

### *La Escala de Likert en la evaluación docente: acercamiento a sus características y principios metodológicos<sup>1</sup>*

Angélica María Fabila Echaury<sup>2</sup>

Hiroe Minami<sup>2</sup>

Manuel Jesús Izquierdo Sandoval<sup>2</sup>

(Recibido: noviembre de 2012, Aceptado enero de 2013)

#### RESUMEN

Las diversas políticas educativas internacionales y nacionales implementadas en los últimos años han propiciado un aumento el número de estudios que analizan la calidad de la práctica educativa de los profesores. Un instrumento empleado frecuentemente, para identificar las percepciones de los estudiantes, profesores y administradores en relación a la práctica educativa lo constituye el cuestionario de escala Likert. Este escrito describe los principios conceptuales y metodológicos que deben sustentar el diseño e implementación de este instrumento en la investigación educativa. Para tal efecto, primero, se realiza un análisis de diversas investigaciones en el ámbito educativo que han empleado el cuestionario Likert para analizar la práctica de los profesores. Segundo, se revisan las bases conceptuales y metodológicas para el diseño del cuestionario y se describen los procedimientos cuantitativos que pueden emplearse para su validación. Finalmente, se revisa la forma en la que las investigaciones, realizadas en torno a práctica educativa, han observado los principios conceptuales y metodológicos descritos en el artículo.

#### ABSTRACT

The educational policies implemented in the last years, at the national and international level, have prompted a growth in the number of research studies that examine the quality and dimensions of the educators' instructional practices. Often, these studies have built upon the implementation of Likert scale questionnaires to examine administrators, educators', and learners' perceptions of the instructional practices of the teachers. This paper describes the conceptual and methodological principles to be observed during the design and implementation of a Likert-scale questionnaire. First, some studies that have

implemented this data collection instrument, to examine educators' instructional practices, are reviewed. Then, key conceptual and methodological issues that underpin the design and implementation of Likert-scale questionnaires are provided. Further, the quantitative procedures that can be used to validate these questionnaire type are described. Finally, the manner in which previous educational research has observed the outlined conceptual and methodological issues is discussed.

#### INTRODUCCIÓN

La evaluación del desempeño docente es un área de la educación que ha ganado interés en las últimas dos décadas, en el ámbito de la política educativa nacional e internacional. Investigadores e instituciones han propuesto diversas perspectivas y formas de acercamiento tanto cualitativa como cuantitativamente, a la evaluación del desempeño docente, a fin de hacer frente a esta actividad, que implica grandes retos para quienes la diseñan y realizan (ver Aquino, Izquierdo, y Echaz, 2013). Uno de los asuntos inacabados respecto a la evaluación docente es la discusión sobre los medios e instrumentos que pueden emplearse para poner en marcha tal proceso de evaluación.

Más aun, la revisión de literatura sobre la evaluación del desempeño docente sugiere que aún no existe consenso en lo concerniente al concepto que define a un

<sup>1</sup> El presente trabajo se realizó en el marco del proyecto "Diagnóstico de las competencias docentes de los profesores de inglés en las secundarias públicas del estado de Tabasco a través de un sistema de evaluación triangular," financiado por CONACYT-Fondo Mixto Tabasco, con clave TAB-2010-C19-144479. Parte de los resultados aquí publicados fueron presentados en el FEL 2011, Foro de Estudios en Lenguas, Internacional realizado en la Universidad de Quintana Roo.

<sup>2</sup> Profesores-Investigadores de la Universidad Juárez Autónoma de Tabasco. División Académica de Educación y Artes. Contacto: angelifa@hotmail.com, hrom57uk@yahoo.com.mx, jesus.izquierdo@mail.mcgill.ca

“buen profesor” (Rueda y Nieto, 1996). Por lo que los instrumentos de evaluación incluyen y ponderan diversas variables e indicadores. No obstante, la evaluación tiene como constante involucrar a los actores del proceso, quienes emiten juicios con respecto al desempeño docente. Así la evaluación, considerada como un conjunto de acciones sistemáticas que tienden a lograr un conocimiento sobre la realización de la actividad docente, involucra con frecuencia a estudiantes, pares y a los mismos docentes.

Con la finalidad de identificar los factores y las dimensiones que han sentado las bases de la evaluación docente en los distintos ámbitos y contextos educativos, los profesores investigadores y estudiantes de la Universidad Juárez Autónoma de Tabasco participantes en el estudio “Diagnóstico de la competencias docentes de los profesores de inglés en secundarias públicas de estado de Tabasco, a través de un sistema de evaluación triangular”, realizaron una pesquisa de estudios relacionados con el desempeño docente.

La revisión de literatura sobre la evaluación del desempeño docente referida, comprendió tanto la búsqueda abierta, mediante el uso de herramientas disponibles en la Internet, como en revistas especializadas en el área de la enseñanza de lenguas, lográndose recopilar 28 artículos. De estos, 22 reportan resultados de estudios empíricos. En relación a los participantes, en 8 de los 22 estudios revisados, los informantes fueron estudiantes y profesores; en 6 estudios, se trabajó sólo con profesores; otros 6 estudios recolectaron datos, sólo estudiantes; finalmente, dos estudios incluyeron directivos para evaluar el desempeño docente.

