

De “quantofrenias” y “artefactos” o la “doxa” en el reino de la “episteme”. Uso y abusos de las escalas de estimaciones sumadas en la evaluación del desempeño docente en contextos de educación superior universitaria

Héctor Mora Nawrath

Departamento de Antropología, Universidad Católica de Temuco
Chile
hectmora@uct.cl

Cita sugerida: Mora Nawrath, H. (2015). De “quantofrenias” y “artefactos” o la “doxa” en el reino de la “episteme”. Uso y abusos de las escalas de estimaciones sumadas en la evaluación del desempeño docente en contextos de educación superior universitaria. *Revista Latinoamericana de Metodología de las Ciencias Sociales*, 5 (1). Recuperado a partir de: <http://www.relmecs.fahce.unlp.edu.ar/article/view/relmecs05n01a03>

Resumen

Este trabajo presenta una reflexión metodológica relativa al uso de escalas de estimaciones sumadas o Likert en la evaluación del desempeño docente en el contexto universitario. Se presentan antecedentes en el marco de las prescripciones técnicas para este tipo de escalamientos, así como un conjunto de observaciones referidas a la pertinencia de su aplicación con fines evaluativos y a sus limitaciones en tanto herramienta para la generación de conocimiento. Se concluye que la escala de Likert puede ser utilizada en contexto evaluativos, atendiendo al conjunto de requerimientos ligados a su aplicación y tratamiento analítico-interpretativo, reconociendo los problemas insalvables que presenta, de manera de sopesar y relativizar la construcción del dato numeral. De este modo, se hace explícita la crítica al carácter “quantofrénico” y “artefactual” que acompaña su aplicación, y que contradictoriamente se inscribe en un discurso que sitúa la evaluación docente en el marco de políticas de calidad en educación superior.

Palabras clave: Quantofrenia; Artefactualidad; Escala de estimaciones sumadas; Medición de actitudes; Evaluación del desempeño docente.

“Quantophrenia” and “artefacts” or “doxa” in the kingdom of “episteme”. Use and abuse of summed estimation scales in teaching performance assessment in contexts of university higher education

Abstract

This work presents a methodological reflection on the use of summed estimation scales or Likert scales in teaching performance assessment in the university context. Information is presented in the framework of the technical prescriptions for scales of this kind, together with a series of observations referring to the relevance of applying them for assessment purposes and their limitations as tools for the generation of knowledge. It is concluded that the Likert scale may be used in the context of assessments, subject to all the requirements associated with its application and analytical-interpretative treatment and recognising the insuperable problems which it presents; this will allow the construction of the numerical datum to be weighed and relativised. This approach throws a critical light on the ‘quantophrenic’, ‘artefactual’ nature of its application, revealing the contradictions of its use in the context of a discourse which places teacher assessment in the framework of policies on quality in higher education.

Keywords: Quantophrenia; artefactuality; summed estimation scales; attitude measurement; teaching performance assessment.

En la rabiosa epidemia de quantofrenia, todo el mundo puede ser un “investigador” y un “indagador científico”, a causa de que todo el mundo puede coger unas cuartillas, llenarlas con toda clase de preguntas, enviar los cuestionarios a todos los sujetos posibles, recibir las respuestas, clasificarlas de este o aquel modo, someterlas a una máquina de tabular, colocar los resultados en varias tablas... y luego escribir un ensayo o un libro lleno de impresionantes adornos de tablas, fórmulas, índices y otras evidencias de una investigación objetiva, esmerada, precisa, cuantitativa (Sorokin, 1957: 258).

Introducción

Como cualquier instrumento, una escala que busca evaluar el desempeño docente¹ debe cumplir con requisitos de calidad y rigor metodológico conforme los marcos procedimentales propios de las técnicas de recolección y análisis que se ponen en escena.² No podría ser de otro modo, si a ello sumamos que dichos instrumentos se encuadran en la política en educación superior que cruza América Latina, formando parte de los mecanismos orientados a asegurar la calidad de las instituciones universitarias y de sus cuerpos académicos (Levy, 1994; Brunner, 1997; Schugurensky, 2000; Alcántara, 2000; Mollis, 2003; Tünnermann, 2003; Chiroleu y Iazzetta, 2005; Llomovatte, 2006). Para el caso de la evaluación del desempeño docente, el instrumento más popular, a pesar de las múltiples críticas y recomendaciones (Gasperoni y Marradi, 2002; Marradi, 2002, 2007), es el cuestionario estandarizado que recoge opiniones-actitudes de los estudiantes con base en escalamientos tipo Likert. Dicha evaluación se levanta sobre el modelo clásico basado en la opinión de los estudiantes (Elizalde y Reyes, 2008), formando parte de lo que Ricco (2001 en Fernández y Coppola, 2008) denomina evaluación tecnológica productivista.

Si la aplicación de estos instrumentos tiene una finalidad evaluativa, debemos advertir que el acto de evaluar no resulta de la pura aplicación instrumental y mecánica, y requiere reflexionar en torno al proceso de construcción del objeto (Bourdieu, Chamboredon y Passeron, 1985 [1973]). Si ello implica la medición de actitudes por medio de opiniones de estudiantes, no podemos desconocer operaciones de tipo epistemológica, teórico-metodológica y técnica que deben ser precisadas y consideradas tanto en la elaboración, como en la aplicación e interpretación de la información.³ Estas debiesen responder a interrogantes tales como: ¿qué se busca analizar o evaluar?, ¿qué claves interpretativas están contenidas en la teoría desde la cual se operativizan los enunciados?, ¿se está midiendo lo que se dice o lo que se quiere medir?, ¿es la técnica de recolección adecuada?, ¿qué límites procedimentales e interpretativos imponen los objetos y las técnicas de análisis?, ¿cómo debe ser interpretada la información?, entre otras⁴.

El problema de fondo se relaciona con la construcción de “artefactos”, cuestión que se manifiesta en el uso meramente procedimental y pragmático de métodos y técnicas, algo no poco común en investigación social cuantitativa y cualitativa. Ello reviste gravedad, si

consideramos que en base a “artefactos” se caracterizan y clasifican poblaciones y se toman decisiones políticas, y que para el caso de las encuestas de opinión del desempeño docente, piezas de la política universitaria, repercuten directamente sobre el estamento académico-docente al estar incluidas en los sistemas de calificación y categorización académica⁵ que se orientan a demostrar-exigir calidad y productividad.

El objetivo de este artículo es discutir la actitud “quantofrénica” (Sorokin, 1957) y “artefactual” (Bourdieu, Chamboredon y Passeron, 1985 [1973]; Bourdieu, 2008 [1984]; Merlliè, 1993 [1989])⁶, que se expresa como retórica cientificista (Kreimer, 2005), y que busca investir de legitimidad a un discurso en tanto opera con números, gráficas y procedimientos estadísticos que fundamentan una representación o una política constitutiva del sistema educativo en general.

Si el discurso tiene como eje la calidad, esta debe exigirse a todo nivel. Por ello, se busca contribuir al debate proponiendo ciertos aspectos que deben orientar la reflexión metodológica sobre los procedimientos de recolección y análisis asociados a la medición de actitudes, y que recupera requisitos en la aplicación y/o interpretación de los resultados propios de la técnica -ligado a la medición de atributos sociales, la interpretación estadística y la aplicación de encuestas-, y aquellas críticas que han surgido a partir de la propuesta original levantada por Rensis Likert.

Esta reflexión se nutre del conocimiento adquirido en diversas experiencias de evaluación en la aplicación y análisis de cuestionarios, y en particular, de la aplicación de cuestionarios para la evaluación docente en la institución de la cual forma parte quien presenta este trabajo. La reflexión que aquí se presenta se ha centrado en la aplicación de encuestas, los cuestionarios con escalamientos tipo Likert, y los análisis estadísticos en el marco de las discusiones sobre medición de atributos sociales.

1. ¿Evaluación del desempeño docente o posicionamiento del sujeto-estudiante?

Para dar cuenta del desempeño docente, pueden ser utilizados varios procedimientos técnicos de recolección y análisis de información. Uno de los más frecuentes es la encuesta on-line a través del cuestionario auto-administrado, siendo en algunos casos denominada “Encuesta de Opinión del Desempeño Docente” (en adelante EODD). En tal caso, el cuestionario resulta ser el dispositivo que media la comunicación y se compone de una escala o constructo dispuesto como batería de enunciados que operacionalizan un “objeto” de opinión o actitud, contemplando distintas dimensiones según la perspectiva teórica orientadora y el objetivo de la escala propuesta. Dichos enunciados o ítems se presentan al

encuestado de modo tal que manifieste su posicionamiento en función de opciones de opinión que se definen en una gradiente bi-polar que corresponde al escalamiento por estimaciones sumadas (summated ratings) o tipo Likert,⁷ esto, más allá que se utilice como expresión de opinión un simbolismo textual -desde muy de acuerdo a muy en desacuerdo- o numeral -por ejemplo, entre 1 a 7 (Likert, 1978 [1932]; Padua y Ahman 1979; Prat y Doval 2003)-.

