

Vol. 20, Núm. 1, 2018

Diseño del formato de escalas tipo Likert: un estado de la cuestión

Likert-Type Scale Format Design: State of Art

Antonio Matas (*) amatas@uma.es

(*) Universidad de Málaga
(Recibido: 7 de abril de 2016; Aceptado para su publicación: 2 de agosto de 2016)

Cómo citar: Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista Electrónica de Investigación Educativa*, 20(1), 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>

Resumen

Las escalas tipo Likert constituyen uno de los instrumentos más utilizados en Ciencias Sociales y estudios de mercado. No obstante, existe cierto debate sobre el formato de estas escalas para obtener datos de calidad. En este artículo se presenta un estudio de revisión sobre los formatos de las escalas tipo Likert y su relación con la calidad de los datos que es posible obtener con ellos. Para ello se ha realizado una revisión sistemática junto con una síntesis del contenido, identificando un conjunto de temas o tópicos relevantes por sus implicaciones metodológicas: sesgos y estilos de respuesta, el número óptimo de alternativas, las razones de elección de la alternativa intermedia y el efecto de la alternativa "No opino". Con base en el análisis y discusión de estos temas se presentan algunas recomendaciones de formato en las conclusiones del estudio.

Palabras clave: Escala Likert, encuestas, datos de investigación, sesgo.

Abstract

Likert-type scales are one of the most commonly used instruments in social sciences and market research. Nonetheless, there is some debate regarding the format needed in these scales to obtain quality data. This study is a review of Likert-type scale formats and their association with data quality. To that effect, a systematic review was performed, along with a synthesis of content, to identify a series of topics that are relevant due to their methodological implications: response bias and styles, the optimal number of alternatives, reasons for choosing the intermediate response, and the effect of the "No Opinion" alternative. On the basis of an analysis and discussion of these topics, the study presents some recommendations for formatting in the conclusions.

Keywords: Likert scale, surveys, survey data, statistical bias.

I. Introducción

El propósito de esta revisión fue examinar el efecto del formato de las escalas tipo Likert en la calidad de los datos a través del análisis de publicaciones recientes. Aunque el concepto de calidad de los datos no es unívoco, en este documento se toma como referencia el acercamiento conceptual de la OCDE (2011) donde se entiende como un concepto multifacético que hace referencia a la idoneidad de un proceso de investigación en función de las necesidades de los usuarios. El concepto de calidad de los datos implica aspectos como la precisión, la oportunidad, la accesibilidad, la interpretabilidad o la coherencia durante su uso y gestión de los mismos (Matus, 2007).

Las llamadas “escalas Likert” son instrumentos psicométricos donde el encuestado debe indicar su acuerdo o desacuerdo sobre una afirmación, ítem o reactivo, lo que se realiza a través de una escala ordenada y unidimensional (Bertram, 2008). Estos instrumentos suelen ser reconocidos entre los más utilizados para la medición en Ciencias Sociales (Cañadas y Sánchez-Bruno, 1998; Dawes, 1975). Este tipo de escala surgió en 1932, cuando Rensis Likert (1903-1981) publicó un informe en el que exponía cómo usar un tipo de instrumento para la medición de las actitudes (Likert, 1932; Edmondson, 2005).

Originalmente, este tipo de instrumentos consistía en una colección de ítems, la mitad expresando una posición acorde con la actitud a medir y la otra mitad en contra. Cada ítem iba acompañado de una escala de valoración ordinal. Esta escala incluía un punto medio neutral así como puntos a izquierda y derecha, originalmente de desacuerdo y de acuerdo, con opciones de respuesta numéricas de 1 a 5. La escala de alternativas aparecían en horizontal, uniformemente espaciadas, al lado del ítem e incluyendo las etiquetas numéricas.

Es importante tener en cuenta que las escalas que utilizan alternativas de respuesta no están vinculadas con el acuerdo o desacuerdo con los ítems, no son escalas Likert en sentido original. No obstante, es frecuente que se les denomine escalas “tipo Likert” por generalización (Nadler, Weston y Voyles, 2015).

Por otro lado, es necesario hacer hincapié en que estas escalas se diseñan a un nivel de medida ordinal (Bertram, 2008; Dawes, 1975; Edmondson, 2005; Likert, 1932; Salkind, 1998). Las evidencias indican que, efectivamente, las distancias psicológicas entre alternativas de la escala no son iguales (Kennedy, Riquier y Sharp, 1996) lo que tiene implicaciones directas en el proceso de medición y, por tanto, en la calidad de los datos, así como en la interpretabilidad de los resultados (Pardo, 2002).