Con respecto a los instrumentos de recolección de datos, la revisión de literatura demostró que el 70% (16 estudios de 22) de las investigaciones analizadas hacen uso de la escala de Likert ya sea de manera única o combinándola con otras técnicas. Debido a la preferencia por el uso del cuestionario tipo escala Likert en los estudios recuperados, la revisión se concentró en los 16 estudios que hacen uso de este instrumento. Doce de los 16 estudios se basan en el uso exclusivo de este instrumento como los casos de Arias García (1984), Bell (2005), Casero Martínez (2008), García Garduño (2003), Hammadou Sullivan (2004), Johnson (1994), Kyriakides et al. (2002), Kyriakides et al. (2006), López-Barajas y

Ruiz Carrascosa (2004), Patrick y Smart (1998), Spooren, Mortelmans y Denekens (2007), y Yonghong y Chongde (2006).

Diversas son las argumentaciones que esgrimen los estudiosos para justificar el uso de la escala Likert para la evaluación. Spooren, Mortelmans y Denekens (2007) argumentan su selección de la escala Likert con base a su facilidad de uso, ya que los resultados pueden ser transformados en porcentajes; su susceptibilidad para realizar la prueba de la confiabilidad mediante el Alpha de Cronbach y la sencillez de interpretación este método estadístico que permite difundir con facilidad el resultado a los usuarios. La facilidad de uso a la que hacen referencia los argumentos de preferencia de la prueba está ampliamente relacionada con la etapa de construcción de la escala, cuyo proceso requiere ser puntualizado.

El objetivo de esta publicación es, en primer lugar, realizar una revisión de los aspectos metodológicos que caracterizan el diseño e implementación de este instrumento de recolección de datos para obtener conocimiento y valoraciones sobre la actividad desempeñada por los profesores a través de la evaluación docente. En segundo lugar, ilustrar cómo estos aspectos metodológicos han sido combinados en diversos estudios para dar solidez metodológica al proceso de la evaluación docente.

## **NATURALEZA Y CARACTERIZACIÓN DE LA ESCALA DE LIKERT**

La recopilación de datos es una etapa esencial para el desarrollo de una investigación; para que dicha etapa se efectúe de manera exitosa, se requiere de un plan estratégicamente diseñado para reunir datos pertinentes sobre los atributos, conceptos y/o variables involucradas en el objeto o situación estudiada. Namakforoosh (2000) identifica tres acciones básicas a través de las que se obtendrá información, datos o respuestas en una investigación: acciones de observación, interrogatorio y simulación.

La escala Likert, de acuerdo con la clasificación de Namakforoosh (2000) puede ubicarse como modalidad del método de interrogatorio. Este método de recopilación de datos es el más conocido especialmente en ciencias sociales y humanas, donde frecuentemente el objeto de estudio lo constituyen seres humanos, quienes tienen la información necesitada o pueden ofrecer testimonios

sobre el asunto estudiado. El interrogatorio puede adoptar diversas formas y realizarse a través de variadas técnicas: Sesiones de grupo, encuestas, entrevistas personales o de grupo, entre otros; lo que hace del interrogatorio una acción de mucha valía para investigaciones de diseño tanto cualitativo como cuantitativo en áreas como la sociología, la psicología (especialmente la social), la educación, la ciencia política, el estudio de mercados, entre muchos otros ámbitos de estudio de las ciencias sociales (Namakforoosh, 2000).

Este método de recolección posibilita la obtención de información variada sobre los sujetos, desde datos muy objetivos y específicos como edad, estado civil, lugar de origen, etc., hasta información más compleja y subjetiva que requiere una mayor elaboración de la persona interrogada, como son sus percepciones, actitudes, representaciones, preferencias, opiniones, a las que se accede a través de "lo dicho" o la expresión de la persona. Una herramienta indispensable para articular el método de interrogatorio en una investigación en cualquiera de sus modalidades, es el cuestionario; definido por Brown (2001:6) en Mackey y Gass (2005:92) como "*cualquier instrumento escrito que presenta a los respondientes una serie de preguntas u oraciones a las que tienen que reaccionar ya sea escribiendo sus propias respuestas o seleccionándolas de entre un número existente de respuestas*".

Sierra Bravo (2001) distingue entre el cuestionario simple, el cuestionario de la entrevista y las escalas socio-métricas. El primero de ellos se refiere al cuestionario que, previa lectura, los encuestados contestan por escrito sin la intervención directa de quienes realizan la investigación. En la entrevista el cuestionario es aplicado a los sujetos por encuestadores (personas que han sido capacitados para la aplicación) y quienes registran las respuestas. Las escalas socio-métricas, por su parte son una forma especial de cuestionarios cuya característica es que las respuesta a las preguntas tienen asignado un valor numérico, lo que permite codificar la información recabada y cuantificar o medir las respuestas obtenidas.

