



INEVA en acción

<http://ineva.uprrp.edu>

Boletín informativo

Volumen 10, Número 1, Diciembre 2015

LA CANTIDAD DE CATEGORÍAS DE RESPUESTA EN UNA ESCALA DE VALORACIÓN

María del R. Medina Díaz, Ph.D., J.D.

Las escalas de valoración¹ son instrumentos de medición utilizados en distintas investigaciones en los campos de la Educación, la Psicología, el Mercadeo y la Salud para determinar las actitudes, creencias y opiniones de las personas acerca de diversos constructos² de interés. Una escala de valoración consiste de un conjunto de estímulos (aseveraciones o afirmaciones) en los cuales las personas asignan o seleccionan una respuesta determinada. La escala Likert, nombrada de esta manera en honor a su creador Rensis Likert (1932), es un tipo especial de estas escalas. Incluye un conjunto de afirmaciones donde las personas indican su grado de acuerdo (endoso, favor u otra percepción similar) con cada una. En ambas escalas, las respuestas están ordenadas en gradaciones en un continuo unidimensional del constructo que se intenta medir (e.g., actitudes). El formato de las respuestas depende de dos componentes primordiales: (a) el número de las categorías de respuestas, considerando si hay un número par o impar, y (b) el rótulo o nombre de cada categoría. Además, se suele asignar un valor numérico a cada categoría de respuesta. Los siguientes dos ítems son parte de una escala de valoración y escala Likert, respectivamente:

¹ Se utilizan distintos nombres para referirse a estas escalas: escala de categorías, escala de estimación, escala de clasificación y *rating scale*, en inglés.

² Constructo es una construcción mental o abstracción creada para explicar determinados fenómenos y comportamientos humanos o sociales (Medina-Díaz, 2010).

1. ¿Cuán importante es la clase de Estadística?

Muy importante	Importante	Poco importante	Sin importancia
3	2	1	0

2. Prefiero los cursos donde se analizan datos numéricos.

Completamente de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Completamente en desacuerdo
5	4	3	2	1

Cada ítem combina un estímulo (pregunta, afirmación o aseveración) y un formato de respuesta determinado (e.g., categorías con adverbios, adjetivos o numerales). Conviene señalar que a menudo se denomina “una escala” al instrumento compuesto por un conjunto de estos ítems.

La cantidad óptima de las categorías de respuesta ha sido un asunto de gran interés y debate en la construcción de las escalas de valoración y Likert, desde la primera década del siglo XX (Cox, 1980). En particular, las investigaciones han indagado los efectos en la confiabilidad y la validez³ de las puntuaciones, pero no hay unos resultados

³ Los datos obtenidos mediante la administración de instrumentos, tales como cuestionarios, escalas y pruebas, se consideran válidos cuando proveen un estímulo adecuado y preciso del constructo que intentan representar y se apoyan en evidencia de diversas fuentes (i.e., contenido, proceso de respuesta, estructura interna, relación con otras variables y consecuencias) (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014).

concluyentes que ofrezcan una respuesta definitiva a la pregunta: **¿cuántas categorías de respuesta se necesitan en una escala?** Este artículo expone, de manera sucinta, algunos de los estudios que se han realizado acerca de este asunto. Concluye que no hay un número “mágico” de categorías de respuestas que sea apropiada para todos las escalas. Sin embargo, en armonía con el principio de parsimonia en la investigación educativa, los requisitos de validez y confiabilidad de las puntuaciones y el constructo que se intenta medir y las características de las personas que van a contestar, recomiendo al menos cuatro categorías.

Cantidad de categorías de respuesta

Teóricamente, la cantidad mínima de categorías de respuesta que podría tener un ítem en una escala de categorías es dos, pero no hay una máxima para todas las circunstancias. Podrían incluir una cantidad numerosa o infinita de posibles respuestas para que las personas seleccionen. Se presume que los ítems que incluyen numerosas opciones permiten mayor amplitud para escoger la que mejor representa su sentir o respuesta al asunto en cuestión (Bending, 1954). Existe, también, la idea que mientras mayor sea el número de categorías, mejor es el poder de discriminación y predicción de los ítems (Pajares, Hartley & Valiente, 2001). Dillman (2007) reporta que las escalas más usadas poseen entre cinco y siete opciones de respuesta. Conforme a la teoría de procesamiento de información, seis o siete categorías de respuesta serían suficientes (Gardner, 1960).

