

**COMENTARIOS Y RESPUESTA A “UN ESQUEMA CONCEPTUAL
PARA ANALIZAR LA VALIDEZ EN LAS INVESTIGACIONES
MEDIANTE ENCUESTA” DE A. GONZÁLEZ Y J.L. PADILLA**

Juana Gómez Benito

Juana Gómez Benito

Universidad de Barcelona

González y Padilla sugieren una propuesta enriquecedora para analizar de modo más completo la validez de una investigación mediante encuesta y como tal podrá ser tenida en cuenta en los estudios aplicados que utilicen esta estrategia metodológica. Comparto el punto de vista de los autores respecto a la consideración de la validez como punto central de toda investigación y, en concreto, de la investigación mediante encuesta. Es cierto que, en este tipo de investigaciones, la preocupación sobre su validez se centra prioritariamente en la adecuación de la muestra y que indudablemente hay otros pilares indispensables en los que se debe apoyar la calidad de los resultados obtenidos. En este sentido, el tema abordado en el artículo es oportuno y sus aportaciones relevantes.

Antes de examinar aspectos concretos relacionados con el esquema conceptual que proponen González y Padilla para analizar la validez en las investigaciones mediante encuesta, considero necesario señalar que como afirman Kahnemann y Tversky (1973), Reichardt (1985) y Mark (1986) la propuesta de cualquier categorización debe ir unida a la recomendación de no adoptarla dogmáticamente. Con esto quiero decir que ha de advertirse a todo investigador que pretenda aplicar la metodología de encuesta, que las tipologías de validez y de fuentes de error asociadas a las mismas simplifican la realidad y que, por ello, nunca deben sustituir al razonamiento crítico del propio investigador. De hecho, todavía carecemos de una estrategia que nos permita estimar de forma precisa la probabilidad y la magnitud de las fuentes de error que podemos encontrar en nuestra tarea investigadora, ya que dicha estimación depende de los objetivos de la investigación y del contexto en el que ésta se desarrolla.

Dicho esto, cabe plantear algunas sugerencias que, a mi juicio, pueden servir para completar dicha propuesta o para reflexionar sobre ella.

Con respecto a la aportación central del artículo, el esquema conceptual planteado, hay que señalar que su principal limitación reside en que no se pueden adscribir todas las investigaciones mediante encuesta a la metodología cuasi-experimental. En la mayoría de los casos, la investigación mediante encuesta enfatiza la generalizabilidad de la población que se desea representar en detrimento de un menor control interno y el control se ejerce de modo indirecto, generalmente a través de la estandarización de las condiciones de la investigación (véase Kish, 1981, para una discusión entre las prioridades de las distintas estrategias metodológicas). En primer lugar, hay que tener en cuenta que una buena parte de las investigaciones mediante encuesta que se llevan a cabo en estos momentos son estrictamente descriptivas. En segundo lugar, no puede obviarse que un porcentaje importante de las encuestas etiquetadas como analíticas ha de proceder de modo indirecto al realizar sus inferencias a partir esencialmente de la variabilidad existente en la población, basándose en gran parte en la teoría y teniendo en cuenta la revisión de hipótesis alternativas que permitan rechazar explicaciones espúreas; incluso las encuestas analíticas son predominantemente transversales, con lo que las relaciones entre variables han de postularse desde la teoría, al igual que posibles secuencias temporales. En general, el investigador de una encuesta analítica plantea que un conjunto de relaciones observables son consistentes con lo que desea predecir de un cierto modelo y que otros modelos alternativos han demostrado su inconsistencia, desestimando algunas explicaciones espúreas de las relaciones observadas. Por lo tanto, aquí la validez de constructo es básica y prioritaria, y enlaza directamente con el concepto de constructo de Cronbach y Meehl (1955) como determinado a través de su relación con otros constructos teóricos y con variables observables, requiriendo una descripción explicativa de la red nomológica en la que está implicado el constructo y el establecimiento de relaciones entre esta red y una serie de medidas observadas.