*Las escalas sociométricas son, en las ciencias sociales, instrumentos de medida constituidos por la disposición conjunta en un cuestionario de distintos elementos o aspectos de una caracte-*

*ristica o actitud a los que se designa convencional o empíricamente un valor numérico, lo que permite cuantificar el grado de esa característica o actitud que se da en un objeto o sujeto determinado (Sierra, 2001: 370)*

La medición puede llevarse a cabo en objetos y cosas concretas; en ciencias sociales y humanas, la medición se puede centrar en cuestiones de mayor complejidad como preferencias, actitudes, opiniones. Aunque el tipo o forma de medición dependerá del objeto o asunto a medir, la designación de escala responde, en su definición más común, a un instrumento que presenta una sucesión de medidas que permiten organizar datos en orden jerárquico. Así, las escalas socio-métricas son útiles y aplicables para medir características diversas; entre las más conocidas y usadas son aquellas que se emplean para medir el grado en que se da una actitud o disposición de ánimo permanente, respecto a cuestiones específicas, llamadas escalas de actitud. En el contexto de las actitudes se reconocen tres tipos básicos de escalas: diferenciales, sumativas y acumulativas, mismas que son también reconocidas por el nombre de sus autores: Thrustone (diferenciales) Likert (sumativa también llamada aditiva) y Guttman (acumulativa). Las escalas antes mencionadas están formadas por series de ítems y en todos los casos la puntuación final de cada sujeto es la suma de todas sus respuestas. La finalidad última es situar el continuo de actitud o rasgo medido.

La escala de Likert es una escala aditiva con un nivel ordinal (Namakforoosh, 2000), constituida por una serie de ítems ante los cuales se solicita la reacción del sujeto. El interrogado señala su grado de acuerdo o desacuerdo con cada ítem, proposición o afirmación relativa al asunto estudiado; a cada posible respuesta se le da una puntuación favorable o desfavorable. La suma algebraica de las puntuaciones de las respuestas que el individuo hace al conjunto de ítems da su puntuación total, por lo que se entiende como representativa de su posición favorable-desfavorable con respecto al fenómeno que se mide.

Una característica de los ítems en la escala de Likert es que las alternativas de respuesta son fijas para todas las proposiciones y todas tienen designado un peso o valor similar o equivalente. Así, la probabilidad de acuerdo o desacuerdo con cualquiera de las series

de ítems favorables o desfavorables, con respecto a un objeto, varía directamente con el grado de actitud de un individuo. Un individuo con una actitud favorable responderá favorablemente a muchos ítems (es decir, estará de acuerdo con muchos ítems favorables al objeto y disientirá a los desfavorables); de un individuo ambivalente puede esperarse que responda desfavorablemente a unos y favorablemente a otros; un individuo con una actitud desfavorable responderá desfavorablemente a muchos ítems.

### CONSTRUCCIÓN DE UNA ESCALA DE LIKERT

Este método de medición de actitudes fue desarrollado en los años treinta por Rensis Likert en 1932 y aunque ya es un método antiguo, se trata de un enfoque vigente y bastante popular, cuya constitución observa varias etapas (Namakforoosh, 2000; Elejabarrieta e Iñiguez, 2008, Méndez y Peña, 2007). Para empezar la construcción de la escala es necesario identificar claramente la variable o actitud que se intenta medir. Para esta formulación se pueden atender dos recomendaciones que hacen Elejabarrieta e Iñiguez (2008). La primera es intentar lograr una definición que contenga los aspectos más relevantes del objeto de actitud. Una vez dado este paso, en segundo lugar, habrá que saber qué campo o campos abarca o toca el objeto de actitud medido, es decir, cuales son los límites de representación de ese objeto.

Una vez identificado el objeto o variable a medir, el proceso de construcción continúa con la etapa de recopilación o elaboración de ítems. Un ítem es una frase o proposición que expresa una idea positiva o negativa respecto a un fenómeno que nos interesa conocer. En esta etapa se trata de recopilar una serie de ítems que expresen un amplio rango de actitudes, desde extraordinariamente positivas hasta extraordinariamente negativas (50%-50%). Cada ítem, recomienda Namakforoosh (2000), debe preguntarse en tiempo presente, expresar una sola idea, ser breve, [Hernández, Fernández y Baptista (2007) puntualizan: no exceder de 20 palabras] ser relevante y estar elaborado de manera que permita aprobar o rechazar las actitudes.

Por su parte Morales Vallejo (2006) reconoce como normas para la redacción de ítems:

- Utilizar expresiones sencillas, fácilmente comprensibles para quienes van a responder y que no puedan interpretarse de diversas maneras.
- Evitar las dobles negaciones.
- Evitar expresiones universales como siempre, nunca, nadie y similares.
- No incluir expresiones que incluyan dos afirmaciones u opiniones ya que se puede estar de acuerdo con una parte del ítem y no con la otra.
- Los ítems deben incluir opiniones no hechos comprobables; cada proposición debe ser opinable y debatible evitando proposiciones con las que todos o casi todos estén de acuerdo. Esta característica de los ítems de la escala de Likert es ampliamente valorada para los casos de evaluaciones, ya que la valoración de los sujetos sobre un hecho evento o persona es opinable y discutible.
- La escala requiere de tantos ítems como sean necesarios para cubrir toda la gama que va desde los muy desfavorables al objeto, hasta los muy favorables, ya que como se ha dicho anteriormente, la suma de respuestas, es lo que permite decidir la posición que una persona ocupa en el hipotético continuum de la actitud. En la construcción del instrumento una premisa importante a considerar es que cada ítem o elemento proporciona una información insuficiente pero necesaria para conocer la posición del respondiente respecto al asunto explorado.