Tampoco hay un consenso entre algunos de los autores de libros en el campo de la Medición Educativa y Psicológica acerca de la cantidad ideal de categorías en los ítems de una escala de valoración y de una escala Likert. En 1954, Guilford sugirió 25. Hopkins y Antes (1990) recomiendan de cuatro a seis categorías para hacer discriminaciones adecuadas, sin que resulte difícil la selección. Colton y Corvet (2007) recomiendan entre tres y siete. Aiken (1996) indica tres a cinco opciones para escalas con los extremos opuestos. Gallagher (1998), por su parte, las extiende a siete.

Del mismo modo, las investigaciones revisadas presentan distintos hallazgos en cuanto a la cantidad ideal. Cabe señalar que se han utilizado distintos criterios para determinarla. La confiabilidad de las

puntuaciones (en particular, la consistencia interna aplicando el coeficiente Alfa de Cronbach) es el criterio principal para evaluar o comparar las cantidades de categorías en distintos instrumentos (Adelson & McCoach, 2010; Chang, 1994; Chomeya, 2010; Contractor & Fox, 2011; Davies, 2008; Dawes, 2008; Flamer, 1983; Hartely & Betts, 2010; Lozano, García-Cueto & Muñoz, 2008; Pajares, Hartley & Valiente, 2001). Algunas indagan acerca de su efecto en la discriminación de los ítems (Garner, 1960) y en muy pocos, la validez de las puntuaciones (Adelson & McCoach, 2010; Chang, 1994; Flamer, 1983; Jacobi & Matell, 1971).

La Tabla 1 presenta 13 estudios acerca del número de categorías en las escalas y la comparación entre escalas con distintas cantidades. En general, los resultados revelan que no necesariamente una gran cantidad de categorías (e.g., 10) produce las mejores estadísticas asociadas con la discriminación de los ítems y la confiabilidad de las respuestas. Dawes (2008), por ejemplo, encontró que las escalas con cinco y siete opciones producen promedios aritméticos de las puntuaciones más altos que las escalas con 10 alternativas. Asimismo, Contractor y Fox (2011) señalan que las escalas con cinco o seis categorías detectan mejor las diferencias en las puntuaciones de grupos de personas que las que tienen siete, nueve o diez opciones. Freund, Tietjens y Strauss (2013), por su parte, favorecen el uso de cuatro categorías con estudiantes de escuelas secundarias.

Cabe mencionar que existen las diferencias en el grado intensidad de las categorías presentadas en la Tabla 1, según lo ilustran los adverbios (e. g., *strongly agree*, *very important*). En cuanto a la rotulación con numerales, Schwartz, et. al. (1991) encontraron que el uso de números positivos, incluyendo al cero (e.g., 5 a 0) sugiere la ausencia o presencia del constructo que se pretende medir. Por el contrario, los números negativos (e.g., -5 a 0) implican la presencia de un opuesto al constructo.

Con respecto a la escala Likert, Chomeya (2010) encontró que seis categorías, en lugar de cinco, producen mayor discriminación de los ítems, consistencia interna y dificultad en la selección de las respuestas de estudiantes universitarios. Lozano, García-Cueto y Muñoz (2008) sostienen que el número de categorías en este tipo de escala debe

fluctuar entre cuatro y siete, considerando la validez y la confiabilidad de las puntuaciones. Por su parte, Hernández Sampieri, Fernández Collado y Batista Lucio (2010, p.252) señalan que “si las participantes tienen poca capacidad de discriminar se pueden considerar dos o tres categorías. Por el contrario, si son personas con un nivel educativo elevado y de gran capacidad de discriminación, pueden incluirse siete o más categorías”. Adelson y McCoach (2010) administraron una escala de actitud hacia la Matemática con cuatro y cinco categorías de respuesta a estudiantes de tercero a cuarto grado. Encontraron que los estudiantes contestaron los ítems de manera similar. No obstante, las características psicométricas de los ítems con cinco categorías resultaron superiores. También, Van Laerhoven, Van der Zaang-Loonen y Derk (2004) destacan la facilidad y preferencia de un grupo de 120 niños y niñas residentes en Holanda por una escala Likert (*very nice, nice, in between, annoying, and very annoying*) en lugar de una gráfica (visual) o numérica (visual-numérica), para contestar siete preguntas sobre sus actividades diarias, sentimientos y opiniones.