Al hilo con este último comentario, y teniendo en cuenta el carácter aplicado de la mayoría de las investigaciones mediante encuesta, sería interesante añadir en el trabajo alguna reflexión en torno a las interrelaciones y prioridades entre los distintos tipos de validez que se plantean en el esquema conceptual para el análisis de la validez en el ámbito de la metodología de encuesta. En la introducción del trabajo se cita a Messick (1995) como abanderado de un concepto único de validez, al tiempo que se critica a otros autores por su consideración compartimentada de los varios tipos de validez. Esta toma de postura no se refleja en el desarrollo del artículo ni en la figura 1 en que, si bien las casillas interrelacionan cada fuente de error con cada tipo de validez, éstas últimas vuelven a aparecer compartimentadas, sin apreciarse planteamientos vinculantes o jerárquicos entre los distintos tipos de validez y, dado el tipo de tabla bidimensional, sin el concepto integrador ni las relaciones anidadas planteadas

por Messick (1981, 1989, 1995) a lo largo de su dilatada trayectoria de teórico de la validez.

En cuanto a algunos comentarios más puntuales, cabe reseñar, en primer lugar, que para tener entidad suficiente como para constituir un tipo de validez en sí misma, la validez de asignación numérica debe definirse de forma menos restrictiva. Parafraseando a González y Padilla:

“Este tipo de validez se relaciona directamente con el proceso de escalamiento diseñado para la obtención de los valores numéricos con los que se *cuantifican* los constructos objeto de medición.(pág. 89)... Es posible que la naturaleza teórica del constructo permita una interpretación *cuantitativa* (i.e. validez de asignación numérica)... (pág. 91)”¹

Esta definición aparenta adolecer de las mismas deficiencias que presentaba la aproximación de Campbell (1928) al problema de la medición, quien planteaba que en la medición debe existir un isomorfismo entre la *cantidad* y las magnitudes de la propiedad a medir. Tal como afirma Muñiz (1998), para ello habría que demostrar que las magnitudes obedecían a los axiomas de *cantidad* desarrollados por Hölder (1901). Al eliminar la restricción de que los números asignados como medidas tuvieran que obedecer necesariamente a las leyes de la cantidad, Stevens (1951) abrió la posibilidad de establecer las hoy clásicas cuatro escalas de medición que se definen mediante cuatro reglas distintas de asignación de los números a los objetos así como las numerosas clasificaciones de escalas que han surgido tras la propuesta de Stevens. Posteriormente, el nuevo enfoque representacional sobre la teoría de la medición que apareció en los años sesenta (véase Schwager, 1991) se caracteriza fundamentalmente por reconocer de forma explícita el papel que juega la teoría en la medición, pasando ésta a formar parte integral de la teoría. En este contexto, la medición se entiende como una modelización de la realidad en la que el sistema relacional formal son los números.

Por otra parte, el planteamiento de la deseabilidad social como una de las fuentes de error del entrevistado que afectan habitualmente a la validez de constructo, podría completarse añadiendo a la misma el sesgo derivado de la “adivinación de hipótesis a partir de claves contextuales”, que se produce cuando los sujetos realizan conjeturas acerca de lo que creen que el experimentador espera encontrar en los resultados de la investigación o acerca del modo en que el experimentador espera que ellos se comporten. Para realizar este tipo de conjeturas utilizan lo que Orne (1962, 1969) denomina “características de la demanda de la situación experimental”, concepto que hace referencia al conjunto de claves contextuales que comunican a los sujetos la hipótesis de investigación y que actúan como factores determinantes de su conducta.

¹ La cursiva es del comentarista

Por último, quisiera valorar muy positivamente la incorporación de la validez ética de la investigación al esquema conceptual para analizar la calidad de la misma ya que, además de salvaguardar los derechos de los distintos agentes objeto de evaluación, optimiza la imagen de los profesionales que realizan investigaciones en este ámbito.