Los enunciados que constituirán los ítems de la escala pueden extraerse por diferentes métodos:

- A.– Con ayuda de literatura sobre el tema; es decir, los ítems se construyen a partir de una revisión de la literatura relacionada con el tema en cuestión.
- B.– Entrevista previa; en este caso, los ítems resultan de las opiniones vertidas por persona(s) que consideremos representativas de la población que vamos a estudiar.
- C.– La propia intuición y "genio" del investigador.

Los enunciados o ítems se presentan con una escala de estimación (*rating scale*) que consiste en una graduación que va desde "totalmente de acuerdo" hasta

"totalmente en desacuerdo", incluyendo grados intermedios, con respecto a la afirmación. Autores como Méndez y Peña (2007) consideran que en efecto cada ítem es una sentencia sobre la cual debe expresarse el grado de acuerdo o desacuerdo; sin embargo, cuando se está evaluando la presencia de una variable, las opciones de respuesta cambian de categorías de acuerdo o desacuerdo a categorías de frecuencia, lo que permite identificar la frecuencia con que se presenta el comportamiento. En cualquier caso la escala de estimación incluye un valor numérico asignado a cada intervalo gradual, cuyo significado debe explicarse al sujeto previamente.

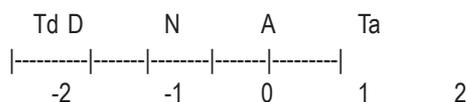
Las categorías de respuesta para los ítems pueden variar en su número y valor asignado. Por lo general, se usan cinco expresiones fijas para cada ítem (Hernández, Fernández y Baptista, 2007; Namakforoosh, 2000; Méndez, 2007) aunque pueden usarse entre dos y siete alternativas de respuesta. Cuando las opciones de respuestas se presentan en número impar se debe repartir la carga positiva y negativa de manera equilibrada por lo cual es necesaria una categoría neutra. En una escala con un número par de alternativas se elimina la opción o categoría neutral o intermedia, para comprometer al sujeto o forzarlo a que se pronuncie de manera favorable o desfavorable (Hernández, Fernández y Baptista 2007); a esta escala se le conoce también como escala de opción forzada.

Los puntos de vista sobre el número de categorías de respuesta son variados y depende de la elección del investigador, aunque la decisión deberá basarse en las características de los sujetos participantes potenciales, especialmente en su capacidad de discriminación. Hernández, Fernández y Baptista (2007) recomiendan que si los participantes tienen poca capacidad para discriminar se pueden considerar dos o tres categorías. Por el contrario si son personas con un nivel educativo elevado y gran capacidad de discriminación pueden incluirse hasta 7 categorías.

Como se mencionó anteriormente, a cada categoría o alternativa de respuesta se le asigna una puntuación o valor numérico, que puede ir de 1 a 5, si fueran 5 las alternativas de respuesta, o bien partir del cero. Según Hernández, Fernández y Baptista (2007), las proposiciones pueden tener dirección: favorable o positiva y desfa-

vorable o negativa. La dirección es muy importante para saber cómo se codifican las alternativas de respuesta. Las puntuaciones más altas, regularmente, se asignan a actitudes favorables o positivas frente al ítem mientras que las desfavorables reciben puntuaciones bajas. Sin embargo, cuando la dirección de los ítems es desfavorable hay que asignarlas puntuaciones de manera inversa.

Autores como Creswell, citado por Hernández, Fernández y Baptista (2010) y Elejabarrieta e Iñiguez (2008) consideran la posibilidad de asignar valores positivos y negativos a las categorías de respuesta partiendo del centro o la posición neutra, como se ilustra:



La notación numérica puede variar, lo que siempre se conserva es que las respuestas se codifican con números íntegros sucesivos, es decir que mantienen una progresión aritmética.

Las expresiones más comunes en las alternativas de respuesta son para manifestar acuerdo o desacuerdo:

Totalmente de acuerdo o muy de acuerdo  
 De acuerdo.  
 Ni de acuerdo, ni en desacuerdo, neutral o indeciso.

En desacuerdo  
 Totalmente en desacuerdo o muy en desacuerdo.

Para expresar frecuencia:  
 Siempre.  
 Casi siempre.  
 A veces.  
 Casi nunca.  
 Nunca.

Las opciones de respuesta pueden colocarse de dos formas: de manera horizontal o de manera vertical. Una condición para la asignación del número de categorías de respuesta, es que tal número no varíe para todas las afirmaciones, respetando el mismo orden o jerarquía de presentación de las opciones para todas las frases.

La presentación de los ítems en la escala debe evitar la agrupación de ítems relativos a un tema. Una buena

opción para integrar los ítems a la escala es seleccionar un ítem de cada aspecto a evaluar (u objetivos) y combinarlos alternando uno por cada aspecto (Méndez y Peña, 2007). Se debe evitar también la secuencia de ítems, especialmente cuando se trabaja con ítems afirmativos y negativos. Estas sencillas estrategias permiten obtener respuestas más personales y eliminan sesgos especialmente por fatiga o por la intuición del participante sobre lo que debe, o es adecuado responder.