Conclusiones

De este breve examen acerca de la cantidad de categorías en una escala, se pueden derivar las siguientes conclusiones: (a) la cantidad idónea depende del constructo que se intenta medir y de las condiciones particulares de la investigación donde se van a utilizar; (b) la cantidad seleccionada debe permitir las diferencias en las respuestas, la observación y la calificación de las personas y (3) un gran número de categorías no necesariamente significa mayor discriminación en las respuestas y confiabilidad de las puntuaciones, cinco y siete opciones producen las mejores estadísticas. No obstante, resulta parsimonioso construir una escala con al menos cuatro categorías, de modo que no se afecte considerablemente la discriminación y la confiabilidad de las respuestas y se exponga a las personas a una carga cognitiva innecesaria y de tiempo para contestar (Tourangeau, Rips & Rasinski, 2000).

Por otra parte, una cantidad impar de categorías de respuesta requiere una en el centro. Esta se recomienda en las circunstancias en que las personas pueden asumir legítimamente una posición neutral (Garland, 1991; Schuman & Presser, 1981;

Medina-Díaz, 2010). En algunos casos esta opción no es necesaria ya que sólo se requiere ubicar o seleccionar la respuesta en dos extremos o polos (e.g., positivo-negativo, aceptable-inaceptable). Nunnally y Bernstein (1994) consideran que este asunto no es tan crítico y que se deben considerar los ítems y las personas que van a contestar en la decisión. Por ejemplo, Weijters, Cabooter y Schillewaert (2010) encontraron, en las respuestas de 1207 hombres belgas a cuestionario electrónico, una tendencia responder de manera favorable cuando se presentaron opciones intermedias y rotuladas.

En fin, las categorías de respuesta no deben ser tan pocas, que se termine con información limitada de lo que se desea, ni tantas que resulte difícil usarlas todas. Lo que se persigue es una cantidad suficiente que capture la mayor información posible de las personas acerca del constructo de interés, reduciendo los errores en seleccionar las respuestas (e.g., error de respuesta fija y de tendencia central). No se debe olvidar que en toda medición hay errores y que el diseño y la construcción de instrumentos contribuyen a éstos. En la práctica, el uso de más categorías podría consumir más tiempo y provocar dificultad para escoger una y por ende, afectar la validez de las respuestas. Recomiendo, pues, que se realice un análisis de las respuestas a los ítems, en el estudio piloto de un instrumento, para verificar la concentración y variabilidad de las respuestas presentadas. Además, las entrevistas cognitivas y el protocolo de pensar en voz alta con un grupo de las personas que responden ayuda a corroborar la certeza de las alternativas seleccionadas. Conviene advertir que las categorías de respuestas son parte del conjunto de ítems que conforma una escala, así que también merece atención la calidad de las aseveraciones y las instrucciones presentadas.

Referencias

- Adelson, J.L. & 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.
- Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behavior, personality, and*