En la década de los cincuenta del siglo XX se desarrolló una línea de investigación centrada en el efecto del formato de las escalas tipo Likert en la calidad de los datos; dicha línea experimentó un resurgimiento en la década de los ochenta y continúa hasta el momento (Velez y Ashworth, 2007).

La investigación en este campo, que metodológicamente tiene una importancia destacada, está justificada por varios motivos. En primer lugar, porque se trata de un instrumento muy utilizados en Ciencias Sociales, Ciencias de la Educación, Ciencias de la Salud, etc., tal como se ha dicho anteriormente. En segundo lugar porque las técnicas actuales suelen ser muy sensibles a la calidad de los datos, según el concepto que se expuso al inicio de este artículo, por lo que es imprescindible conocer qué tipo de dato registra cada instrumento (Dawes, 2008). Y en tercer lugar, porque no parece haber un acuerdo en cuanto al formato óptimo de las escalas (Adelson y McCoach, 2010).

El objetivo de este estudio fue hacer una revisión sistemática de las publicaciones más recientes o significativas, que de alguna u otra forma hayan estudiado el efecto del formato de las escalas tipo Likert en la calidad de los datos. En segundo lugar, se buscaron recomendaciones que pudieran servir tanto para el diseño y uso de las escalas, como base para profundizar en el tema en futuras investigaciones.

II. Metodología

Para el propósito del estudio se llevó a cabo una revisión sistemática de la literatura. Este tipo de estudios suele utilizar el meta-análisis (Higgins y Green, 2011); sin embargo, el meta-análisis no fue apropiado en este caso, ya que sólo es aplicable cuando los datos son homogéneos entre estudios (Eysenck, 1995). Cuando los datos, los tamaños de las muestras y las variables son de naturaleza heterogénea es más apropiada la síntesis de tipo no estadístico (Bland, Meurer y Maldonado, 1995). Puesto que la literatura sobre la forma y el diseño del formato de las escalas tipo Likert presenta un alto nivel de heterogeneidad en términos de cuestiones analizadas, diseños de investigación, modelos teóricos, variables, instrumentos, etc. (ver resultados) se optó por una síntesis no estadística o revisión sistemática.

2.1 Estrategia de búsqueda

Se realizó una exhaustiva búsqueda en la literatura especializada a través de dos estrategias: por un lado se buscó en las bases de datos digitalizadas Scopus y ERIC (Education Resources Information Center), y por otro, se llevó a cabo una estrategia "manual", consistente en la búsqueda sistemática de referencias temáticas relevantes identificadas, principalmente, en la sección de referencias de los artículos recuperados en la primera oleada de búsqueda. Esta consiste en una revisión exploratoria de las referencias de los artículos recuperados, con el fin de identificar publicaciones significativas o relevantes para la investigación.

A continuación se revisaron los resúmenes (*abstracts*) de los artículos potencialmente relevantes. La revisión fue realizada por dos jueces independientes especialistas en Métodos de Investigación en Ciencias Sociales. La selección final de los artículos fue puesta en común por los jueces. Los desacuerdos fueron debatidos hasta conseguir un consenso sobre la selección final.

2.2 Criterios de inclusión y exclusión

Los criterios de inclusión fueron:

- Que las palabras clave incluyeran los siguientes términos: escalas Likert (*Likert scale*), número de alternativas (*number of choices*), y alternativa intermedia u opción intermedia (*midpoint*).
- Tratarse de un artículo de investigación, revisión o comunicación científica. Se excluyeron monografías y ensayos para centrarse principalmente en trabajos de tipo empírico.
- Estar vinculado al ámbito de las Ciencias Sociales y Ciencias de la Salud, por ser estos los más afines al campo donde suelen utilizarse este tipo de instrumentos (Dawes, 1975).

En la búsqueda en las bases de datos se estableció la fecha de inicio de recuperación en 2010 para obtener los resultados más actualizados. Sin embargo, en la segunda estrategia de búsqueda no se limitó el tiempo para garantizar una recuperación de la literatura clave vinculada al tema.