Otra etapa en la construcción de la escala es la conocida como escala-piloto, en ésta se distribuyen los ítems a una muestra de personas seleccionada al azar con características similares a quienes se les aplicara la escala final. Tras aplicar la escala-piloto a los sujetos, se obtiene la puntuación global para cada sujeto, resultado de la suma de las puntuaciones equivalentes a las respuestas elegidas. Esta puntuación global permite estimar la posición que el sujeto ocupa en el continuum hipotético de actitud. Lo que permitirá dar inicio a la cuarta etapa, conocida como determinación de la validez y fiabilidad de la escala.

### FIABILIDAD Y VALIDEZ DE LA ESCALA

Una vez aplicada la escala piloto, es necesario determinar las bases para la elección de los ítems para la escala final. Estas bases pueden obtenerse de dos formas: una es el método de consistencia interna y otra forma es el método de análisis de ítems. Ambos métodos tienen como finalidad encontrar ítems que consistentemente separen a los sujetos que tienen una actitud alta “muy favorable” de los que tienen una actitud baja o “de sfavorable”(Namasforoosh,2000). Aquellos ítems que reciben respuestas favorables por parte de individuos que, como grupo, no responden a la mayoría de los otros ítems de forma favorable (o viceversa) se descartan, considerándose que no detectan las mismas actitudes que los otros ítems.

El primer método de análisis de ítems (Namasforoosh, 2000) también llamado de grupos extremos (Elejabarrieta e Iñiguez, 2008), se vale de las puntuaciones globales de los sujetos participantes para estimar el Poder Discriminativo de los ítems en base a la comparación de las varianzas de las medias a través de la T de Student; este método ha sido desestimado por algunos autores por considerarlo inaplicable por la naturaleza ordinal de los

datos obtenidos vía la escala de Likert (véase Romano, Kromrey, Coraggio, y Skowronek, 2006).

El segundo método de análisis es conocido como método de consistencia interna, también llamado método de correlación ítem-test (Morales, 2006; Elejabarrieta e Iñiguez, 2008). Este método calcula la correlación de cada ítem con la suma de todos los demás para establecer la consistencia interna de los ítems (fiabilidad). La consistencia interna se establece a partir de la prueba Alfa de Cronbach (índice de consistencia interna) que presenta valores entre 0 y 1; los valores cercanos a la unidad son aceptables ya que indican que se trata de un instrumento fiable, es decir que sus mediciones son estables y consistentes.

Existen dos formas de calcular el índice de consistencia interna: mediante la varianza de los ítems que se obtiene al aplicar la siguiente fórmula:

$$\alpha = \left[ \frac{K}{K-1} \right] \left[ 1 - \frac{\sum_{i=1}^K S_i^2}{S_t^2} \right],$$

Donde

$S_i^2$  es la varianza del ítem i,

$S_t^2$  es la varianza de la suma de todos los ítems y

K es el número de preguntas o ítems.

(Morales, 2006)

O bien mediante las correlaciones entre los ítems, cuyo cálculo obtiene con de la siguiente forma:

$$\alpha = \frac{np}{1 + p(n-1)},$$

Donde

n es el número de ítems y

p es el promedio de las correlaciones lineales entre cada uno de los ítems.

(Morales, 2006 )

El proceso se simplifica si se corre a través de un paquete estadístico como el *Statistical Package for the Social Sciences* (SPSS). Éste permite calcular no sólo la consistencia interna de la escala, sino también identificar las variables o ítems que aumentan o disminuyen la consistencia interna. Esta información es útil para identificar elementos que puedan ser excluidos de la escala bajo un criterio cuantitativo; aunque es recomendable que esta decisión se haga también valorando aspectos cualitativos del ítem a partir del análisis y valoración del investigador.

El índice Alfa de Cronbach posibilita la estimación de consistencia interna de un test cuando éste está formado por un conjunto de ítems que se combinan aditivamente para hallar una puntuación global; es decir, cuando la puntuación que se interpreta es la global. Tal como sucede en la escala de Likert. Cuando se hace el cálculo a través del procedimiento de correlaciones entre ítems es necesario también que todos los ítems que integran el test midan la característica deseada en una sola dirección; es decir, las alternativas de respuesta deben ir en un mismo sentido, de ahí la necesidad de adecuar la asignación de valores a las proposiciones favorables y desfavorables.

Posterior a la prueba de Alfa de Cronbach, la validación del instrumento se completa con el análisis factorial. El análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de ellas. Estos grupos homogéneos se forman con las variables que se correlacionan mucho entre sí (Guisande, 2006). Aplicar un análisis factorial exploratorio a los resultados de la encuesta piloto permite encontrar grupos de variables con significado común. El análisis factorial permite identificar las dimensiones o factores que están siendo medidos por los ítems de la escala o test y los agrupa, asociando los ítems con los factores a los que "cargan".

El análisis factorial puede ser exploratorio cuando no se ha determinado aun cuáles son los factores que presentes en la escala, o confirmatorio cuando de acuerdo con la teoría subyacente del estudio se han identificado previamente los factores a medir. El análisis factorial con extracción de componentes principales se usa cuando los factores son independientes; mientras que si estos son relacionados la extracción es a través de la máxima verosimilitud. Con el análisis de componentes principales, un número relativamente pequeño de componentes explica la mayor parte de la variación total de las variables originales. De tal manera que los últimos factores o componentes, que explican menos la variación, pueden ser eliminados con pérdidas mínimas de información (Guisande, 2006).