- attitudes*. New York: John Wiley.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amoo, T. & Friedman, H. H. (2001). Do numeric values influence subjects' response to rating scales? *Journal of International Marketing and Marketing Research*, 26, 41-46.
Recuperado de <http://academic.brooklyn.cuny.edu/economic/friedman/numeric/htm>.
- Bending, A. W. (1954). Reliability and the number of rating scale categories. *Journal of Applied Psychology*, 38, 38-40.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18, 205-215.
- Chomeya, R. (2010). Quality of Psychology test between Likert scale 5 and 6 points. *Journal of Social Sciences*, 6 (3), 399-403.
- Colton, D. & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation*. San Francisco, CA: John Wiley & Sons.
- Contractor, S. H. & Fox, R. J. (2011). An investigation of the relationship between the number of response categories and scale sensitivity. *Journal of Targeting, Measurement and Analysis for Marketing*, 19, 23-35.
- Cox, E.P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, 17, 407-422.
- Davies, R. S. (2008). Designing a response scale to improve average group response reliability. *Evaluation and Research in Education*, 21(2), 134-146.
- 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, 61-77.
- Dillman, D.A. (2007). *Mail and internet surveys: The tailored design method* (2a. ed.). Hoboken, NJ: John Wiley & Sons.
- Flamer, S. (1983). Assessment of the multitrait-multimethod matrix validity of Likert scales via confirmatory factor analysis. *Multivariate Behavioral Research*, 18, 275-308.
- Freund, P.A., Tietjens, M. & Strauss, B. (2013). Using rating scales for the assessment of physical-self-concept: Why the numbers of response categories matters. *Measurement in Physical Education and Exercise Science*, 17, 249-263.
- Garner, W. R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67, 343-352.
- Hartley, J. & Betts, L. R. (2010). Four layouts and a finding: The effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology*, 13, 17-27.
- Hernández Baeza, A. Muñiz, J. & García Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicotema*, 12(2), 288-291.
- Hernández Sampieri, R., Fernández-Collado, C. & Baptista Lucio, P. (2010). *Metodología de la investigación* (5a. ed.). Mexico, D. F.: McGraw-Hill/Interamericana Editores S. A. de C.V.
- Hopkins, C. D. & Antes, R. L. (1990). *Classroom measurement and evaluation* (3a. ed.). Itasca, IL: F. E. Peacock.
- Gallagher, J. D. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ:

- Merrill/Prentice Hall.
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70.
- Guilford, J.P. (1954). *Psychometrics methods* (2a. ed.). New York. McGraw-Hill.
- Jacobi, J. & Matell, M.S. (1971). Three-point Likert scale are good enough. *Journal of Marketing Research*, 8, 495-500.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5-55.
- Lozano, L. M., García-Cueto, E. & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79.
- Medina Díaz, M. (2010). *Construcción de cuestionarios para la investigación educativa*. San Juan, PR: ExPERTS Consultants.
- Muraki, E. (1990). Fitting a polytomous ítem response model to Likert-type data. *Applied Psychological Measurement*, 14, 59-71.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3a. ed.). New York, NY: McGraw-Hill, Inc.
- Pajares, F., Hartley, J. & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33, 214-221.
- Schuman, H. & Presser, S. (1981). Questions and answers in attitude surveys: Experiments on questions form, wording, and context. San Diego, CA: Academic Press.
- Schwarz, N., Knäuper, B., Hippler, H-J., Noelle-Neuman, E. & Clark, L. (1991). Rating numeric values may change the meaning of scale values. *Public Opinion Quarterly*, 55, 570-582.
- Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The Psychology of Survey Response*. New York: Cambridge Press.
- Van Laerhoven, H., Van der Zaang-Loonen, H. J. & Derk, B. H. F. (2004). A comparison of Likert Scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica*, 93, 830-835.
- Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category levels. *International Journal of Research in Marketing*, 27, 236-247.

¿TRIANGULACIÓN VIENE DE TRES?

Nellivette Prieto Vázquez
Careli Sánchez Pabón

Con el propósito de ofrecer una mirada metodológica del concepto triangulación es necesario estudiar sus antecedentes. Se aportará una discusión en términos lingüísticos, así como una definición dirigida al ámbito de la investigación.

El Diccionario de la Real Academia Española (RAE) define triangular como: “ligar por medio de triángulos ciertos puntos determinados de una comarca para levantar el plano de ella”. Cabe destacar que la definición que ofrece el diccionario está enmarcada en el contexto geográfico y en la medición de superficies, y no en los paradigmas de investigación. La triangulación ha sido utilizada en otros ámbitos tales como la navegación y la milicia, por lo que se le ha otorgado varias definiciones: (1) “la división de un área en triángulos para fines de topografía y (2) la medición y la cartografía de un área a través de triángulos con una base, altura y ángulos conocido” (Oppermann & McKercher, 2000, p. 1). Rodríguez-Ruíz (2005) establece que las medidas utilizando en la triangulación deben ser correctas e interrelacionadas (p. 1).