Estas escalas buscan medir o determinar una actitud, es decir, una disposición evaluativa de carácter interno que un sujeto expresa frente a un fenómeno, objeto específico, persona o idea. Al tener un referente cognitivo y reflejar una predisposición, la actitud puede ser entendida como expresión exteriorizada del sujeto, la que se manifiesta a través de su posicionamiento en relación a un conjunto de enunciados que representan su opinión, y que deben poseer cierto grado de consistencia (Thurstone, 1978; Summer, 1978; Cortada de Kohan, 2004).⁸ Dicha actitud puede ser inferida a partir de un conjunto de posicionamientos que expresan intencionalidad, y que dan cuenta de “la suma total de inclinaciones y sentimientos, prejuicios o distorsiones, nociones preconcebidas, ideas, temores, amenazas y convicciones de un individuo acerca de cualquier asunto específico” (Thurstone, 1978:,158).

Su carácter evaluativo implica una carga afectiva y motivacional que se pone en juego - implica dirección e intensidad-, y que opera desde categorías que actúan como referentes comparativos acerca de un dominio o aspecto de la realidad, siendo resultado del conocimiento o la interacción social, lo que resulta en asignaciones de sentido positivo o negativo (Sherif y Sherif, 1978). En esta dirección, la actitud connota juicios, valoraciones y predisposiciones a la acción que van más allá de la mera descripción del objeto, implicando respuestas de aceptación o rechazo a través de modos de conducta característicos, consistentes y selectivos expuestos en un periodo de tiempo determinado (Sherif y Sherif, 1978; Osgood, Suci y Tannenbaum, 1978).

El tipo de escalamiento mencionado indica que la propiedad a medir -actitud- es el resultado de la suma de las opciones de opinión dispuestas en la escala como si fuese un continuo respecto del cual el sujeto se sitúa, exigiendo propiedades de amplitud y consistencia en las respuestas actitudinales. Los sujetos se encontrarán ordenados en función del “acuerdo” o “desacuerdo” respecto de cada una de las proposiciones, siempre que estén relacionadas con la actitud que pretenden medir. Es una escala centrada en el sujeto, debido a que supone que la variación en las respuestas se debe a las diferencias que existen entre estos (Likert, 1978 [1932]).

De lo anterior se puede extraer un primer cuestionamiento a su aplicación en el marco de la evaluación docente: la insistencia en definir este tipo de procedimientos como evaluaciones de su desempeño. Si bien la selección de una alternativa de respuesta frente a un enunciado puede ser tomada como opinión, lo que estos tipos de escalamientos están orientados a determinar es la evaluación o actitud del sujeto conforme sus marcos interpretativos. Por ello no es vano insistir en que la propuesta de Likert buscaba escalar sujetos a partir de reacciones “afectivas” en función de los enunciados propuestos, no siendo su finalidad evaluar objetos.

Por ejemplo, si la escala contempla 19 enunciados y 5 opciones de respuesta, el menor puntaje posible sería 19 y el mayor 95; esto quiere decir que las puntuaciones teóricas se distribuyen entre estos límites, y que cualquier puntuación concreta tomará un valor en dicho rango.⁹ Ello puede llevar erróneamente a pensar que la puntuación -de entre 19 y 95 o entre 1 y 7- resulta equivalente a la calificación o evaluación que se otorga al docente. Sin embargo, dicha puntuación no es más que una representación de la posición del sujeto-estudiante en la escala y no del docente. Advertir que, conforme a los objetivos analíticos que se persiguen, existen una serie de técnicas alternativas orientadas al escalamiento de objetos -escala de Thurstone- y sujetos-objetos, estas últimas, denominadas escalas de Guttman y Coombs (Morales et. al., 2003; Prat y Doval, 2003; Cortada de Kohan, 2004).¹⁰

2. Supuestos de las escalas de estimaciones sumadas o de Likert

Como la mayoría de los procedimientos técnico-metodológicos, la escala de Likert presenta una serie de supuestos y requerimientos básicos que deben ser considerados, entre otros, por sus efectos en términos interpretativos: a) las actitudes de los sujetos pueden inferirse a través de preguntas o enunciados que dan cuenta de opiniones que expresan pensamientos, sentimientos, creencias y conductas probables; b) las opiniones tienen el mismo significado para todos los sujetos; c) los sujetos indican su acuerdo o desacuerdo con las opciones de opinión, las que pueden ser codificadas con número en función de la gradiente definida, de manera de posibilitar el escalamiento de los sujetos; d) todos los ítems dan cuenta del mismo rasgo con la misma intensidad, y a través de la corrección de la orientación del enunciado, todas las respuestas están en la misma dirección; e) la escala exige que se demuestre unidimensionalidad y monotonicidad, es decir, que mide el mismo rasgo o concepto, y por tanto, la posibilidad de aceptar o rechazar un ítem debe ser monótonica (Likert, 1978 [1932], Morales et. al., 2003; Cortada de Kohan, 2004).

Conforme a lo anterior, se recomiendan una serie de consideraciones para la construcción adecuada de escalas de actitud, y en particular para la redacción de enunciados (Padua y Ahman, 1994; Morales et. al., 2003; Prat y Doval, 2003), los que se describe a continuación:

Procedimientos para la construcción de escalas

1. Delimitación de los objetivos de la escala o de la actitud a medir.
2. Selección teórica de los ítems.
3. Elaboración y redacción de los ítems.
4. Determinar el modo y número de respuestas.
5. Preparar la clave de corrección.
6. Preparar preguntas adicionales orientadas al estudio de validez.
7. Obtención de datos de una muestra.
8. Análisis de los ítems (discriminación).
9. Comprobar la fiabilidad.
10. Configuración del instrumento de medida.
11. Comprobar validez y otros análisis.

Procedimientos para la redacción de enunciados

1. Todas las afirmaciones deben guardar alguna relación con la actitud en estudio.
2. Las opiniones deben reflejar la actitud presente del sujeto y no sus actitudes pasadas. Los enunciados deben aparecer expresados en tiempo presente.
3. Las afirmaciones con doble significado son ambiguas. Cada afirmación debe expresar una sola idea. Cada proposición debe tener una sola interpretación posible.
4. Evitar enunciados que solo son aplicables a grupos restringidos de personas.
5. Las afirmaciones no deben tener conceptos relacionados de forma confusa. Utilizar un lenguaje claro, simple y directo.
6. Evitar el vocabulario de jergas, excepto cuando sirvan a un objetivo particular.
7. Evitar frases interpretables como hechos. Cada proposición debe ser opinable y debatible.
8. Evitar afirmaciones con las cuales la mayoría, o casi nadie estaría de acuerdo.
9. Las frases deben ser cortas; no más de 20 palabras.
10. Evitar palabras que implican universalidad (todos, siempre, ninguno, nunca, etc.).
11. Utilizar con cuidado los adverbios. (simplemente, habitualmente, etc.).
12. Las frases deben ser sencillas, sin cláusulas dependientes.
13. Evitar el empleo de formas negativas complejas (frases con doble negación).

14. Cada afirmación debe permitir respuestas en toda la gama de intensidad de la actitud medida.
15. Combine los ítems formulados de manera positiva o negativa en una proporción aproximada de 50% y 50%.

Muchas unidades que se encargan de procesar información en el marco de la evaluación docente no consideran estos aspectos, lo cual se agudiza con la automatización informática que acompaña a estos procesos. El objetivo, al parecer, es la producción de números, promedios y desviaciones para cumplir con la aplicación de estos instrumentos en el marco de las acreditaciones institucionales, olvidando el esfuerzo analítico e interpretativo que se debe desplegar.

Con el objetivo de evaluar la redacción de 19 enunciados para una escala que busca medir el “Desempeño Docente”, se detallan los principales problemas con base en las 15 consideraciones antes descritas.