### **TRIANGULACIÓN DE LOS RESULTADOS OBTENIDOS EN LA ESCALA LIKERT**

En los estudios de la evaluación docente revisados, se detectó que, de los 12 estudios que se basan en el

uso exclusivo del cuestionario tipo Likert, cuatro optaron por recoger los datos en más de un grupo de personas. Esto fue con la finalidad de realizar la validación de sus resultados a través del proceso de triangulación de la información. Ésta permite realizar acercamientos a la realidad tomando en cuenta una diversidad de puntos vistas. La triangulación se puede definir como el uso de dos o más métodos para la recolección de datos en investigaciones sobre distintos aspectos del comportamiento humano (Cohen y Manion, 1994: 233). A partir de la literatura revisada, se describe a continuación el empleo de la escala Likert para medir el desempeño docente. En la descripción, se ponen en relieve principalmente los procesos empleados para validar el cuestionario Likert.

Luna Serrano y Rueda Beltrán (2001) utilizaron los datos cuantitativos y cualitativos, combinando un cuestionario de tipo Likert con entrevistas semi-estructuradas. En el caso de Arkan et al. (2008), se solicitaron descripciones del profesor eficaz mediante dos preguntas abiertas. Posteriormente se administró un cuestionario Likert elaborado en base al resultado anterior, con una escalade cuatro puntos (escala forzada). Al-Mutawa (1997) combinó una valoración en salones de clase por evaluadores externos y una auto-evaluación.

Arias García (1984) solicitó a varios grupos de estudiantes de semestres iniciales, intermedios y avanzados, una lista de características y factores sobre los cuales consideraran que debieran ser evaluados los profesores. Se elaboró un cuestionario con base en la información obtenida y se aplicó a estudiantes de grupos diferentes a los que se habían obtenido los factores y los comportamientos descriptivos. El mismo cuestionario fue aplicado también a los profesores. En el caso de Luna Serrano y Rueda Beltrán (2001), los autores triangularon la recogida de datos con la participación de estudiantes y profesores universitarios, además del uso de dos instrumentos: un cuestionario de tipo Likert y entrevistas semi-estructuradas.

### **USO DE ESCALA LIKERT EN ESTUDIOS DE EVALUACIÓN DOCENTE**

En los 11 estudios que emplearon el cuestionario Likert, se identificaron diversas combinaciones de los procedimientos metodológicos presentados anteriormente con respecto a la construcción de los cues-

tionarios, la validación de los ítems, y la triangulación. A continuación se ejemplifican algunos casos.

Kyriakides *et al.* (2002) realizaron una entrevista grupal a 14 maestros de una escuela primaria de Cipro en busca de las características de un profesor efectivo. Con base en los resultados, se elaboró un cuestionario con 51 características, con una escala de importancia de siete puntos (1= importancia mínima y 7= importancia máxima). El instrumento se administró al mismo grupo de maestros y se realizó un análisis de conglomerados, mediante el cual se identificaron ocho categorías. Se determinó la ponderación de los encuestados sobre las categorías (media mayor de 5.50). Asimismo, estos resultados fueron presentados a los maestros para su confirmación.

En la segunda etapa, se hizo un estudio a nivel nacional administrando el instrumento a un 20% de maestros de primaria de Cipro seleccionados aleatoriamente. Se realizó la validación del cuestionario mediante dos procedimientos estadísticos. Primero, se verificó Alfa de Cronbach para determinar la congruencia interna entre los ítems del cuestionario (para una mayor información, consulte Larson-Hall, 2010: 170-175). El segundo procedimiento consistió en un análisis factorial exploratorio para identificar las categorías conceptuales en las cuales los ítems se agrupaban (para mayor información, consulte Field, 2000: 619-679). En este caso, se practicó el Análisis de Componentes Principales como método de extracción. Referente al primer análisis los resultados arrojaron un alfa de 0.87. En el segundo análisis los investigadores observaron que los ítems se agrupaban en siete categorías, en las que se obtuvo un alfa entre 0.70 y 0.83. El estudio finalmente reporta la correlación entre los ocho criterios generados en el contexto de una institución y las características del profesor eficaz registradas en la literatura por un lado, y por el otro, estos ocho criterios y los resultados derivados de la muestra representativa nacional.

Patrick y Smart (1998) organizaron un meta-inventario de 72 ítems sobre las características de un profesor efectivo, con una escala de cinco puntos. Para construir éste, se utilizaron las opiniones de los estudiantes y los ítems cuantitativos tomados de otros inventarios existentes. El instrumento fue administrado a 266 estudiantes

de la carrera de psicología. Los investigadores procuraron, antes que nada, la pertinencia de someter los datos recogidos a los procedimientos estadísticos, revisando la normalidad de los mismos y transformándolos. En consecuencia, sólo 46 ítems fueron analizados mediante la prueba de alfa de Cronbach y el análisis factorial. El instrumento obtuvo un alfa de 0.97. Y los ítems agruparon en tres categorías. En seguida, se practicó el análisis de varianza, el cual permitió reducir los 46 ítems hasta 24.