El uso de la triangulación en las ciencias sociales fue originado en el 1959 en la obra de Campbell y Fiske. Esta obra fue dedicada al operacionismo múltiple y su propósito fue la validación de los

resultados de la investigación. Por su parte, en el campo de la educación ha sido utilizada para comparar datos, perspectivas o teorías de otras investigaciones, contextos, instrumentos y métodos (Rodríguez-Sabiote, Pozo-Llorente & Gutiérrez-Pérez, 2006; Hussein, 2009). Su propósito es identificar tendencias detectadas en un determinado grupo de observaciones (Rodríguez-Ruíz, 2005). Morse (1985), citado en Hussein (2009), afirma que existen múltiples usos de la triangulación. Por ejemplo, un maestro podría comparar las puntuaciones de las evaluaciones de sus estudiantes, con la asistencia a clases y los niveles en los que se encuentran según las pruebas estandarizadas del estado. La triangulación de esos datos le permitiría al maestro comparar todas las evidencias, corroborar resultados y llevar a cabo la toma de decisiones del proceso de enseñanza aprendizaje, fundamentada en el análisis de la información.

Tipos de triangulación

La literatura destaca varios tipos de triangulación, sin embargo solo se discutirán los propuestos por Denzin (1978), ya que es uno de los autores que más ha contribuido al tema.

I. **Triangulación de datos** - Se refiere a que los investigadores pueden emplear eficientemente los mismos métodos, pero deliberadamente tomar este método para tantas áreas diferentes como sea posible. La triangulación de datos puede ser realizada a base de los siguientes tres subtemas: tiempo, espacio y personas, pero a su vez el análisis de personas, tiene tres niveles (a) análisis de agregados, (b) análisis interactivo y (c) análisis colectivo.

- a. **Tiempo** - Se aplica la máxima variación en tiempo cuando el investigador recoge datos sobre el mismo objetivo en distintos tiempos por ejemplo: diferentes horas del día o semanalmente, mensualmente o anualmente, para luego ser comparada.
- b. **Espacio** - El espacio también puede ser considerado como una unidad de análisis al realizar un análisis ecológico, por ejemplo observaciones en el hogar, observaciones en el ámbito escolar, entre otras.

c. **Personas** - Esta se subdivide en tres niveles, pero en todas se comparan entre las unidades de estudio.

- i. **análisis de agregados** - En este la unidad de estudio es el individuo. Se seleccionan individuos por sus características individuales, pero estos no tienen relaciones entre sí. Por ejemplo, un investigador podría examinar las actitudes del personal de un hospital hacia la muerte.

- ii. **análisis interactivo** - La unidad de estudio seleccionada son pequeños grupos y se presta atención a las interacciones que ocurren en el mismo. Un investigador podría llevar a cabo un estudio sobre cómo se generan actitudes hacia la muerte a raíz de las conversaciones o encuentros entre el personal de un hospital.

- iii. **análisis colectivo** - La unidad de observación es una organización, grupo o hasta una sociedad completa. Un investigador podría examinar las características de un hospital (su organización) y como dictan ciertas actitudes y prácticas en los miembros de esta entidad.

II. **Triangulación de investigadores** - Esta consiste en que se utilicen múltiples observadores (investigadores) que compartan el mismo rol y sus aportaciones sean equivalentes. Esto ayuda a eliminar los posibles sesgos que pueda tener el investigador como persona.

III. **Triangulación de teorías** - Esta requiere que se examinen múltiples perspectivas o hipótesis. Se debe recoger datos que apoyen y que contradigan las perspectivas o hipótesis estudiadas, con el fin de evaluar su utilidad y poder.

Conclusión

Desde que se introdujo el uso de la triangulación en las ciencias sociales surgieron ideas equivocadas de su significado (Hussein, 2009). Puede que el

fundamento de la mala interpretación del método sea el desconocimiento o la asociación del término con las definiciones otorgadas en otras disciplinas como lo es la medición. Cabe señalar que, la triangulación puede ser utilizada con el propósito de realizar un análisis de la información para aportar a la credibilidad del estudio o hacer referencia a múltiples fuentes de datos en la misma investigación para fines de validación (Hussein, 2009). Sin embargo, en ocasiones puede interpretarse que para emplear este tipo de estrategia de análisis es necesario tener tres componentes. Por ejemplo, el investigador podría deducir que para llevar a cabo la triangulación en la investigación cualitativa se necesitan tres estrategias de recopilación de datos; como sería la observación, la entrevista y el grupo focal. No obstante, dicha interpretación es un concepto erróneo del término, ya que el investigador puede emplear todas las estrategias que desee para llevar a cabo la triangulación. Es decir, que no necesariamente tiene que limitarse a tres tipos de fuentes de datos para analizar la información.

