(108), 117-144.

Manuscrito recibido en Octubre de 2006
Aceptado para publicación en Septiembre de 2007