La investigación de Spooren, Mortelmans y Denekens (2007) se basó en un cuestionario tipo Likert creado para evaluar los docentes en la Universidad de Antwerp, Holanda, con la intención de mejorar la calidad de la instrucción y facilitar la toma de decisiones para la basificación de los profesores. La versión inicial del cuestionario contaba con 165 ítems derivados de una revisión de la literatura y otros instrumentos existentes. Estos ítems estaban organizados en ocho dimensiones, divididas en 22 sub-dimensiones. El instrumento creado se aplicó en siete cursos (433 estudiantes). Se estimó el alfa de Cronbach del cuestionario para cada curso. Posteriormente, a partir de un análisis factorial exploratorio, se verificó la validez y la confiabilidad, lo que permitió reducir a 31 el número de ítems y a 10, el número de subdimensiones. En la tercera fase, la nueva versión del cuestionario se aplicó a 566 estudiantes, quienes evaluaron ocho cursos en dos momentos con un intervalo de una o dos semanas. Para revisar la confiabilidad de los ítems en ambos momentos, se creó una tabla de frecuencia para cada ítem en donde los resultados de la primera aplicación fueron cruzados con los de la segunda aplicación. Finalmente se calculó Kappa de Cohen, Kendall's Tau B y coeficiencia de correlación de Spearman. Estos análisis demostraron que los 31 ítems y las 10 dimensiones eran confiables.

Al-Mutawa (1997) elaboró un formato para evaluar las competencias de los profesores de Inglés en las primarias de Kuwait. Este instrumento se utilizó en dos momentos: valoración por cuatro evaluadores externos y auto-evaluación de los profesores mismos. La última versión del instrumento fue sometido a la revisión de los especialistas de educación y se determinó la confiabilidad del instrumento mediante la prueba de alfa de Cronbach (0.85) y la confiabilidad entre los evaluadores (0.97).

Arias García (1984) elaboró un cuestionario acerca de las características y factores sobre los cuales debi-

eran ser evaluados los profesores. Este cuestionario se construyó a partir de las opiniones de los estudiantes. Tras la primera revisión, se redujeron a 11 factores con 198 ítems. El instrumento se aplicó a 219 estudiantes y 50 profesores, quienes evaluaron los ítems en relación a los factores de interés. A partir de esta evaluación, se conservaron 10 factores y 50 ítems con los cuales se obtuvo más del 70% consenso entre los respondientes y un 50% en términos de la clasificación de los factores. Para verificar la efectividad del cuestionario para diferenciar entre buenos profesores y los regulares, el investigador pidió a estudiantes que evaluaran 27 buenos catedráticos y 16 regulares. Los resultados de los dos grupos de catedráticos, se analizaron utilizando la prueba de Mann-Whitney; la consistencia interna se verificó a través del coeficiente de Alfa.

## CONCLUSIÓN

La escala Likert representa una alternativa valiosa para la recolección de datos en investigaciones cuantitativas que pretenden obtener información sobre las predisposiciones, actitudes, valoraciones y opiniones y que tiene una población sobre un asunto en particular. Una consideración generalizada en torno a las escalas Likert es que es una escala de fácil construcción y aplicación que permiten medir o registrar información de compleja sobre los individuos a partir de la acumulación de respuestas sobre un tópico. La escala es un instrumento que soporta varias formas de validación lo que la hace una prueba confiable una vez sometida a este proceso.

Los cuestionarios tipo escala Likert son ampliamente utilizados en materia de evaluación docente ya que ésta suele exigir la realización de valoraciones al desempeño de los profesores desde diversos aspectos tales como: el conocimiento de la materia, el manejo de contenidos y organización del curso, capacidad y actitudes para atender estudiantes, entre otros sobre los que los sujetos pueden expresar sus opiniones. De acuerdo con los estudios revisados sobre evaluación de desempeño docente, existe uso preponderante de la escala tipo Likert como único instrumento de valoración o en combinación con otras técnicas. Para lograr la triangulación en los estudios, algunos investigadores recurren no sólo al uso de más de un instrumento si no a recoger información de más de un grupo de los sujetos de investigación. El

instrumento Likert puede ser validado cuantitativamente, cualitativamente o en forma combinado de las dos.

En los estudios revisados por los autores, se detectó que el proceso de construcción y validación de la escala Likert presenta algunas características constantes. Sin embargo, cada estudio puede implicar un diseño distinto de acuerdo a la información que se busque obtener y al diseño metodológico de la investigación. En consecuencia, estos estudios ilustran la diversidad de posibilidades en la construcción, validación e inclusión de sujetos se describen algunos de los estudios que fueron recuperados en la recolección de literatura sobre evaluación docente. No obstante, los estudios analizados también nos permiten observar que el uso de cuestionarios Likert para evaluar el desempeño docente debe cumplir con una construcción metodológica; esto con la finalidad de que los resultados obtenidos gocen de un rigor científico que facilite una documentación precisa, sistematizada y replicable de la realidad observada.

## SEMBLANZA DE LOS AUTORES

Profesores Investigadores de la División Académica de Educación y Artes de la Universidad Juárez Autónoma de Tabasco, México.