El propósito de la triangulación es la disminución del sesgo en la investigación. Además, sirve para argumentar y fortalecer los hallazgos de una investigación (Oppermann, 2000). Si se compara con el paradigma cuantitativo, podría relacionarse con el término validez, ya que la mayoría de sus beneficios están relacionados con la confirmación. Se caracteriza como una influencia para demostrar la veracidad del contenido (Hussein, 2009). Por ejemplo, un maestro podría utilizar la triangulación para analizar la información que recopila en la sala de clases. Para tomar decisiones fundamentadas en los datos, el docente puede compilar la información a través de la evaluación formativa, *assessment*, asistencia a clases u otros parámetros. Esto fortalecería sus hallazgos y los argumentos para pasar juicio sobre la ejecución del estudiante. La triangulación sirve para desarrollar fases secuenciales de estudios cualitativos y cuantitativos a lo largo del tiempo. Al conseguir variedad de fuentes que revelen la veracidad de la información, el estudio fortalece su credibilidad. Es por eso que la triangulación es muy importante en el proceso de recopilar y corroborar la información encontrada obtenida en un estudio (Lucca & Berríos, 2009).

Referencias

- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work*, (1), 1-12. Recuperado de http://jcs.w.no/local/media/jcs.w/docs/jcs.w_issue_2009_1_8_article.pdf
- Lucca Irizarry, N. & Berríos Rivera, R. (2009). *Investigación cualitativa: Fundamentos, diseños y estrategias* (2a. ed.). Cataño, PR: Ediciones SM.
- Oppermann, M. & McKercher, B. (2000). Triangulation: A methodological discussion. *International Journal of Tourism Research*, 2(2), 141-145.
- Rodríguez-Ruiz, O. R. (2005). La triangulación como estrategia de investigación en ciencias sociales. *Revista Madrid*, 31(2) 1-9.
- Rodríguez-Sabiote, C., Pozo-Llorente, T. & Gutiérrez-Pérez, J. (2006). La triangulación analítica como recurso para la validación de estudios de encuesta recurrentes e investigaciones de réplica en Educación Superior. *RELIEVE*, 12(2), 1-13. Recuperado de http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_6.htm.

INVESTIGACIÓN Y ACCIÓN: DEFINICIÓN Y MODELOS DEL PROCESO

Sarai Torres Ruiz
Rose M. Vincenty Colón, M.Ed.

Los educadores tienen el cometido de descubrir si una estrategia de enseñanza-aprendizaje funciona para el bienestar de sus estudiantes, hecho que puede ser logrado mediante la investigación en acción. Autores como Corey (1949), Creswell (2015), McMillan (2012) y Mills (2000) definen investigación en acción como una investigación sistemática conducida por docentes, directores y consejeros para determinar las consecuencias de una práctica educativa con el fin de mejorarla. En otras palabras, durante el proceso de enseñanza-aprendizaje el docente lograría ver cambios en el razonamiento, conocimiento y comportamiento del alumno (López, 2005). Cabe señalar que Kurt Lewin, 1930, acuñó el término *action research* en el ambiente social. En esa época, su fin era mediar con la escasez de la carne, el reconocimiento aéreo para la Segunda Guerra Mundial, entre otros (Mills, 2000).

Mills (2000) indica que la investigación en acción es una actividad creativa en curso que nos expone a sorpresas en el camino, pero también está compuesta de conductas y eventos que son ordenados y tienen causas detectables. Hoy en día, la investigación en acción ha tomado auge en la educación ya que se utiliza cuando se tiene un problema específico en el salón de clase. Por tanto, permite que el docente reflexione ante sus propias prácticas. Asimismo, al ser una actividad creativa, el docente tiene a su alcance descubrir si la estrategia funciona para el bienestar de sus alumnos y el suyo.

La investigación en acción consiste de una espiral de ciclos descrito como modelo de; ciclos de acción reflexiva, espirales de acción y flujograma (Latorre, 2003). En 1946, Kurt Lewin describe el ciclo de acción reflexiva, el cual consiste de: planificación, acción y evaluación de la acción (Latorre, 2003). Comienza con una "idea general" sobre un tema de interés elaborando un plan de acción. Se reconoce el plan, sus posibilidades y limitaciones y se lleva a cabo el primer paso de acción evaluando el resultado. Luego, se revisa el plan general a la luz de la información y se planifica el segundo paso de

acción sobre la base del primero (Latorre, 2003).