Tabla 1: Evaluación de enunciados

Enunciado	Comentarios
Cuando procede, muestra la relación del curso con el futuro profesional	El problema se desprende de la pertinencia del enunciado en función de ciertas materias y avances del estudiante. ¿Cómo se sitúa frente al enunciando un estudiante de un curso de matemáticas en una carrera de kinesiología, donde además el docente es matemático y no kinesiólogo? Por otro lado, el condicional “Cuando procede” introduce cierta ambigüedad. ¿Un estudiante de una materia en el primer año, puede discernir con claridad lo que solicita el contenido del enunciado?
Muestra interés por el desarrollo integral de sus estudiantes.	Hay ambigüedad, lo que no permite situar correctamente el objeto de la actitud; ¿cómo interpreta un estudiante la frase “desarrollo integral”? Se deja al arbitrio evocar la noción de integralidad, la que puede estar compuesta por todos los atributos que subjetivamente le asigne. No todos pueden entender de igual forma “desarrollo integral”.
Las evaluaciones las realiza con pautas claras, objetivas y conocidas.	Se debió privilegiar un enunciado simple y directo, ya que todos los adjetivos pueden ser resumidos en: pautas claras y conocidas.
Nos entregó y explicó el programa del curso o guía de aprendizaje.	En este enunciado también se observa ambigüedad. Habría que distinguir entre entregar -que es proporcionar al estudiante el programa- y explicar -que es detallar en qué consiste la asignatura en términos de contenidos y evaluaciones-, lo que implicaría dos acciones diferentes. Por otro lado, respecto de la primera afirmación, esta no permite graduar la gama en la intensidad de las opiniones, puesto que solo son posibles dos acciones: entrega o no entrega. La respuesta al enunciado, más que un aporte, genera una distorsión en la escala; en este sentido podemos comparar tres puntuaciones en el enunciado para su aplicación 2012 y 2013 en tres evaluaciones diferentes de un mismo docente: 7,0 sin desviación (lo que demuestra que todos los respondientes puntuaron 7); 5,9 con desviación 1,67; 5,58 con desviación 1,53.

Es puntual en el inicio y término de las clases.	Este tipo de enunciado exige un esfuerzo en el sujeto por determinar en un lapsus considerable de tiempo aquellas veces en que se comienza y finaliza en el tiempo establecido (considerar que algunas cátedras implican hasta 32 sesiones en un semestre académico). Puede ocurrir que de las 32 sesiones, el docente hubiere llegado impuntual a las tres últimas, y que en base a ello el estudiante manifieste su opinión.
Desarrolla todos los contenidos del programa o guía de aprendizaje a lo largo del semestre.	El supuesto en el que se basa este enunciado es muy discutible, ya que presupone la obligación de cumplir con el programa, y no en la administración de su contenido dependiendo del resultado de aprendizaje y avances en una cátedra. No se trata de pasar contenidos, sino de estimular el aprendizaje. En esta misma dirección, un programa debería ser concebido como una guía o ruta de referencia, ya que la concreción depende de las características de los estudiantes y la normalidad de un año académico.
Nos estimula en el uso de la bibliografía del programa del curso o guía de aprendizaje.	Este enunciado puede ser suprimido o reorientado, en tanto lo que supone debiera ser parte de una práctica interiorizada por el estudiante. A ello se suma que el manejo de la bibliografía en una asignatura se examina en las evaluaciones que el docente establece.

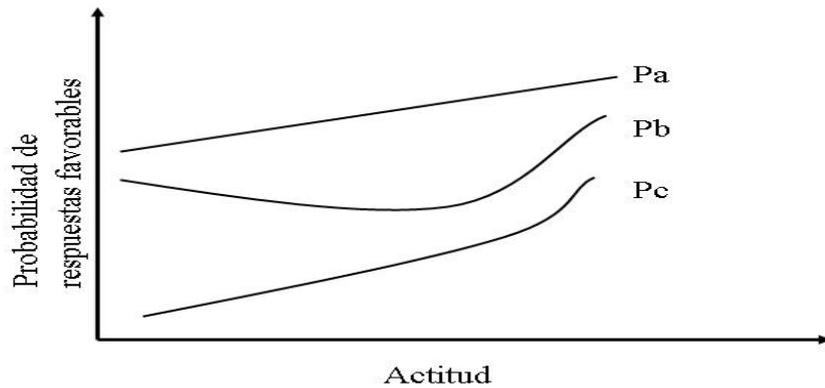
El ejemplo anterior permite evidenciar que más de un tercio (36%; 7/19) de los enunciados presenta algún tipo de problema en su formulación. A ello se suma el no considerar el requisito de combinar enunciados con orientación positiva y negativa; esta recomendación no es menor, ya que su incorporación fue realizada para evitar que el sujeto mecanice el proceso de respuestas (llevarlo a reflexionar sobre el enunciado) una vez advertida la orientación de los enunciados. Ello da cabida a las denominadas respuestas set, fenómeno frecuente en este tipo de técnicas, lo que será tratado más adelante.

Por último, es conveniente hacer mención al supuesto de monotonía implicado en este tipo de escalas, dada su relevancia en la evaluación de la calidad e interpretación de la escala. Una función es monótona -entre dos conjuntos-, si conserva el orden dado a través de funciones monótonas crecientes y decrecientes (o simplemente crecientes y decrecientes), o de funciones que conservan e invierten el orden ($f: p \rightarrow q$). En el marco de escalas de actitud, la monotonía implica que la probabilidad de que un sujeto valore positiva o negativamente un enunciado es monótona respecto de la posición que ocupa en el continuo de la escala, es decir, que crece o decrece en función de lo favorable o desfavorable de la actitud (Elejabarrieta e Iñiguez, 1984; Ghiglione y Matalon, 1989; Cortada de Kohan, 2004).

Por ejemplo, si suponemos que la actitud existe, y por tanto los sujetos se ordenan según sean más o menos favorables al “objeto”, cada vez que el ítem se encuentre bien relacionado con la actitud global, la probabilidad de que un sujeto proporcione la respuesta favorable a cada pregunta o ítem particular será más alta. Si observamos la figura 1, el eje Y representa la probabilidad de respuestas favorables y el eje X la actitud; la curva que

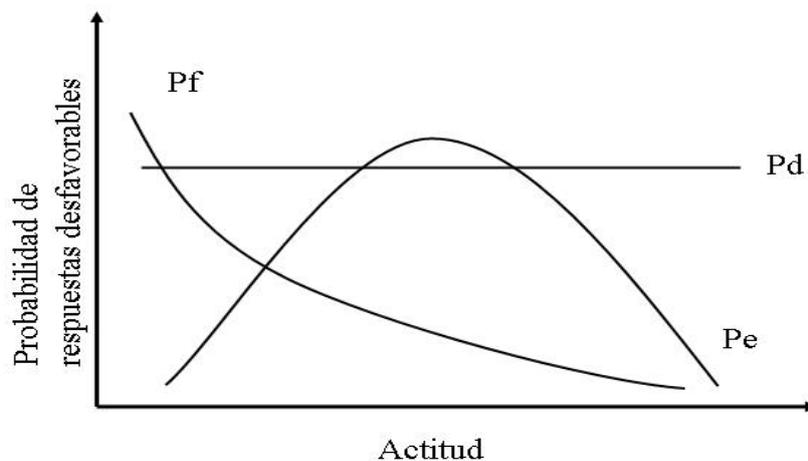
corresponde a la pregunta (Pa) grafica la probabilidad de respuesta favorable a una pregunta moderada o neutra, respuesta que incluso los sujetos poco favorables pueden proporcionar. La pregunta (Pb) es intermedia, en tanto que (Pc) es apropiada, pues solo provoca una respuesta favorable de parte de los sujetos que poseen una actitud global muy favorable.

Figura 1



Fuente: Elaboración propia

Figura 2



Fuente: Elaboración propia

Por otro lado, la figura 2 expresa otras posibilidades o distribuciones conforme la determinación de actitudes desfavorables; algunas de ellas debieran ser rechazadas ya que implica que los enunciados no permiten discriminar entre actitudes positivas o negativas, en tanto otras deberían ser ajustadas. Si observamos la distribución (Pd), ella muestra una recta que grafica una proporción constante de respuestas favorables y desfavorables, lo que no permite determinar la actitud. (Pf) representa un error de codificación, si consideramos

que la orientación del conjunto de enunciados es positiva, pudiendo corregirse a partir de la inversión de los valores. (Pe) es la respuesta que obtendría si se manifestara una posición promedio, lo que implicaría la imposibilidad de precisar la intensidad de las valoraciones en los sujetos; esta tendencia deja en evidencia problemas de discriminación -o de acomodación en la opción intermedia-, y puede obedecer a un error en la formulación del enunciado. Este tipo de respuesta no podría ser sumada al total de la escala, y es muy posible que las pruebas de correlación ítem-test, y las de fiabilidad detecten lo anterior.