2.3 Análisis

Los jueces expertos clasificaron el contenido de los artículos aplicando un proceso de categorización *a posteriori*. Para garantizar la calidad de la clasificación ambos jueces siguieron una estrategia de saturación, consistente en la revisión crítica de los resultados, tomados provisionales, al mismo tiempo que se trata de encontrar argumentos en contra de dichos resultados (Hopkins, 1989). Este análisis fue realizado de forma independiente por cada juez. Posteriormente, llevaron a cabo un proceso de triangulación (Tójar, 2006), que consistió en una puesta en común para establecer un consenso sobre los temas tratados en la literatura revisada. El proceso concluyó con la identificación de los distintos temas (tópicos) tratados en los artículos (ver la sección de resultados).

III. Resultados

De las bases de datos se seleccionaron 16 artículos a los que se añadieron 12 en la segunda estrategia de búsqueda. Del total se eliminaron 17 por no cumplir alguno de los criterios de inclusión. El número final de artículos seleccionados fue de 11, de los cuales uno era de revisión, otro de simulación numérica y el resto empíricos.

Las características de las muestras empleadas en los estudios empíricos analizados son muy diversas. Así, dos estudios tienen muestras grandes (superiores a los 15 mil participantes), cuatro tienen muestras entre los 500 y mil participantes, y sólo dos estudios tienen muestras menores a 200 participantes. Además, las características sociodemográficas presentan una elevada variabilidad entre investigaciones.

Se procedió a valorar la calidad metodológica de los artículos seleccionados tomando como guía los siguientes criterios (adaptado de Agli, Bailly y Ferrand, 2015) que frente a otros sistemas de valoración basados en escalas (Silva, Valdivia Arancibia, Iop, Gutierrez Filho y Silva, 2013) presentan una mayor flexibilidad y facilidad de adaptación a distintos estudios:

- Adecuado tamaño muestral en función del objetivo de estudio.
- Descripción del método de muestreo.
- Descripción clara de los criterios de inclusión o exclusión de casos.
- Una descripción adecuada de los participantes.
- Uso de medidas fiables y válidas.
- Uso de técnicas analíticas apropiadas.
- Discusión y conclusiones coherentes con los resultados obtenidos.

Teniendo en cuenta los criterios anteriores, todos los artículos seleccionados fueron valorados como de alta calidad.

3.1 Categorías temáticas o tópicos

A continuación se exponen las temáticas que aparecen vinculadas a la calidad de los datos, su interpretación, utilidad y, en definitiva, las que garantizan que los resultados de los estudios satisfagan los objetivos planteados en esos mismos estudios. Las temáticas que surgieron de la literatura revisada son: los sesgos en las respuestas, la opción o alternativa intermedia, los argumentos de los participantes para la elección de una respuesta, y la inclusión o no de una alternativa que registre el no tener opinión sobre el ítem.

Sesgos. Un tema recurrente es el relativo a los sesgos, principalmente los conocidos como estilos de respuesta. Un estilo de respuesta es la tendencia sistemática del encuestado a contestar sesgadamente con relación al constructo de interés (Paulhus, 1991). Los estilos de respuesta en este tipo de escalas pueden agruparse en cuatro grupos, tal como sugieren He y Von De Vijver (2015):

1. Estilo de Respuesta Aquiescente (ERA): es la tendencia a elegir opciones que están de acuerdo con el objeto del ítem, independientemente de su contenido.
2. Estilo de Respuesta Extrema (ERE): es la tendencia a elegir las alternativas extremas.
3. Estilo de respuesta intermedia (ERI): es la tendencia a elegir las alternativas moderadas, evitando las extremas.
4. Respuesta socialmente deseable (RSD): es la tendencia a responder de la manera que se espera que sea mejor vista por los demás. Se puede entender como una forma de deseabilidad social (Nadler, Weston y Voyles, 2015).

En la literatura se recoge también el llamado sesgo negativo. Este sesgo consiste en registrar una mayor frecuencia de respuestas en las alternativas de la derecha de la escala (normalmente los que están a favor del sentido del ítem). Aunque el sesgo negativo ha sido identificado habitualmente en las escalas Likert

(Peterson y Wilson, 1992) parece que un mayor número de alternativas lo corrige (Dawes, 2008).