## REFERENCIAS:

- Aquino, S., Izquierdo, J., y Echalaz, B. (2013), Evaluación de la práctica educativa: una revisión de sus bases conceptuales. *Revista Actualidades Investigativas en Educación*, Vol. 13, No. 1. Costa Rica: Universidad de Costa Rica.
- Arias Galicia, F. (1984), El inventario de comportamientos docentes (ICD): Un instrumento para evaluar la calidad de la enseñanza. *Revista Perfiles Educativos*, Vol. 23 No. 4. México: CESU/UNAM
- Bell, T. (2005), Behaviors and attitudes of effective foreign language teachers: Results of a questionnaire study. *Foreign language annals*, Vol. 38, No. 2. E.U.: ACTFL
- Casero Martínez (2008), Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación educativa*, Vol. 26 No. 1. España: Asociación interuniversitaria de investigación en pedagogía.
- Bisquerra Alzina, R. (2004). *Metodología de la investigación Educativa*. Madrid: Editorial La Muralla. S.A.
- Carifio, J.; Rocco, P.(2008). *Resolving the 50-year debate around using and misusing Likert scales*. Blackwell Publishing Ltd.
- Cohen, L. y Manion, L. (1994). *Research Methods in Education*, London: Routledge.

- Elejabarrieta, F. e Iñiguez, L., (2008). Construcción de escalas de actitud tipo Thurstone y Likert. Revista electrónica La sociología en sus escenarios No. 17: Centro de Estudios de Opinión, Universidad de Antioquia,. Disponible en: <http://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/6820>.
- Field, A. (2000). "Discovering Statistics Using SPSS". London: SAGE Publication.
- García Garduño (2003). "Profesores Universitarios y su efectividad docente. Un estudio comparativo entre México y Estados Unidos". *Revista Perfiles Educativos*. Vol. 25 No. 100. México: CESU/UNAM.
- Guisande, Castor, et al., (2006). *Tratamiento de datos*. España: Ediciones Díaz Santos.
- Hammadou Sullivan (2004), "Identifying the best foreign language teachers: teacher standards and professional portfolios". *The Modern Language Journal*. Vol. 88, No. 3 (Autumn, 2004), pp. 390-402. E.U.: Blackwell Publishing
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio. P., (2007). *Metodología de la Investigación*. México, Mac Graw Hill.
- Johnson, S. (1994), A national assessment of secondary-school principals' perceptions of teaching- effectiveness criteria. *Communication Education*, Volumen 43, January, 194. E.U.: NCA.
- Kyriakides, L., Campbell, R. J. and Christofidou, E. (2002). Generating Criteria for Measuring Teacher Effectiveness Through a Self-Evaluation Approach: A Complementary Way of Measuring Teacher Effectiveness, School Effectiveness and School Improvement, 13: 3, 291-325. Inglaterra: Routledge
- Kyriakides, L., Demetriou, D. and Charalambous, C. (2006). "Generating criteria for evaluating teachers through teacher effectiveness research". *Educational Research*, 48: 1, 1-20. Inglaterra: Routledge
- Larsen-Hall, J. (2010). *A Guide to Doing Statistics*. New York: Routledge.
- López-Barajas y Ruiz Carrascosa (2005). "La evaluación de la Docencia Universitaria. Dimensiones y variables más relevantes". *Revista de Investigación educativa*. Vol 23 No. 1. España: Asociación interuniversitaria de investigación en pedagogía.
- Luna Serrano y Rueda Beltrán (2001). "Participación de académicos y estudiantes en la evaluación de la docencia". *Revista Perfiles educativos*. Vol 23 No. 93. México: CESU/UNAM.
- Mackey, A., y Gass, S. (2005). *Second Language Research. Methodology and Design*. Nueva Jersey: Lawrence Erlbaum Associates, Publishers. LEA.
- Méndez ,M., Peña, J., (2007). *Manual Práctico para el diseño de la Escala de Likert*. México UANL/Trillas
- Morales, P., (2006). *Medición de Actitudes en Psicología y Educación*. Tercera edición. España: Universidad Pontificia Comillas de Madrid.
- Muñoz, L. M., Ríos, M. P. y Abalde, E. (2002) Evaluación docente vs. Evaluación de la calidad. *Revista electrónica de Investigación y Evaluación Educativa*, 8 (2), 103-134. [http://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2\\_4.htm](http://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2_4.htm).
- Namakforoosh (2000). *Metodología de la investigación*. México: Limusa
- Patrick y Smart (1998). An Empirical Evaluation of Teacher Effectiveness: the emergence of three critical factors', *Assessment & Evaluation in Higher Education*, 23: 2, 165-178. Inglaterra: Routledge.
- Romano, J., Kromrey, J., Coraggio, J., y Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? Paper presented at the Annual Meeting of the Florida Association of Institutional Research, Cocoa Beach, Florida.
- Rueda, M. y Nieto, J. (comps.). *La evaluación de la docencia universitaria*. México, Facultad de Psicología-UNAM.
- Sierra Bravo, R. (2001). *Técnicas de Investigación Social*. España: Paraninfo.
- Spooren, Mortelmans y Denekens (2007), 'Student evaluation of teaching quality in higher education: development of an instrument based on 10 Likert-scales'. *Assessment & Evaluation in Higher Education*. Vol. 32:6, 667-679. Inglaterra: Routledge.