Para finalizar, debo poner de manifiesto que mis consideraciones y sugerencias solo pretenden enriquecer una propuesta que, como he señalado al principio del presente escrito, considero harto interesante y que, sin lugar a dudas, entraña una profunda reflexión sobre un tema fundamental y poco analizado. Por ello, quiero felicitar sinceramente a sus autores.

Respuesta: Andrés González Gómez y José Luis Padilla García

Resulta grato comenzar agradeciendo la atención que la profesora Gómez ha prestado a nuestra propuesta. Sus comentarios han estimulado una relectura del trabajo original que nos ha revelado argumentos e ideas poco claras o expuestas de manera incompleta. Nuestra respuesta sólo pretende apuntar e insistir en algunas ideas sugeridas por sus comentarios.

Primero, el trabajo no pretende proponer una nueva tipología o categorización de la validez. La concepción unitaria de la validez en torno a la validez de constructo es aceptada hoy día por la práctica totalidad de los profesionales de la medición. Insistir en esta aproximación unitaria pretende además evitar análisis reduccionistas sobre la calidad de la información obtenida, en este caso, mediante la encuesta. Ahora bien, parece evidente que el estudio sobre la validez de una interpretación de las mediciones no puede ser *holístico*. La concepción unitaria de la validez no puede dejar al investigador aplicado o usuario de la encuesta ante la magnitud de un proceso de validación global, sin fin, en el que deba obtener evidencias para sustentar todos y cada uno de los argumentos que hay detrás de la interpretación deseada de las mediciones. No parece aventurado apuntar que la necesidad de priorizar o al menos ordenar la búsqueda de evidencias está detrás, por ejemplo, de la propuesta de Messick (1988) sobre las “facetas de la validez” o de las “perspectivas” de Cronbach (1988). Compartiendo, por tanto, la advertencia de la profesora Gómez sobre el uso dogmático de cualquier tipología, la propuesta de un esquema conceptual para analizar la validez en las investigaciones mediante encuesta sólo pretende organizar la búsqueda de evidencias.

El “aroma tipológico”, valga la expresión, de nuestra propuesta procede de estar sustentada en parte sobre la categorización de las amenazas a la validez elaborada por Cook y Campbell (1979). Basta pensar que la propuesta se hubiera “suavizado” si sustituimos la etiqueta “validez interna” por la expresión “análisis de las relaciones causales”. De cualquier modo, la amplia repercusión de la categorización de Cook y

Campbell (1979) garantiza la atención hacia los problemas de validación a los que la encuesta no es ajena.

El núcleo del comentario de la profesora Gómez se centra en la propuesta de adscribir todas las investigaciones mediante encuesta a la metodología cuasi-experimental. Sus dudas sobre la pertinencia de esta adscripción se centran en que la mayoría de las encuestas que se realizan son descriptivas y en que "...las encuestas etiquetadas como analíticas han de *proceder de modo indirecto* al realizar sus inferencias a partir esencialmente de la variabilidad existente en la población" (pág. 100).

Existe una confusión muy extendida que emana de la mezcla de tipologías elaboradas con diferentes criterios; por un lado, taxonomías elaboradas atendiendo al diseño de la investigación: observacional o no, selectiva o no y experimental, cuasi-experimental y no experimental; y por otro, una clasificación por su objetivo o finalidad: investigaciones descriptivas o analíticas (González, 1998). No es conveniente mezclar ambos criterios taxonómicos ya que por ejemplo es posible realizar una investigación analítica bajo la modalidad observacional, selectiva o experimental.

Básicamente, una investigación se puede considerar una encuesta si se asienta sobre un proceso de obtención de muestras. Ahora bien esa misma investigación se podría también considerar como un experimento, un cuasi-experimento o un no-experimento, dependiendo de como cumpliera los criterios de asignación aleatoria y manipulación.