Número de alternativas u opciones de respuesta. Con relación al número de alternativas existe una línea de trabajo centrada en analizar cómo el número de alternativas de respuesta afecta a las propiedades psicométricas de las escalas tipo Likert (Cox, 1980; Bishop, 1987; Dawes, 2008; Lozano, García-Cueto y Muñiz, 2008; Oaster, 1989). En 1933, Pemberton fue uno de los primeros investigadores en esta línea. La conclusión era que la confiabilidad aumentaba cuando las alternativas se incrementaban de cinco a siete. En general, estas conclusiones se han confirmado a través de estudios posteriores, según la literatura consultada. No obstante, este incremento es apreciable entre cuatro y siete alternativas, siendo menos apreciable a partir de siete alternativas (Cummins y Gullone, 2000; Dillman, 2007).

En menor medida se ha estudiado el efecto del número de alternativas sobre estadísticos como la media o la varianza. En este sentido, Dawes (2008) no encontró diferencias significativas en medias, varianzas, ni curtosis al comparar tres versiones de un mismo instrumento con escalas de cinco, siete y diez alternativas.

La alternativa intermedia. Entre los tópicos estudiados con relación al formato, destacan los estudios que analizan los argumentos a favor y en contra de incluir una opción intermedia (opción de indecisión o ambivalencia) entre las alternativas de respuesta. Estos argumentos están estrechamente vinculados a los sesgos de tendencia central y deseabilidad social, tal como se ha puesto de manifiesto en varios estudios (Baka y Figgou, 2012; Johns, 2005).

Diversos estudios han puesto en evidencia que la elección de la alternativa intermedia puede darse por muy diversos motivos, al margen del posicionamiento real de la persona (Velez y Ashworth, 2007). Esta diversidad de argumentos puede generar un aumento de la frecuencia de elección de la alternativa intermedia, provocando la aparición del sesgo de tendencia central.

Existen evidencias de que la alternativa intermedia suele ser elegida entre el 20% y el 50% en las escalas Likert (Johns, 2005). En un estudio más reciente, Baka y Figgou (2012) registraron hasta un 88.65% de respuestas intermedias dadas por una muestra de 71 voluntarios a una escala sobre cuestiones políticas.

Ante este sesgo algunos investigadores toman la decisión de usar escalas pares, sin alternativa intermedia. Sin embargo, las razones para incluir o no incluir esta alternativa son causa de debate (Ayidiya y McClendon, 1990; Baka y Figgou, 2012; Bishop, 1987; Moors, 2008). Así, Garland (1991) afirmaba que la eliminación de esta alternativa disminuía el sesgo de deseabilidad social. En encuestas de orden político se encontraba que la frecuencia de elección de opciones extremas era más equitativa cuando no se incluía la alternativa intermedia (Schuman y Presser, 1981).

Otros estudios apuntan a que la presencia de la alternativa intermedia en la escala tiene un efecto moderado en la validez (Scherpenzeel y Saris, 1997). Sin embargo, algunos autores argumentan que la eliminación de la alternativa intermedia obliga al encuestado a posicionarse a favor o en contra del ítem. Por tal motivo, estos autores recomiendan la inclusión de las categorías intermedias, argumentando que es preferible no forzar esta elección para evitar inducir un error en los datos (Converse, 1970; Schuman y Presser, 1981; Sudman y Bradburn, 1989).

En este debate también intervienen los estudios que apuntan a que la elección de la alternativa intermedia está afectada por el contenido de la escala de la encuesta (Klopfer y Madden, 1980). En este sentido, hay que prestar atención a la legibilidad y claridad del ítem (Baka y Figgou, 2012) un aspecto poco tratado en la literatura consultada (Saris, Veld, y Gallhofer, 2004).

La legibilidad (*readability*) se puede entender de diversas formas, entre otras, como dificultad del lenguaje y como claridad del lenguaje. La dificultad del lenguaje puede valorarse a través del índice de Rudolf Flesch (1948); y la claridad, por su parte, que tiene que ver con la ambigüedad del lenguaje, suele valorarse a través de jueces expertos de reconocido prestigio que emiten un informe razonado y

justificado sobre las características del discurso. En ambos casos, la calidad de los datos se puede ver afectada. De hecho, se ha registrado cómo una legibilidad difícil está relacionada con un aumento de elección de las alternativas intermedias; es decir, que ante la dificultad de comprender el ítem, los encuestados eligen con más frecuencia las opciones intermedias de la escala (Velez y Ashworth, 2007).

Argumentos de selección. Tal como se ha adelantado, las razones para elegir la alternativa intermedia pueden ser muy diversas. Aunque los estudios al respecto no son muchos, destacan las consecuencias que de ellos se extraen para el diseño de escalas. Entre ellos, Johns (2005) recoge cuatro tipos de electores de las alternativas intermedias:

1. Los ambivalentes. Es el grupo de encuestados que eligen la opción intermedia porque realmente se posicionan entre los dos extremos.
2. Los indiferentes. Son aquellos que no tienen interés por el tema o no se implican en contestar de forma competente.
3. Los aquiescentes. Es el grupo de encuestados que tratan de responder lo que piensan que los demás esperan que contesten.
4. Los que evitan u ocultan opiniones socialmente no deseables. Es el grupo que evita elegir opciones que están mal vistas entre su grupo de iguales o de referencia.

Por su parte, Hernández, Espejo, Gómez-Benito y González-Romá (2001) realizaron una revisión de los argumentos que, según la literatura, pueden llevar a los encuestados a seleccionar las alternativas intermedias. Una de ellas era la de sentirse en el punto medio por indiferencia o ambivalencia, que sería las razones originales previstas en las escalas Likert. Sin embargo, otras razones eran las siguientes:

- La falta de comprensión del enunciado.
- La falta de competencia o información relativa al enunciado.
- El recelo a revelar cuestiones personales.
- Y la falta de familiaridad o baja frecuencia de uso de los ítems empleados en una lista de adjetivos.

Más recientemente, Baka y Figgou (2012) llegaron a la conclusión de que existen dos tipos de encuestados que eligen la alternativa intermedia. Por un lado los que no tienen formada una opinión y, por otro, los indecisos. Estos autores identificaron cuatro grandes argumentos (entre sus encuestados) para elegir la opción intermedia: ambivalencia (el 44% de las respuestas intermedias), no tener información sobre el tema (el 33% de las respuestas intermedias), discuten o no están de acuerdo con el tema del ítem o del cuestionario (el 30% de las respuestas intermedias), y desconocimiento o indiferencia sobre el tema (el 25% de las respuestas intermedias).

Alternativas “sin opinión”. Otro tópico presente en la literatura consultada es el relativo a la inclusión de la alternativa “sin opinión”, “no opino” o “no sé”. Esta opción suele colocarse al margen de la escala ordinal. Los resultados obtenidos por Baka y Figgou (2012) concluían que los encuestados eligen esta alternativa principalmente cuando no tienen información o conocimiento sobre el tema que trata el ítem. Otras razones que hallaron fueron que los encuestados elegían esta opción cuando no tenían interés por el tema de la escala, no entendían lingüísticamente el ítem (por la complejidad o desconocimiento de las palabras usadas, o por falta de claridad en la redacción) o que el encuestado tenía una perspectiva totalmente opuesta al argumento racional del ítem.

Las evidencias halladas en la literatura consultada ponen de manifiesto cierta interacción entre la alternativa intermedia y la opción “sin opinión” o “no se” (SO/NS). En los estudios se llega a la conclusión de que la opción SO/NS es socialmente menos deseable que la alternativa intermedia (Johns, 2005; Nadler, Weston y Voyles, 2015). Así, cuando el encuestado no comprende el ítem, y además no se dispone de la opción SO/NS, la persona que contesta puede plantearse tres escenarios (Velez y Ashworth, 2007):

- No responder al ítem.
- Elegir una alternativa cualquiera, tanto a favor como en contra. La distribución de esta elección sería azarosa.
- Elegir la alternativa intermedia.

En cualquiera de los escenarios, la calidad de los datos se vería afectada, puesto que la alternativa intermedia no sería una respuesta típica de los encuestados con niveles intermedios de opinión.

IV. Discusión y conclusiones

El objetivo de este artículo ha sido ofrecer una revisión de la literatura científica sobre el formato de las escalas tipo Likert y su posible efecto en la calidad de los datos.

En los artículos revisados aparece reiteradamente una serie de sesgos vinculados a estas escalas. La mayor parte de los sesgos se han asociado con los estilos de respuesta. Esto puede implicar que las personas tienden a reaccionar de una determinada manera al ítem, no sólo por el contenido y el formato, sino también por su propio estilo de afrontar este tipo de tareas. Destaca en la literatura la coherencia entre los argumentos de selección de la opción intermedia hallados por Johns (2005) con los estilos de respuesta recogidos por He y Von De Vijvar (2015).

Otro aspecto destacado en la literatura es el número óptimo de alternativas. Se trata de un debate aún sin resolver. Teniendo en cuenta los estudios sobre este tópico y efectos del contenido de los ítems, algunos autores sugieren que las escalas tipo Likert deberían tener cuatro opciones de respuesta, añadiendo la opción SO/NS (Nadler, Weston y Voyles, 2015) o bien, escalas de siete opciones para el caso de aquellas que se administran vía Internet para evitar una excesiva interpolación, tal como sugiere Finstad (2010) en las conclusiones de su investigación.

Por otro lado, Dawes (2008) apunta en sentido contrario, puesto que no encuentra diferencias en los estadísticos descriptivos de posición, dispersión y forma en tres versiones de un instrumento que analiza. Sin embargo, cabe señalar que dos de las versiones que utiliza este autor coincidían con el número óptimo de alternativas comentado anteriormente (de cinco y siete). En este sentido, sería necesario replicar el estudio de Dawes (2008) con versiones de escala que no se ajusten al número óptimo de alternativas.

Otro aspecto a comentar con relación al número óptimo de alternativas es el relativo al tipo de metodología usado en los estudios. Se observa que los estudios de simulación, como el realizado por Lozano, García-Cueto y Muñiz (2008) recomiendan la utilización de escalas de alternativas pares, en concreto de cuatro opciones. Sin embargo, los estudios empíricos suelen señalar las bondades de las escalas impares, o bien, de escalas con la alternativa de no opinión. Esto sugiere que los datos simulados no recogen toda la cualidad de los datos recogidos de muestras reales. Esta cuestión debe tenerse en cuenta para estudios sobre el tema.

Además de los sesgos más recurrentes en la literatura (tendencia central, deseabilidad social, etc.) existen otros aspectos a considerar en el diseño de las escalas. Un artículo se ocupa especialmente del tema de la legibilidad (Vékez y Ashwartz, 2007); dada su potencial trascendencia, sus argumentos y sugerencias deberían tenerse en cuenta, a pesar de que en el resto de literatura sólo son citados de forma ocasional (Baka y Figgou, 2012).

La principal conclusión es el número relativamente bajo de investigaciones publicadas sobre el tema, a pesar de que el diseño o forma de las escalas tipo Likert es fundamental para obtener datos de calidad. No obstante, a partir de la literatura consultada se han podido identificar algunos aspectos sobre la forma más recomendable para la elaboración de este tipo de escalas. Entre ellos destacan los siguientes:

- Usar escalas de cinco alternativas junto con una opción de “No tengo opinión”, “No opino” o “Sin opinión”.
- Cuidar y adaptar el lenguaje, tanto en los términos usados como en la estructura gramatical, al nivel sociocultural de la población diana.

Por último, con relación a futuras investigaciones, aparecen cuestiones a tener en cuenta: con relación al objetivo de investigación, estudiar con mayor profundidad los argumentos que las personas tienen para elegir las distintas alternativas; con relación a la metodología de estudio, desarrollar principalmente estudios que se pueden completar con estudios de simulación.

En cualquier caso, el debate sobre el formato y contenido de los ítems en las escalas tipo Likert sigue abierto.

Referencias

Adelson, J. L. y McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement*, 70(5), 796-807.

Agli, O., Bailly, N. y Ferrand, C. (2015). Spirituality and religion in older adults with dementia: a systematic review. *International Psychogeriatrics*, 27(5), 715-72.

Ayidiya, S. y McClendon, M. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, 54, 229-247.

Baka, A. y Figgou, L. (2012). ‘Neither agree, nor disagree’: a critical analysis of the middle answer category in voting advice applications. *International Journal Governance*, 5(3/4), 244-263.

Bertram, D. (2008). *Likert Scales... are the meaning of life*. Topic report: Recuperado de <http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf>

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-232.

Bland, C. J., Meurer, L. N. y Maldonado, G. (1995). A systematic approach to conducting a non-statistical meta-analysis of research literature. *Academic Medicine: Journal of the Association of American Medical Colleges*, 70, 642-653.

Cañadas, I. y Sánchez-Bruno, A. (1998). Categorías de respuestas en escalas tipo Likert. *Psicothema*, 10(3), 623-631.

Converse, P. E. (1970). Attitudes and non-attitudes: continuation of a dialogue. En E. R. Tufte (Ed.), *The quantitative analysis of social problems* (pp. 168-189). Reading, MA: Addison-Wesley.

Cox, E. (1980). The optimal number of response alternatives in a scale: A review. *Journal of Marketing Research*, 17, 407-422.

Cummins, R. A. y Gullone, E. (2000). *Why we should not use 5-point Likert scales: The case for subjective quality of life measurement*. Actas de la II Conference on Quality of Life in Cities (pp. 74-93). National University of Singapore.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*, 50(1), 1-19.

- Dawes, R. M. (1975). *Fundamentos y técnicas de medición de actitudes*. México: Limusa.
- Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method*. Nueva York: John Wiley y Sons.
- Edmondson, D. R. (2005). Likert scales. A history. *Conference on Historical Analysis y Research in Marketing Proceedings*, 12, 127-133.
- Eysenck, H. J. (1995). Problems with meta-analysis. En I. Chalmers y G. G. Altman (Eds.), *Systematic reviews* (pp. 64-74). Londres: BMJ Publishing Group.
- Finstad, K. (2010). Reponse interpolation and scale sensitivity: evidence against 5-point scales. *Journal of Usability Studies*, 5(3), 104-110.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Garland, R. (1991). The midpoint on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70.
- He, J. y Von De Vijver, J. R. (2015). Effects of a general response style on cross-cultural comparisons. Evidence from the teaching and learning international survey. *Public Opinion Quarterly*, 79, 267-290. doi:10.1093/poq/nfv006
- Hernández, A., Espejo, B., Gómez-Benito, J. y González-Romá, V. (2001). Escalas de respuesta tipo Likert: ¿Es relevante la alternativa "indiferente"? *Metodología de Encuestas*, 3(2), 135-150.
- Higgins, J. y Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Recuperado de www.handbook.cochrane.org
- Hopkins, D. (1989). *Investigación en el aula*. Barcelona: Promociones y Publicaciones Universitarias.
- Johns, R. (2005). One size doesn't fit all: selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties*, 15(2), 237-264.
- Kennedy, R., Riquier, C. y Sharp, B. (1996). Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement and Analysis for Marketing* 5(1), 56-70.
- Klopfers, F. y Madden, T. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin*, 6, 97-101.
- Likert, R. (1932). A technique for the measurement of attitude. *Archives of Psychology*, 140, 5-55.
- Lozano, L., García-Cueto, E. y Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79.
- Matus, C. (2007). *Dimensiones de la calidad según OECD y EUROSTAT*. Recuperado de http://www.ine.cl/canales/menu/publicaciones/estudios_y_documentos/estudios/dimensionesdecalidad_4.pdf
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42(6), 779-794.
- Nadler, J., Weston, R. y Voyles, E. (2015). Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, 142(2), 71-89.
- Ooster, T. R. F. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills*, 68, 549-550.

- Pardo, A. (2002). *Análisis de datos categóricos*. Madrid: UNED.
- Paulhus, D. (1991). Measurement and control of response biases. En J. Robinson, P. Shaver y L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-51). EUA: Academic Press.
- Pemberton, E. (1933). A technique for measuring the optimum rating scale for opinion measures. *Sociology and Social Research*, 17, 470-472.
- Peterson, R. y Wilson, R. (1992). Measuring customer satisfaction: fact and artifact. *Journal of the Academy of Marketing Science*, 20(1), 61-71.
- Salkind, N. (1998). *Métodos de Investigación*. México: Prentice Hall.
- Saris, W., Veld, W. y Gallhofer, I. (2004). Development and improvement of questionnaires using predictions of reliability and validity. En S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin y E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (275-297). San Francisco, CA: John Wiley y Sons.
- Scherpenzeel, A. y Saris, W. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research*, 25, 341-383.
- Schuman, H. y Presser, S. (1981). *Questions and answers in attitude surveys*. San Diego, CA: Academic Press.
- Silva, F., Valdivia, B., Iop, R., Gutierrez, P. y Silva, R. (2013). Escalas y listas de evaluación de la calidad de estudios científicos. *Revista Cubana de Información en Ciencias de la Salud*, 24(3). Recuperado de <http://www.acimed.sld.cu/index.php/acimed/article/view/438/318>
- Sudman, S. y Bradburn, N. M. (1989). Measuring attitudes: recording responses. En S. Sudman y N. Bradburn, (Eds.), *Asking questions: a practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass Publishers.
- Tójar, J. C. (2006). *Investigación cualitativa*. Madrid: La Muralla.
- Velez, P. y Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods*, 1(2), 69-74.