Por ejemplo, si pensamos en la actitud que presentan un conjunto de estudiantes que han asistido a la misma clase y “evalúan el desempeño” de un docente en particular, distribuciones como aquellas que expresa la figura 2 u otras distribuciones -como del tipo campana de gauss invertida- no permitirían pensar en la validez y fiabilidad de la escala, lo que implicaría desechar el instrumento.¹¹

Respecto a lo descrito, podríamos preguntar ¿Cómo interpretan lo anterior las instancias institucionales orientadas a la evaluación? ¿Es realmente posible decir algo sobre el objeto?, ¿se cumple con los requisitos de validez y fiabilidad?, ¿se examina el comportamiento de los enunciados y de las escalas? Esto nos hace pensar profundamente en la legitimidad de las políticas de evaluación cuando se presentan estas disyuntivas; una aproximación mecánica a la aplicación y análisis de cuestionarios -producir número- hace imposible advertir ciertos problemas en el comportamiento de la distribución, por lo cual no se puede garantizar el cumplimiento de los supuestos asociados a la interpretación de la escala y a la formulación adecuada de los enunciados.

3. El cuestionario como dispositivo semántico y los marcos de la experiencia.

Algunos problemas de la escala de Likert

Al aplicar un cuestionario, no podemos olvidar que el conjunto de enunciados contenidos incluyen una clave interpretativa que se desprende de una teoría. Por ello, la redacción y selección de enunciados involucra siempre un esfuerzo importante de operacionalización teórica en lo cual la comprobación de validez y fiabilidad es necesaria (Lazarsfeld, 1973 [1965]).¹² Para el caso de la evaluación docente, el constructo debería operativizar una concepción del desempeño docente conforme a una teoría de la educación o del área pedagógica, esto, en el marco del modelo educativo que cada institución pone en marcha. Además, se torna fundamental considerar en dicha elaboración los marcos de la experiencia del encuestado, en tanto las respuestas resultan de la interpretación del sujeto respecto de lo que cada enunciado o pregunta contiene. Ello implica una cuestión de carácter gnoseológica ligada a la comprensión e interpretación del instrumento en tanto dispositivo

de comunicación que moviliza o pone en juego el conocimiento tácito del encuestado (Cicourel, 1968).

Considerando lo anterior, no es aconsejable dar por sentada la ininteligibilidad de un cuestionario, es decir, que un sujeto sitúa el contenido en su experiencia e interpreta lo allí dispuesto sin mediar tensiones, sean éstas semánticas o derivadas de las tipificaciones que utiliza como referentes evaluativos ligados, en este caso, al desempeño docente.¹³ En tal sentido, el cuestionario debe ser considerado en su doble dimensión: como instrumento de recolección de información y como dispositivo semántico. Relevar lo anterior, como señala Navarro (1995), nos lleva a situar la encuesta social por cuestionario en una perspectiva post-positivista; la puesta en escena de un texto que plasma por medio del lenguaje una determinada interpretación elaborada y propuesta por un sujeto epistémico “experto” que es interpretada por un sujeto cognoscente. Lo anterior implica que pueden tener espacio diversas interpretaciones, e incluso, dar lugar a tensiones que se producen por la disputa de sentido no solo en función del nivel explícito representado por preguntas o enunciados -¿por qué se incluyen ciertas preguntas?, ¿por qué se pregunta sobre un tópico específico?- , sino también por aquello que no está pero que debió estar presente -lo que debió haberse preguntado y no se preguntó- o lo que se busca pesquisar -qué se busca medir o qué se debe responder-. En esta dirección, hay que consignar que hay un “texto” que fluye en paralelo a la aplicación de un cuestionario, y que toma forma a través de los comentarios que emite el encuestado.

Refiriéndonos al cuestionario analizado, frente a enunciados como “Facilita nuestro aprendizaje con la utilización de diferentes estrategias de enseñanza”, “cuando procede, muestra la relación del curso con el futuro profesional”, “es claro en la exposición de los contenidos tratados en clase”, se esconden dilemas semánticos ligados a la situación social del respondente y su marco de experiencia. ¿Qué es una estrategia de enseñanza y cuáles considero, desde mi subjetividad, como óptimas o deseables?, ¿cómo imagina lo concerniente al futuro profesional un estudiante que se encuentra en sus primeros años de formación?, etc.

Las situaciones descritas permiten ser pesquiasadas por medio de procedimientos técnicos que pueden situarse entre aquellos que ayudan a definir la validez y confiabilidad. La tensión estructura-semántica se puede explorar a través del análisis sociosemántico (Navarro, 1995) en la fase de puesta a prueba del instrumento o, del mismo modo, se puede incluir como parte del registro que puede llevar el encuestador en la situación de encuesta.

Asociado al problema de interpretación de los enunciados, y en particular para las técnicas que se basan en comentarios de frases con pleno sentido, Marradi (2007) advierte el fenómeno de la reacción parcial al enunciado o reacción al objeto. Comparando comentarios y reacciones a los enunciados propuestos en una escala, se ha determinado que entre un 5% y 10% de ellas se efectuaban en función de un término y no del enunciado completo y, en específico, respecto del objeto designado por ese término. En ello juega la carga emotiva o emocional que evoca un término específico.

De la subjetividad y cualidad interpretativa-reflexiva del encuestado se desprende otro problema, las “respuestas set” (Marradi, 2007). Por lo general, los enunciados se presentan en baterías de preguntas, lo que hace posible que el encuestado aprenda el esquema de organización del instrumento y responda a cada enunciado como si fuese parte de una única afirmación o dimensión, sin realizar una lectura reflexiva uno a uno.¹⁴ Este es un problema que Likert también detectó, para lo cual propuso invertir el sentido de las valoraciones de los enunciados, combinando orientaciones negativas y positivas. Al respecto, autores como Cook y Sellitz (1978) proponen: a) introducir enunciados o reactivos que no correspondan al objeto actitudinal; b) dar garantías de anonimato; c) hacer hincapié en las contestaciones honestas; d) incorporar reactivos desfavorables para evitar disposiciones a contestar uniformemente en función de respuestas de orientación favorables; e) incorporar reactivos de control.

Lo descrito con anterioridad puede ser rotulado como posibles fuentes de distorsión, las que no necesariamente son advertidas por el analista, más aún, si opera por medio de la automatización en la aplicación y procesamiento de la información, por ejemplo, mediante cuestionarios on-line. Sin embargo, si bien son posibles de determinar o controlar, su ocurrencia es ineludible al ser efectos asociados a la técnica de cuestionario, en tanto tienen lugar debido a que: a) cada sujeto puede concebir de manera distinta -de quien propone la escala- la propiedad en la cual debe valorar su estado; b) cada sujeto puede equivocarse de buena fe en su posicionamiento; c) cada sujeto puede manejar conscientemente la imagen que proyecta a través del instrumento, en los marcos de la deseabilidad social o de la búsqueda de revancha; d) un sujeto puede errar en el entendimiento del mecanismo de funcionamiento de la escala y por tanto confundir lo que cree que debe responder (Marradi, 2007).

Sumado a lo anterior, se pueden consignar otro tipo de problemas; no pocos estudiantes han manifestado responder cuestionarios de manera aleatoria. Aun cuando ello es fácil de detectar -por ejemplo con un análisis de distribución, correlación entre los enunciados o de validez del constructo-, la automatización del proceso no lo advierte. En estimaciones

sumadas ello no se hace explícito si solo se atiende a la puntuación índice -que es la suma de todas las opiniones-, y no se explora el comportamiento de cada una de las puntuaciones en la distribución. Lo anterior se relaciona con el grado de compromiso con el objeto de evaluación y con el cuestionario, y se puede atribuir a: a) el alto número de frases que componen la batería a la cual se somete al encuestado -además, con base en una estructura monótona-; b) los acervos de conocimiento y las esferas de interés de los encuestados que implican la baja motivación por lo que se consulta; c) no tener opiniones o conocimientos sobre los temas que se presentan (Marradi, 2007).

Para el caso tratado, se han podido pesquisar expresiones del tipo: “No sabemos para qué contestamos si nunca pasa nada con la evaluación”, “no me interesa”, “es aburrido y me quita tiempo”, “preferiría poder expresarme y dar mi opinión”. La manifestación más clara de lo anterior es la baja receptividad para responder los cuestionarios, que se suma al problema que presentan encuestas cuyo cuestionario se aplica sin presencia directa del encuestador -como las telefónicas, postales u on-line-. En muchos casos resulta común encontrar que los porcentajes de respuesta a los cuestionarios no alcanzan al 50% del total de estudiantes, siendo posible incluso encontrar tasas de no respuesta que se aproximan al 80%. Aun cuando no existe consenso relativo respecto a cuál es la tasa de no respuesta admisible -ello depende por lo general del tipo de investigación-, la robustez de los datos y las posibilidades de realizar cualquier estimación, disminuyendo los sesgos, es a todas luces impensable para tasas de no respuesta en torno al 50%, aun más, para poblaciones pequeñas o inferiores a 50 casos. Con fines probabilísticos, la tasa de no respuesta no debería superar al 12% en estudios de mercado y un 11% en investigaciones sociales (Moser en Vivanco, 2005), en tanto su cálculo -tasa de no respuesta- puede resultar fundamental para evaluar el tamaño de la muestra en este tipo de diseños.

Para alcanzar tasas aceptables de respuestas, algunas instituciones han instalado la política de condicionar la inscripción de asignaturas, una medida impositiva que aumenta la probabilidad de respuestas aleatorias. Sin embargo, se hace necesario articular preguntas mucho más profundas: ¿Es posible pensar una evaluación con tasas de respuesta tan bajas?, ¿es significativo pensar en una tasa de respuesta de 50%?, ¿es la obligatoriedad de la respuesta una medida prudente?, ¿se ha reflexionado o indagado en las dimensiones sociológicas que están detrás del problema de la no respuesta?

4. El problema de la medición de atributos sociales y la interpretación del resultado en la escala de Likert

El concepto de medición ha experimentado ciertas adecuaciones que se han derivado de su aplicación en el campo de las Ciencias Sociales.¹⁵ Su acepción clásica, de orientación psicofísica, se desprende de la propuesta de Campbell,¹⁶ quien la define como la relación isomórfica que se establece entre las propiedades de un objeto y su representación a través de cantidades o magnitudes (Torgenson, 1971 [1958]; Muñiz, 1998).¹⁷ Stanley Stevens (1946) discute la concepción naturalista de Campbell -buscar números que representen cantidad o magnitud presente en un objeto- para sostener que lo fundamental en este campo radica en la unicidad de la representación, es decir, determinar una estructura que defina la invariabilidad de la escala. Medir consistiría, según Stevens, en asignar números a objetos según ciertas reglas, donde los números asignados no dan cuenta necesariamente de cantidades, sino de relaciones de representación en función de distintas propiedades de las variables o conceptos (nominal, ordinal, de intervalos y razón). De este modo, se amplía el concepto clásico de medición, en tanto la asignación de números debe conducirse a través de reglas (diferenciación escalar o nivel de medición), lo que hace posible medir variables que no poseen relación directa con cantidades o magnitudes como es el caso de niveles nominales y ordinales.

Si bien la propuesta de Stevens (1946) se puede interpretar como un avance en lo que se refiere a la teoría de la medición, y sirvió de base a la teoría representacional (Jáñez, 1989; Asún, 2006),¹⁸ no ha estado exenta de críticas. Con énfasis distintos, tanto Campbell como Steven¹⁹ tratan las relaciones entre sistemas empíricos y formales como axiomáticas (Muñiz, 1998); isomorfismo y axiomas matemáticos deben estar presentes para llevar a cabo la medición, tornándose en una perspectiva idealista. Por otro lado Marradi (2002, 2007), en concordancia con la perspectiva de Torgenson (1971 [1958]), señala que el término medición se debe reservar a objetos que exhiben propiedades métricas (de intervalo y de razón en la clasificación de Steven), siendo coherente para dar cuenta de objetos con propiedades no métricas, operaciones de conteo y clasificación. Lo que plantea Marradi (2002, 2007) ofrece restricciones para el tipo de operaciones que se pueden realizar al analizar una variable no numérica, pero resulta interesante de considerar conforme al tratamiento habitual para los datos que se producen por aplicación de escalas de estimaciones sumadas o de Likert. Ello además resulta un buen punto para reflexionar sobre el efecto métrico que produce la estimación sumada, lo que conduce a lo que Barriga (1998) ha denominado la “fetichización” del número o quantofrenia (Sorokin, 1957), y que lleva a que el investigador, ilusoriamente, interprete las puntuaciones como si fuesen producto de estimaciones sobre variables numéricas.²⁰

Desde cualquier punto de vista, la escala de Likert resulta de la suma de una serie de números naturales que se asignan como referentes de opinión valorativa conforme un conjunto de enunciados. Podríamos agregar que en estricto rigor, la variable es cualitativa y ordinal, donde el número -o más bien el numeral- resulta de una asignación arbitraria -podría ser cualquier número- que solo busca cumplir con el principio de nominación-exclusividad en base a la diferenciación jerárquica. La escala se construye conforme a un baremo que resulta de la suma de puntuaciones posibles de acuerdo a la cantidad de preguntas y opciones-opiniones que se presentan al sujeto.²¹ Para Marradi (2002, 2007), el problema surge de la utilización de técnicas e interpretaciones cardinales para dar cuenta de puntuaciones ordinales, debido a que carece de sentido atribuir características métricas a atributos que solo indican posición y no magnitud. En el caso de escalas de Likert, el sujeto es ubicado en un baremo según la puntuación obtenida, debiendo interpretar dicha puntuación como una referencia-localización y no como una magnitud, dado que las distancias entre cada rótulo no pueden ser consideradas iguales o equidistantes (Marradi y Macri, 2012). Por otro lado, los sujetos pueden percibir nítidamente una distancia discriminante entre los extremos o polos evaluativos pero no del todo entre las gradientes intermedias del continuo (Marradi, 2002, 2007).

Por ello, y para efectos evaluativos, la posición del sujeto es referencial respecto a la actitud de otros sujetos en la escala. Carecería totalmente de sentido basar la interpretación en números enteros o decimales, haciendo uso de sus supuestas propiedades numéricas. Que un estudiante demuestre una actitud de 6 en una escala, no tiene que ver con que el docente obtenga esa puntuación, y por otro lado, que 6 es una evaluación el doble de superior que una representada por un 3. En este contexto, hablar siquiera de asignación decimal (6,2) resulta poco convincente, y no representa más que la fetichización mediante la cual se le otorga al número una función que no puede cumplir.

En estricto rigor, las posibilidades analíticas para variables ordinales se restringen a la moda, la mediana, porcentajes y relaciones a través de procedimientos que trabajen con coeficientes ordinales (Marradi, 2002, 2007). Si bien es común la utilización de estadísticas multivariadas (componentes principales) para analizar validez del constructo y otro tipo de test o pruebas para variables métricas, se debe ser consciente de que el número es solo referencial y apoya la interpretación.²²

Otro problema que se relaciona con utilizar el escalamiento tipo Likert con fines evaluativos, tiene relación con el efecto curvilíneo o acampanado que exhibe la distribución de los datos cada vez que se adicionan todas las propiedades en todos los sujetos. Esto se conoce como efecto del tipo doblar el continuum, que transforma la distribución en una herradura (Marradi,

2002, 2007). Ello se produce debido a que: a) no todos los sujetos reaccionan del mismo modo a todos los enunciados que se les presentan, b) un enunciado puede ser desaprobado por motivos opuestos. Es más, se ha demostrado que en una distribución existe la tendencia, por parte de los sujetos, a ubicarse en posiciones medias y no extremas de una gradiente, lo cual también repercute en el efecto de curvilinealidad. En base a ello, ¿es realmente posible evaluar el desempeño si por definición, y por efectos de la técnica, es poco probable que podamos encontrar posiciones extremas?, y por otro lado, ¿se vulnera el principio de monotonía de la escala de actitud?

Retomando la discusión sobre la medición -o análisis de datos numerales para dar cuenta de fenómenos sociales-, se hace necesario plantear otros problemas derivados de las aproximaciones evaluativas a través de escalas tipo Likert. En la actualidad es ampliamente aceptada la idea de que la gran mayoría de los fenómenos sociales no son susceptibles de observación directa, debido a su alto grado de abstracción. De allí que Carmines y Zeller (1979) propongan concebir la medición como “el proceso de vincular conceptos abstractos con indicadores empíricos”, lo cual se realiza mediante un plan explícito y organizado para clasificar (y frecuentemente cuantificar) los datos disponibles en indicadores que se derivan del concepto o fenómeno que se busca referir. En este proceso, el instrumento de medición o de recolección de datos juega un papel central pues sin él no hay observaciones clasificadas. La definición sugerida incluye dos consideraciones: a) nos remite a la dimensión empírica, estableciendo que el centro de atención es la respuesta observable, sea esta una alternativa de respuesta en un cuestionario, una conducta grabada vía observación o una respuesta dada a un entrevistador; b) el interés se sitúa en el concepto subyacente, que es representado por la respuesta a nivel empírico.