No obstante, para argumentar de manera más apropiada debemos centrarnos en la taxonomía más clara, aquella que organiza de forma exhaustiva y exclusiva las investigaciones en experimentales, cuasi-experimentales y no experimentales. La "aleatorización" y la manipulación de las variables antecedentes o independientes son los criterios definitorios de una investigación experimental (León y Montero, 1993). ¿Puede realizarse una investigación experimental mediante una encuesta? Aunque poco frecuente la respuesta debe ser afirmativa. Pensemos por ejemplo en una investigación para conocer los efectos de dos sistemas de retribución sobre el rendimiento de los funcionarios de la administración. Se podría seleccionar de la población de funcionarios una muestra al azar para participar en el estudio y formar con ellos los grupos de comparación. La decisión sobre qué grupos recibirán los diferentes sistemas de retribución completa el cuadro de una investigación experimental mediante encuesta. Sirva el reciente ejemplo de Durán, Lacaci, Baena y Manzano (1999) en cuyo trabajo realizaron un *experimento mediante encuestas* con el siguiente esquema: 1) se pretende comprobar en qué medida diferentes estrategias de información previa a la población incrementan la tasa de respuestas; 2) se consideran dos variables independientes con dos niveles cada una de ellas: utilización de cartel (si-no) y utilización de carta personalizada (si-no), lo que genera cuatro condiciones experimentales; 3) se obtiene una muestra de viviendas mediante el procedimiento de las rutas aleatorias; 4) se asignan aleatoriamente las viviendas seleccionadas a las cuatro condiciones experi-

mentales. Ésta es la descripción ortodoxa de una investigación experimental que, además, se basa en el diseño y aplicación de una encuesta.

La escasez de investigaciones experimentales mediante encuestas hace que prefiramos su inclusión dentro de la metodología cuasi-experimental. Los ejemplos de encuestas cuasi-experimentales son más numerosos ya que se incluirían bajo esta etiqueta todas las investigaciones dirigidas a establecer relaciones en las que no sea posible la aleatorización. El esquema conceptual propuesto en el trabajo se basa en un concepto más amplio y laxo del término “manipulación”, un concepto que incluya la selección que hace el investigador de los niveles de las variables de los que recoge información y que contemple los *procedimientos indirectos* a los que la profesora Gómez alude al referirse a las encuestas analíticas. Por ejemplo, un investigador interesado en conocer los cambios en los hábitos de consumo debidos a la edad, “manipula” la edad al seleccionar sus niveles, con los que agrupa a los integrantes de la muestra.

Coincidimos en el papel básico y prioritario de la validez de constructo para las encuestas analíticas desde la perspectiva de contraste de modelos, pero no pensamos que cuestione nuestra propuesta ya que lo mismo se podría decir de toda investigación: ni siquiera las encuestas descriptivas se libran de la interpretación que hace el investigador de la información a partir de un modelo explícito o implícito, más o menos elaborado. Por ejemplo, en una encuesta sobre el uso de métodos anticonceptivos, el investigador se limitará a dar el porcentaje de parejas que usan el preservativo o interpretará que dicho porcentaje es “bajo”, “normal”, “insuficiente”, “reflejo de la influencia de la religión”, etc.

Por otra parte, no puede ser sino debido a la brevedad con la que en el trabajo original presentamos la denominada “validez de asignación numérica”, la causa de que la profesora Gómez nos atribuye una definición restrictiva de la medición. Restricción en la línea de Campbell (1928) al imponer un isomorfismo entre la *cantidad* y la *magnitud* de la propiedad a medir y superada por los trabajos de Stevens y el papel asignado a la teoría en el proceso de medición por el hoy día dominante enfoque representacional. No es ese el propósito de este apartado del trabajo original.

Nuestra intención al incluir la validez de la asignación numérica era llamar la atención sobre la necesidad de preocuparse por las propiedades de los números con los que se cuantifican, o si se quiere “representan”, las variables medidas en la encuesta. La mayoría de las mediciones realizadas con las encuestas pertenecen a lo que ya Torgerson (1958) denominó “medidas de índice”, aquellas en las que la medida depende de un cuerpo teórico que justifica la relación entre el constructo y los indicadores del mismo elegidos de forma arbitraria por el investigador. Un ejemplo típico de medida de índice es la que se realiza para medir el constructo “clase social”. El investigador puede medir la clase social a través de indicadores como el nivel de renta, el lugar de residencia, el tipo de trabajo, etc., pero siempre hay algo de arbitrariedad y su uso sólo está justificado desde una perspectiva operacionalista y por

razones de utilidad práctica. La referencia a la asignación numérica pretende que el usuario de la encuesta cuestione si es posible, por ejemplo en este caso, "ordenar" o hablar de "distancias" entre clases sociales a partir de mediciones por indicadores del constructo "clase social".