Kemmis, 1989, apoyándose en el modelo de Lewin, elabora un modelo para aplicarlo a la enseñanza (Latorre, 2003). El proceso está integrado por cuatro fases o momentos interrelacionados: planificación, acción, observación y reflexión. Cada uno de estos forman conjuntamente una espiral autoreflexiva de conocimiento y acción (Latorre, 2003). La duración de cada ciclo dependerá del tema de investigación (Yuni & Urbano, 2012).

Al igual que Kemmis, John Elliot toma como punto de partida el modelo cíclico de Lewin. Elliot, 1993, crea un modelo compuesto por tres momentos: (1) elaborar un plan, (2) ponerlo en marcha y (3) evaluarlo; (1) rectificar el plan, (2) ponerlo en marcha y (3) evaluarlo, y así sucesivamente (Latorre, 2003). Whitehead, 1991, propuso un modelo que permite mejorar la relación entre la teoría educativa y autodesarrollo profesional, este espiral de ciclos se compone de: sentir o experimentar un problema, imaginar la solución del problema, poner en práctica la solución imaginada, evaluar los resultados de las acciones emprendidas y modificar la práctica a la luz de los resultados (Latorre, 2003).

Conclusión

La investigación en acción es una estrategia flexible que debe llevarse a cabo como parte natural del proceso de enseñanza-aprendizaje. Ésta se puede considerar una herramienta al alcance del docente, ya que le brinda la oportunidad de evaluar e interpretar la reacción del estudiante ante tareas que se realizan en el salón de clase y de ser necesario hacer modificaciones para el bienestar del ambiente educativo. Es importante para el(la) educador(a) investigador(a) determinar si la investigación en acción es el diseño apropiado en su sala de clase.

Referencias

- Corey, S. M. (1949). Curriculum development through action research. *Educational Leadership*, 1(3), 147-152.
- Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4a. ed.). Boston, MA: Pearson Education, Inc.

- Latorre, B. (2003). *La investigación-acción: Conocer y cambiar la práctica educativa*. Barcelona, España: Editorial Grao.
- López, A. (2005). *Introducción a la educación* (5a. ed.). Hato Rey, Puerto Rico: Publicaciones Puertorriqueñas, Inc.
- McMillan, J. H. (2012). *Educational research: Fundamentals for the consumer* (6a. ed.). Boston, MA: Pearson.
- Mills, G. E. (2000). *Action research: A guide for the teacher researcher*. Upper Saddle River, New Jersey: Prentice Hall.
- Yuni, J. A. & Urbano, C. A. (2012). *Mapas y herramientas para conocer la escuela: Investigación etnográfica e investigación-acción*. Córdoba, Argentina: Editorial Brujas.

Si desea citar alguno de los artículos presentados en este boletín, recomendamos utilizar el formato que especifica el Manual de estilo de publicaciones de la *American Psychological Association* (2010). A continuación se presenta un ejemplo de cómo citar un artículo de un boletín electrónico.

Vázquez, J. P. (2007, marzo). Estudio de Evaluabilidad. *INEVA en acción*, 3(1). Recuperado el 1 de marzo de 2007, de <http://ineva.uprrp.edu/boletin/boletin8.pdf>

Las opiniones vertidas en esta publicación son de los (las) autores(as) y no representan las del Programa de INEVA. Las personas interesadas en escribir para esta publicación deben comunicarse con la Junta Editora a nuestra dirección electrónica (ineva@uprrp.edu). De la misma manera pueden enviarnos sus comentarios y sugerencias acerca de esta publicación.

JUNTA EDITORA

Carlos R. Carrasquillo Ríos
Luz G. Cátala Rivera
Angélica Cortés Ortiz
Elba Ortiz Laureano
Nellivette Prieto Vázquez
Careli Sánchez Pabón
María B. Serrano Abreu
Sarai Torres Ruiz
Rose M. Vincenty Colón
Dra. Claudia X. Alvarez Romero
Dr. Víctor E. Bonilla Rodríguez

Conceptuación gráfica
Víctor E. Bonilla Rodríguez, Ph. D.