Según lo anterior, un instrumento de medición adecuado es aquel que registra datos observados que representan “verdaderamente” los conceptos o las variables que el investigador tiene en mente. Por ello, resulta clave para la medición tanto la operacionalización conceptual -para llegar a definir los indicadores empíricos- como la aplicación del instrumento en la población, ello, de modo de establecer si la referencia del indicador condice con la establecida a través del concepto. En definitiva, es importante la relación entre indicador y el concepto fundamentado desde la teoría; una medición válida y fiable alcanza la correspondencia entre estas dos dimensiones. Esto nos lleva a concebir la medición como una acción selectiva que involucra distinguir teóricamente los rasgos característicos que pueden describir un concepto, y a pensar cómo desarrollar un proceso en pro de identificar la forma de representación numérica de éste. Lo anterior sugiere que la llamada operacionalización es clave a la hora de medir, es decir, al establecer la conexión

entre conceptos abstractos y sus propiedades a través de indicadores empíricos de tipo numérico (Blalock, 1970; Lazarsfeld, 1973 [1965]; Latiesa, 1994).

Por ello, un aspecto fundamental para el establecimiento de cualquier medición remite al cumplimiento de los requisitos de validez y fiabilidad, esto, en tanto dichas pruebas contribuyen a garantizar el control de los errores de medición.²³ Lo anterior, como ya se advirtió, surge del hecho de que la gran mayoría de las mediciones sociales son indirectas (Blalock, 1970), debido al alto grado de abstracción que presentan los conceptos-variables con los que trabajamos. Si por definición aceptamos que no existen mediciones perfectas - todas poseen un margen de error-, la tarea del analista se remite a determinar cuán grande es el error contenido en las mediciones que realiza (Latiesa, 1994)²⁴. Ello hace necesario establecer el control de los factores transitorios o errores aleatorios (fiabilidad) y factores estables o errores sistemáticos (validez), los que pueden estar contenidos tanto en los instrumentos de recolección de información como en la observación del fenómeno y en la selección de la muestra.

Por ello, informar una actitud comprende un proceso, que a diferencia de lo que se cree, no puede ser definido exclusivamente como una actividad técnica de tipo mecánica. El proceso técnico de construcción de escalas,²⁵ que se denomina escalamiento, debe cumplir con los requisitos de operacionalización del constructo, es decir: a) un trabajo de elaboración conceptual, b) identificación de ítems o enunciados -indicadores- y, c) puesta a prueba de la validez y fiabilidad en la medición, con el objetivo de analizar las proyecciones interpretativas del constructo propuesto. Como señaló Lazarsfeld (1973 [1965]), la relación entre un indicador y el rasgo que se toma como indicación es de probabilidad; es decir, el rasgo no reproduce de manera invariable la conducta. Por tanto, siempre es posible encontrar inconsistencias entre medidas diferentes de un rasgo hipotético, siendo la tarea buscar la combinación que represente la mejor inferencia en relación con las múltiples operaciones empíricas acerca de la característica que subyacen a un constructo - modelación-.

Medir comprende un proceso complejo que no tiene que ver con computar números que se asignan como propiedades de una variable o enunciado. El principio de probabilidad, indica que cada acción de medir requiere un análisis exhaustivo en base al cual evaluar la consistencia de las formulaciones a nivel empírico. En el marco de la evaluación docente, esto exige observar el comportamiento de las escalas tras cada medición, examinando las respuestas dadas para cada enunciado, así como para el conjunto.

A modo de conclusión: ¿Qué tuvo que haber sucedido para aceptar ser clasificados o medidos a través de artefactos?

En el artículo “Estadística, orden natural, orden social” Juan Sánchez Carrión (2001: 36) se pregunta: ¿qué circunstancias se han tenido que dar para que la gente acepte ser resumida por un número abstracto que incluso puede que no coincida con la opinión, valoración, nivel de conocimientos, etc. de ninguna persona en particular? Podemos retomar cierto sentido de la pregunta y plantear ¿qué ha tenido que suceder para que un conjunto de académicos acepten ser “medidos” o “evaluados” por artefactos productores de números, sobre los cuales hay serias dudas de la calidad y rigor involucrado en su construcción? Tras esto, podemos encontrar presente la simple complicidad, el aval desde la lógica de la mecánica tecnicista, la *illusio* o fe en el juego, la ingenuidad frente a la retórica científicista o la retórica de la calidad, etc. Por ello, el título de este trabajo no es antojadizo. La “doxa” en el reino de la “episteme” busca interpelar. En las universidades se teje todo un discurso académico y científicista que apela a la calidad y las exigencias del método, mientras en su seno se implementan ciertos procedimientos para justificar una política llena de “quantofernia” y “artefactualidad”, sin mediar un cuestionamiento al proceso de construcción del dato que involucra toda referencia al mundo social (Merlliè, 1993 [1989]).

En este sentido, muchos de los problemas y requisitos técnicos en el contexto de la evaluación docente no son advertidos. Tenemos que considerar que todo método genera y/o impone ciertas condiciones que dan lugar a determinadas situaciones de interacción o marcos teatrales (Goffman, 2006 [1975]) y al cumplimiento de ciertos supuestos y requerimientos técnicos. Todo ello demanda una voluntad que nos dirija hacia la indagación, algo que va más allá de la razón técnica que se manifiesta en la aplicación de un instrumento según una prescripción: ¿es necesaria la metodología!

No se trata de estar en contra de la evaluación sino a favor de una propuesta de mayor calidad. Como señala Merlliè 1993 [1989], los “artefactos” se pueden enderezar, y para ello la reflexión metodológica exige la revisión retrospectiva para visualizar qué han producido. Si bien se hace difícil pensar en formas alternativas a la encuesta por cuestionario -que ofrecen un procesamiento automatizado para la producción de números a bajo costo y en poco tiempo-, más aún si no se está dispuesto a renunciar a la lógica de la cuantificación; lo menos que podemos hacer es exigir el cumplimiento de los criterios técnicos y las recomendaciones metodológicas que llenan las bibliotecas y bases de datos de revistas especializadas de nuestras universidades. En este sentido, lo que se presenta en este trabajo no es nada nuevo, y esto es lo lamentable, porque reflexiones de este tipo han cumplido más de 50 años.

Agradecimientos

El autor agradece los comentarios realizados a la primera versión del artículo a los académicos del Departamento de Antropología de la Universidad Católica de Temuco, José Díaz Diego y Mario Samaniego Sastre.

Notas

1 La encuesta corresponde al método, el cuestionario al instrumento de registro o recolección, y la escala, al constructo teórico operativo que se orienta a caracterizar el posicionamiento de sujetos respecto de ciertos enunciados que se supone dan cuenta de un aspecto de la “realidad”.

2 De no considerar esto como necesario, sería mucho más rápido y menos costoso solicitar al estudiante “más responsable y de mejor rendimiento académico” de una carrera, que nos dé su opinión acerca del desempeño pedagógico del docente considerando una serie de aspectos.

3 En *El oficio del sociólogo*, Bourdieu, Chamboredon y Passeron (1985 [1973]) sentencian la necesidad de aplicar una jerarquía de actos en la construcción del objeto: conquista, construcción y comprobación.

4 Una reflexión sobre los aspectos operativos ligados a la construcción del objeto pueden ser encontrados en Henríquez y Barriga (2005).

5 De ello se desprende un sistema de recompensas ligadas a la jerarquización, los cargos, las distinciones, los recursos, los incentivos, como diría Bourdieu (2008 [1984]), lo relativo a la acumulación de capital social y simbólico para lograr hegemonía en el campo.

6 La *quantofrenia* se hace presente cuando se argumenta la utilización de información representada por números o numerales, muestras probabilísticas y análisis estadísticos, en tanto permite legitimar un discurso político, social, institucional e incluso académico, brindando de este modo un cariz de cientificidad, una retórica. Esta misma actitud da lugar a la creación de artefactos técnicos, procedimientos objetivados en cuestionarios y en operaciones estadísticas (se hace algo) que carecen, en su construcción, de un conocimiento y una reflexión respecto de sus bases teórico-metodológicas y requerimientos técnicos.

7 Ello implica que cada puntuación se interpreta de igual modo como una opción evaluativa, y que la actitud es el resultado de la sumatoria del total de puntuaciones posibles en la escala.

8 Thurstone sostiene que “En casi cualquier situación de medición se postula un continuo abstracto... la cosa medida de ese continuo se realiza generalmente por medios indirectos, a través de uno o más índices. La verdad se infiere solamente por consistencia relativa de los diferentes índices, ya que nunca se conoce de modo directo. Y nos enfrentamos al mismo tipo de situación cuando intentamos medir la actitud. Necesitamos postular una variable de actitud que es, prácticamente semejante a todos los atributos mensurables de la esencia de un continuo abstracto, y debemos encontrar uno o más índices, los cuales nos satisfarán en el grado en que tengan consistencia interna” (1978: 159).

[9](#) Lo mismo procede para una escala que no involucre conceptos de opinión, sino números; al solicitar a estudiantes posicionarse frente a 19 enunciados con la asignación de valores que van de 1 a 7, la puntuación para la calificación más alta para los 19 sería 7, y para la más baja sería 1 (suma de todas las puntuaciones dividida por el total de enunciados).

[10](#) El primer tipo persigue el objetivo de clasificar los estímulos o juicios -representados en ítems o reactivos- considerando la preferencia de los sujetos, de manera de identificar un orden o jerarquía entre éstos. El segundo tipo, sitúa a los sujetos a través de las respuestas que efectúan en una escala de actitud, de manera de conocer el lugar que ocupan respecto de otros en el continuo del rasgo que se busca medir. El tercer tipo, se encuentra centrado en las respuestas, y considera tanto la variabilidad de los estímulos o juicios como de los sujetos. Otra clasificación propuesta por Morales (2003) establece la diferencia entre respuestas que se piden a los sujetos, a decir, seleccionar alternativas, o responder a una pregunta en términos absolutos. Para el primer tipo, el sujeto debe comparar las alternativas presentadas para escoger unas respuestas; pueden aportar o no datos ipsativos (una respuesta depende de otra, si escoge A no puede optar por B). En el segundo tipo, cada sujeto da respuesta a una pregunta o estímulo de manera independiente (sin necesidad de comparar respuestas); estos no ordenan ni escogen, sino responden a los ítems enunciados según el grado de acuerdo o desacuerdo que manifiesten frente a él. Cada respuesta refleja un valor numérico y la puntuación de cada sujeto es la suma de todas las respuestas.

[11](#) Este tipo de distribución podría dar cuenta de un problema de discriminación de la escala -o de respuestas aleatorias o intención-, al constatar la existencia de posiciones extremas entre un número considerable de sujetos -y una serie de otros problemas asociados a la validez y consistencia de esta-.

[12](#) En la teoría clásica del test, los ítems retenidos para la “medición” se consideran como aquellos seleccionados de un universo de los cuales estos forman parte de una muestra representativa y atingente.

[13](#) Una crítica a estas suposiciones se encuentra en el trabajo de Schatzman y Strauss (1955), así como en la denuncia de la falsa neutralidad de las técnicas realizada por Bourdieu, Chamboredon y Passeron en *El oficio del Sociólogo*.

[14](#) Esto puede ser observado en un instrumento cuando hay cruce de columnas, donde se evidencian set de, por ejemplo, muy de acuerdo y por otro lado, en desacuerdo.

[15](#) Al respecto, Muñiz (1998) reconoce tres perspectivas presentes en la medición psicológica: la de orientación psicofísica y de actitudes -ambas concebidas como escalamiento de estímulo-, y la de escalamiento de sujetos o test.

[16](#) Campbell establecía una distinción entre medición fundamental y medición derivada. La primera dependía de las leyes que relacionan las cantidades del constructo, donde los números pueden asignarse de acuerdo con leyes naturales para de esta manera representar sus propiedades (ejemplo la longitud, la resistencia, el volumen, etc.). La medición derivada se obtiene indirectamente, mediante una relación con medidas extensivas que se define a partir de mediciones fundamentales (un ejemplo es el de la densidad, la cual se obtiene de una relación entre peso y volumen). Para este autor, en psicología, y por tanto respecto de fenómenos sociales, es imposible efectuar mediciones de tipo fundamental ya que se carece de operaciones empíricas que se vinculen con magnitudes.

[17](#) Como señala este autor, “La aproximación de Campbell al problema de la medición era ciertamente restrictiva y tomada al pie de la letra dejaría fuera incluso muchas de las mediciones físicas, por lo cual ya fue criticada por el propio Bertrand Russell...” (1998: 15).

[18](#) La teoría representacional sostiene que los símbolos asignados a los objetos deben representar las relaciones percibidas entre los atributos de los objetos, *si lo que se pretende es que dicha asignación constituya una medición*.

[19](#) Para Campbell eran imperativas las relaciones empíricas, y para Stevens las propiedades de la escala.

[20](#) La noción de reglas de asignación de Stevens, que admite la medición de variables cualitativas (nominales y ordinales), establece serias restricciones a las operaciones que tienen lugar en psicometría -como es el caso del escalamiento de Likert-, en las cuales la numeración ordinal de categorías se puede considerar representación explícita de magnitud y distancia. Ello vulneraría el principio de representación isomórfica que opera en la vinculación entre las propiedades del fenómeno y el sistema de reglas.

[21](#) Existen por lo menos, otras dos formas para su cálculo, siendo su original el denominado método de sigma.

[22](#) La escala de estimaciones sumadas o de Likert se orienta a escalar a sujetos a un nivel de medición denominado ordinal, aun cuando es común que se trabaje como si fuera de intervalo (Marradi y Macri, 2012). De hecho, el objetivo de Likert fue precisamente idear una escala que reprodujera en su distribución características similares a las que expresa una escala numérica, cuestión que permite -con ciertas restricciones interpretativas-, operar con técnicas de análisis para variables cuantitativas. En este caso, los resultados se interpretan como orientadores o se atiende a posibilidades heurísticas que ciertos procedimientos hacen posibles. Como advierte Marradi (2002, 2007), más allá de que el tratamiento de datos siga procedimientos adecuados para atributos numéricos, la interpretación de los resultados debe siempre consignar su "naturaleza ordinal", y por ende, clasificatoria, no métrica. Más allá de discutir si son adecuadas este tipo de técnicas, hay que insistir en las precauciones interpretativas. De todos modos, el análisis de correspondencias múltiples puede ser una alternativa para explorar la estructura de los datos en variables cualitativas, dado que opera bajo distancia ji-cuadrado, que considera la variabilidad en base a distribuciones porcentuales.

[23](#) Respecto al tamaño muestral para prueba del instrumento, Cea D'Ancona (2004) señala que se exige un tamaño elevado, debido a que las muestras pequeñas ofrecen mayor probabilidad de estimaciones de coeficiente de correlación poco fiables. Citando a Comery (1973), la autora presenta una escala que va desde un tamaño muy pobre (50 casos), pobre (100 casos), justo (200 casos), bueno (300) y muy bueno (500), siendo el mínimo recomendado entre 200 y 300 casos. Otros autores proponen como regla la existencia de al menos 5 casos por variable considerada en el análisis.

[24](#) Una breve, pero precisa aproximación a las fuentes de error puede encontrarse en el texto de esta autora (ver referencias en bibliografía).

[25](#) Una escala comprende un conjunto de ítems (expresados a través de enunciados en forma de proposiciones, frases, preguntas) que permite medir el nivel o el grado que alcanza un determinado atributo; es un instrumento de medición que hace uso del nivel semántico y de la codificación numérica.

Anexo: Escala utilizada en la en la determinación de opiniones al desempeño docente.

Dimensión	Enunciado
Metodología y preparación	Es claro en la exposición de los contenidos tratados en clase.
	Cuando procede, muestra la relación del curso con el futuro profesional
	Facilita nuestro aprendizaje con la utilización de diferentes estrategias de enseñanza.
	Nos estimula el autoaprendizaje.
	Relaciona las materias que presenta con los conocimientos previos que tenemos.
	Cuando nos equivocamos, lo aprovecha como fuente de aprendizaje.
	Fomenta el trabajo grupal y la cooperación entre los estudiantes.
Interacción con los estudiantes	Fomenta el diálogo sobre los contenidos tratados en clase.
	Está abierto a considerar nuestros puntos de vista en el desarrollo de las clases.
	Muestra interés por el desarrollo integral de sus estudiantes.
Evaluación	Las evaluaciones las realiza con pautas claras, objetivas y conocidas.
	Revisamos nuestros aciertos y errores en las evaluaciones para aprender de ellos.
	Realiza las evaluaciones en base a los contenidos y/o competencias trabajadas.
Administración	Nos entregó y explicó el programa del curso o guía de aprendizaje.
	Es puntual en el inicio y termino de las clases.
	Cumple el horario de atención de estudiantes.
	Desarrolla todos los contenidos del programa o guía de aprendizaje a lo largo del semestre.
	Realiza las evaluaciones en las fechas programadas
	Nos estimula en el uso de la bibliografía del programa del curso o guía de aprendizaje.

Bibliografía

Alcántara, A. (2000). Tendencias mundiales en la educación superior. El papel de los organismos multilaterales. En D. Cazés, L. Porter y E. Ibarra (coords.). *Reconociendo a la Universidad: sus Transformaciones y su Porvenir* (pp. 81-102). (Tomo I Estado, Universidad y Sociedad: entre la Globalización y la Democratización) México. CEIICH-UNAM.

Asún, R. (2006). "Medir la realidad social: El sentido de la metodología cuantitativa. En M. Canales (Editor). *Metodología de la Investigación Social*, Santiago: Lom.

Barriga, O. (2008). La geometría del espacio social. Una forma de superar la brecha entre el número y la palabra. En N. Cohen y Piovani, J. *La metodología de investigación en debate* (pp. 219-241). La Plata: Editorial de la Universidad Nacional de La Plata.

Blalock, H. (1971). *Introducción a la investigación social*. Buenos Aires: Amorrortu.

Bourdieu, P. (2008 [1984]). *Homo academicus*. Buenos Aires: Siglo XXI.

Bourdieu, P., Chamboredon, J. y Passeron, J. (1985 [1973]). *El oficio del sociólogo*. Buenos Aires: Siglo XXI.

Bruner, J. (1997). Educación superior, integración económica y globalización. *Perfiles Educativos*, Tercera Época, 19, (76-77), 6-15.

Carmines E. y Zeller R. (1979). *Reliability and Validity Assessment*. Beverly Hills, CA: Sage.

Cea D'Ancona, M. (2004). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Síntesis.

Cicourel, A. (1968). *The Social Organization of Juvenile Justice*. Londres: John Wiley & Sons.

Cortada de Kohan, N. (2004). *Teoría y Métodos para la construcción de escalas de actitud*. Buenos Aires: Lugar.

Cook, S. y Selltiz, C. (1978). "Enfoque de indicadores múltiples en la medición de actitudes". En G. Summer (Editor). *Medición de actitudes* (pp. 361-387). México: Trillas.

Chiroleu, A. e Iazzetta, O. (2005). La reforma de la educación superior como capítulo de la reforma del Estado. Peculiaridades y trazos comunes. En E. Rinesi, G. Soprano y C. Suasnábar (Comps.). *Universidad: reformas y desafíos. Dilemas de la educación superior en la Argentina y el Brasil* (pp. 15-38). Buenos Aires: Prometeo.

Elejabarrieta, F. e Iñiguez, L. (1984). Construcción de escala de actitud tipo Thurst y Likert. Universidad Autónoma de Barcelona. Barcelona. Recuperado de <http://antalya.uab.es/liniguez/Materiales/escalas.pdf>

Elizalde, L. y Reyes, R. (2008). Elementos claves para la evaluación del desempeño de los docentes. *Revista Electrónica de Investigación Educativa*, Número especial, 2-13.

Fernández, N. y Coppola, N. (2008). La evaluación de la docencia universitaria en Argentina. Situación, problemas, perspectivas. En *Revista Iberoamericana de Evaluación Educativa*, 1 (3), 97-123.

Gasperoni, G. y Marradi, A. (2002). Presentación. En G. Gasperoni y A. Marradi. *Construire il dato 3. Le escale Likert* (pp.7-14). Milano: Franco Angelini

Goffman, I. (2006). *Frame Analysis*. Madrid: Siglo XXI.

Ghiglione, R. y Matalon, B. (1989). *Las encuestas sociológicas. Teoría y práctica*. México: Trillas.

Henríquez, G. y Barriga, O. (2005). El rombo de la investigación. *Cinta de Moebio*, (23). Recuperado de <http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=10102302>

Jañez, L. (1989). *Fundamentos de la psicología matemática*. Pirámide: Madrid.

Kreimer, P. (2005). Sobre el nacimiento, desarrollo y demolición de los papers. En D. Golombek. *Demoliendo paper: La trastienda de las publicaciones científicas*. Buenos Aires: Siglo XXI.

Latiesa, M. (1994). "Validez y fiabilidad de las observaciones sociológicas". En M. García-Ferrando, J. Ibáñez y F. Alvira (eds.). *El análisis de la realidad social. Métodos y técnicas de investigación social* (pp.275-333). Madrid: Alianza.

Lazarsfeld, P. (1965). De los conceptos a los índices empíricos. En R. Boudon, y P. Lazarsfeld. *Metodología de las ciencias sociales. Conceptos e índices* (pp. 35-46). Baelona: Laia.

Levy, D. (1994). Introducción. En D. Levy (coord.). *La educación superior dentro de las transformaciones políticas y económicas de los años noventa. Informe del grupo de trabajo sobre educación superior de la asociación de estudios latinoamericanos*. Buenos Aires: CEDES, Argentina. Recuperado de <http://bibliotecavirtual.clacso.org.ar/ar/libros/argentina/cedes/levy.rtf>

Likert, R. (1978 [1932]). Una técnica para medir actitudes. En G. Summer (Editor), *Medición de actitudes* (pp. 182-193). México: Trillas.

Llomovatte, S. (2006). Para una crítica del modelo de la triple hélice: universidad, empresa y Estado. En S. Llomovatte, F. Juarros, J. Naidorf, y A. Guelman. *La vinculación universidad-empresa: una crítica desde la universidad pública*. Buenos Aires: Miño y Dávila Editores.

Marradi, A. (2002). Le scale Likert e la razione all'oggetto. En G. Gasperoni y A. Marradi. *Costruire il dato 3. Le escale Likert* (pp.15-52). Milano: Franco Angelini.

Marradi, A. (2007). Clasificación, conteo, medición, construcción de escalas. En A. Marradi, N. Archenti y J. Piovani. *Metodología de las ciencias sociales*. Emecé: Buenos Aires.

Marradi, A. y Macri, E. (2012). Sono equidistanti le categorie di una scala Likert? Alcune risultanze di ricerca. *Cambio. Revista sulle trasformazioni sociali*, 2 (3), 171-188.

Merlliè, D. (1993 [1989]). La construcción estadística. En P. Champagne, R. Lenoir, D. Merlliè y L. Pinto. *Iniciación a la práctica sociológica*. México: Siglo XXI.

Mollis, M. (2003). Presentación. En M. Mollis (Comp.). *Las universidades en América Latina: ¿reformadas o alteradas? La cosmética del poder financiero*. (pp. 9-15). Buenos Aires: Clacso.

Muñiz, J. (1998). La medición de lo psicológico. En Revista *Psicothema*, (10), 1, 1-21.

Navarro, P. (1995). La encuesta como texto: un enfoque cualitativo. Ponencia presentada en el V Congreso Español de Sociología. Granada, septiembre de 1995.

Padua, J. y Ahman, I. (1994). Escalas para la medición de actitudes". En J. Padua. *Técnicas de investigación aplicadas a las Ciencias Sociales*. México: Fondo de Cultura Económica.

Prat, R. y Doval, E. (2003). Construcción y análisis estadístico de escalas. En J. P. Lévy y J. Varela (Eds.). *Análisis multivariante para la Ciencias sociales* (pp. 44-89). Madrid: Pearson Educación.

Sánchez, J. (2001). Estadística, orden natural y orden social. Revista *Papers*, 63 (64), 33-46.

Schatzman, L. y Strauss, A. (1955). Social Class and Models of Communication. *The American Journal of Sociology*, 60, (4), 329-338. Recuperado de <http://www.jstor.org/stable/2772025>

Sherif, M y Sherif, C. (1978). La actitud como estrategia personal del individuo: el enfoque de la implicación y juicio social de la actitud y del cambio de actitud. En G. Summer (Editor). *Medición de actitudes* (pp. 361-387). México: Trillas.

Schugurensky, D. (2000). Autonomía, heteronomía, y los dilemas de la educación superior en la transición al siglo 21; caso Canadá. En R. Rodríguez (Coord.). *Reformas en los sistemas nacionales de educación superior*. Coruña: Netbiblo.

Sorokin, P. (1957). *Achaques y manías de la sociología moderna*. Madrid: Águila.

Stevens, S. (1946). On the theory of scale of measurement, *Science*, 103 (2684), 667-680. Recuperado de <http://www.sciencemag.org/content/103/2684/677.full.pdf?sid=7f32b588-95ec-4c0d-b242-89c1059641b2>

Summer, G. (1978). Introducción. En G. Summer (Editor). *Medición de actitudes* (pp. 361-387). México: Trillas.

Thurstone, L. (1978). Las actitudes pueden medirse. En G. Summer (Editor). *Medición de actitudes* (pp. 157-173). México: Trillas.

Torgenson, W. (1971 [1958]). La naturaleza de la medición. En: M. Mora y Araujo, et. al. *Medición y construcción de índices* (pp. 83-124). Buenos Aires: Nueva Visión.

Tünnermann, C. (2003). *La universidad latinoamericana ante los retos del siglo XXI*. México: Unión de Universidades Latinoamericanas.

Vivanco, M. (2005). *Muestreo estadístico. Diseño y aplicaciones*. Santiago: Editorial Universitaria.

Recibido: 7 de diciembre de 2013

Aceptado: 1 de junio de 2014

Publicado: 1 de junio de 2015