Se puede argumentar con razón que el enfoque representacional reconoce de forma explícita el papel de la teoría en la medición considerándola parte integrante de la misma, pero puede olvidarse que el enfoque representacional es también axiomático en el sentido que establece las condiciones que debe cumplir un sistema empírico (propiedades objeto de la medición) para ser representado por un sistema numérico (Jañez, 1989). No parece exagerado pedir que el responsable de la encuesta reflexione al menos sobre el tipo de relaciones que se pueden representar entre los valores de la variable que pretende medir. No necesariamente todas las mediciones se deben realizar por indicadores. Para la medición de variables actitudinales o constructos específicos se dispone de esquemas conceptuales que pueden ayudar en esta tarea; por ejemplo, la teoría de la medida conjunta (Luce y Tuckey, 1964) resulta útil para medir variables a un nivel de intervalos a partir de relaciones con otras variables medidas a un nivel ordinal.

Terminamos como al principio, agradeciendo los comentarios y sugerencias de la profesora Gómez, cuyos trabajos despiertan el interés de cualquier persona dedicada a la metodología de las ciencias sociales, y a quien los autores no pueden sino agradecer el haber sido objeto de su atención.

Referencias

- Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. Londres: Longmans Green.
- Cook, T. D. y Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Cronbach, L. J. (1988). Five perspectives on validity argument. En H. Wainer (Ed.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. y Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Durán, A.; Lacaci, P.; Baena, S. y Manzano, V. (1999) Estrategias para aumentar la tasa de respuestas en las encuestas a domicilio. Comunicación presentada en el VI Congreso de Metodología de las Ciencias Sociales y de la Salud. Oviedo: Universidad de Oviedo.
- González, A. (1998). *Apuntes de metodología*. Documento no publicado. Universidad de Granada.
- Hölder, O. (1901). Die axiome de quantität die lehre von mass. *Berichte ueber die Verhandlungen der Königlich Sächsischen Gessellschaft der Wissenschaften zu Leipzig, Mathematisch-Psysische Class*, 53, 1-64.

- Jañez, L. (1989). *Fundamentos de Psicología Matemática*. Madrid: Pirámide.
- Kahnemann, D. y Tversky, A. (1973). On the Psychology on prediction. *Psychology Review*, 80, 237-251.
- Kish, L. (1981). Representación, aleatorización y control. En F. Alvira, M. D. Avia, R. Calvo y J.F. Morales (Eds.), *Los dos métodos de las ciencias sociales*. (2 ed.). Madrid: CIS.
- León, O. G. y Montero, I. (1993). *Diseño de investigaciones*. Madrid: McGraw-Hill.
- Luce, R. D. y Tuckey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.
- Mark, M. M. (1986). Validity typologies and the logic and practice of quasi-experimentation. En W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 47-66). San Francisco: Jossey-Bass.
- Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, 89, 575-588.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. En H. Wainer (Ed.), *Test validity* (pp. 33 - 45). Hillsdale, NJ: Erlbaum.
- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational Measurement*. New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50, 741-749.
- Muñiz (1998). La medición de lo psicológico. *Psicothema*, 10, 1-21.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. En R. Rosenthal y R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143-179). New York: Academic Press.
- Reichardt, C. S. (1985). Reinterpreting Seaver's study of teacher expectancies as a regression artifact. *Journal of Educational Psychology*, 77, 231-236.
- Schwager, K. W. (1991). The representational theory of measurement: an assessment. *Psychological Bulletin*, 110 (3), 618-626.
- Stevens, S. S. (1951). Mathematics, measurement and psychophysics. En S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley.