



METODOLOGÍA DE INVESTIGACIÓN EN EDUCACIÓN MÉDICA

El diseño de investigación en educación: conceptos actuales

Educational research design: current concepts

Lucy María Reidl Martínez

Coordinadora del Consejo Académico del Área de las Ciencias Sociales. Profesora Titular C Definitiva, División de Estudios de Posgrado, Facultad de Psicología, Universidad Nacional Autónoma de México.

Recepción 3 de octubre 2011; aceptación 31 de octubre 2011

El presente manuscrito ofrece un panorama global de varios conceptos relevantes para la investigación en educación aplicables a las ciencias de la salud, destacando las características (tradicionales y novedosas) de las diferentes aproximaciones en la búsqueda del conocimiento.

Paradigmas de la investigación

El paradigma tradicional de investigación científica se refiere a las relaciones causales que se dan en un sistema cerrado. El sistema se ha aislado porque se controla la influencia de las variables extrañas, reduciendo la posibilidad de dar explicaciones alternativas al evaluar las hipótesis planteadas por el investigador.¹

El análisis causal en la investigación científica se localiza en un extremo de un continuo, y en el contrario se encuentra el análisis causal de los sistemas abiertos. Los sistemas abiertos son aquellos donde no se puede controlar el efecto de todas las variables extrañas (diseños correlacionales, no experimentales y de observación pasiva), y donde pueden existir alternativas diversas a la relación causal establecida entre las variables de interés.

Se plantea la posibilidad de utilizar un paradigma unificado de las relaciones causales en la investigación científica, que incluya la especificación de las condiciones de

los sistemas abiertos, para proporcionar un ajuste aceptable a la forma en que ocurren los eventos en el ambiente natural. La manera de hacerlo es por medio de métodos que permitan estudiar el comportamiento en el hábitat natural y cultural de los organismos.

La investigación debe asegurar tres tipos de validez: la interna, la externa y la ecológica. La primera es la que infiere si la relación entre dos variables es causal o no; la segunda se refiere a la generalización de las conclusiones experimentales más allá de las personas, escenarios y tiempos del estudio experimental, y la tercera hace referencia al grado en el que los hallazgos de investigación permiten predecir el fenómeno en el mundo real.²⁻⁵

Frente a una realidad que es cada vez más compleja y cambiante, la investigación forzosamente lleva a realizar un abordaje multidisciplinario, frente a casi cualquier tipo de problema, entre ellos, el de la educación. Obliga a conocer, reflexionar y decidir si se deben adoptar tradiciones de investigación modernas, para conocer una realidad mucho más amplia; para ello se proponen como alternativa de abordaje y desarrollo de la ciencia, los modelos mixtos de investigación.^{6,7}

Los paradigmas son visiones del mundo o sistemas de creencias que guían a la investigación; se reconoce la existencia de tres tipos de ellos: a) la aproximación

Correspondencia: Edificio de los Consejos Académicos. Ciudad Universitaria. Coyoacán, CP 04510. México DF. Teléfonos: 5622 1565 y 5622 1535. Fax: 5622 1502. Correo electrónico: lucym@unam.mx

positivista/empirista/cuantitativa, b) la aproximación constructivista/fenomenológica/cualitativa, y c) la aproximación pragmática, que utiliza una metodología mixta, cualitativa y cuantitativa.^{6,8}

Para el pragmatismo, la pregunta de investigación es más importante que el método a utilizar o el sistema de creencias (paradigma) que subyace a dicho método. El método es secundario a la pregunta de investigación y la decisión con respecto a usar un método cualitativo o cuantitativo (o ambos) depende de la pregunta de investigación, de su planteamiento y de la fase del ciclo de investigación en la que se encuentra. Surge de la necesidad de hacer investigación que comprenda y mejore la condición humana, que comunique resultados que favorezcan una toma de decisiones más informada. Reconoce que el mundo es complejo, estratificado y difícil de entender, y que es por ello que se requiere combinar las metodologías de investigación.

El tema de nuestro interés es la educación, que ha sido descrita de diversas maneras que llevan a verla como un proceso multidireccional mediante el cual se transmiten conocimientos, valores, costumbres y comportamientos; ésta incluye la vinculación y la concienciación cultural, moral y conductual. Se asimilan y aprenden conocimientos, normas de conducta, modos de ser y formas de ver el mundo. Se trata de un proceso de socialización formal de los individuos para desarrollar capacidades físicas e intelectuales, habilidades, destrezas, técnicas de estudio y formas de comportamiento ordenadas de acuerdo a un fin social (valores, trabajo en equipo, regulación, etc.). Esto es lo que se tiene que investigar: qué, cómo, cuándo y dónde hacer para formar profesionistas, y el porqué de cada una de estas estrategias y no otras, en función de lo que favorece o pone en riesgo el proceso educativo.

Medición y método de investigación

La función del método de investigación es guiar el proceso de investigación: cómo hacer la pregunta de investigación (en virtud de lo que se busca determinar si hay diferencias entre grupos o sujetos), si existen relaciones entre variables (dos o más), si esta relación es concurrente o predictiva. El procedimiento inicia con la búsqueda bibliográfica en revistas especializadas, nacionales e internacionales. Se debe establecer la perspectiva que se tomará: interdisciplinar o multidisciplinar.

La principal fuente de invalidez de cualquier investigación es el desconocimiento de la forma en que se miden y evalúan las variables de interés y los modelos de investigación existentes. Ello lleva a conocer las teorías de construcción de instrumentos y las características que deben reunir: confiabilidad y validez. Y por otro lado, a conocer los diversos diseños de investigación cuantitativos, cualitativos y modelos mixtos. Aunado a ello es preciso saber también de análisis estadísticos adecuados y pertinentes en cada caso.

La medición puede darse en diferentes niveles: nominal, ordinal, intervalar y de razón. El nivel nominal clasifica, usando categorías dicotómicas (como las de hombre-mujer) o politómicas (como azul-verde-rojo-amarillo). El ordinal, ordena de menos a más, como cuando se hace una fila de personas en función de la estatura de las

mismas. El intervalar cuenta con un cero arbitrario como en los casos de los intervalos subjetiva o aparentemente iguales, como la escala centígrada de temperatura, donde el 100 corresponde a la temperatura en que hierve el agua y el 0 a aquella en donde se congela (al nivel del mar). La medición de razón es aquella en la que existe un cero real, como por ejemplo la escala de grados Kelvin.

Medir se define como la asignación de números de acuerdo a ciertas reglas. Estas reglas son los modelos de medición, y entre los más usados se encuentran el modelo de la medición del error, el de dominio-muestra y el de pruebas paralelas. El de la medición del error parte de dos premisas fundamentales: cuando se mide algo, siempre se sobrestima o se subestima al azar, en diferentes cantidades; si esto sucede un número infinitamente grande de veces, la suma de esas mediciones es cero. Y es por ello que se requiere de un número relativamente grande de preguntas, reactivos, aseveraciones o afirmaciones.

Las premisas del modelo dominio-muestra señalan que existe un número infinitamente grande de posibles reactivos que miden una característica o variable «X», y que si se obtuviera una muestra aleatoria de este universo de reactivos, ésta sería una buena representación del mismo, y es lo que constituye una prueba, cuestionario, escala o cualquier instrumento que mide el atributo o variable de interés. El modelo de pruebas paralelas es una extensión del anterior, en el sentido de pensar que se pueden tomar dos muestras al azar del universo de aseveraciones, y las pruebas por lo tanto son iguales y miden lo mismo, de donde proviene el concepto de confiabilidad de formas paralelas.

Otros modelos de medición derivan de supuestos como son la ley del juicio comparativo y la del juicio categórico, desarrolladas por Thurstone.^{9,10} De estas leyes surgen los procedimientos de pares comparados (un modelo discriminativo) y los de los intervalos aparentemente iguales (estimaciones subjetivas). En el primer caso se emplean jueces "expertos", reales o creados (definiendo el constructo para que ellos sepan de qué se trata) y se sigue un conjunto de pasos hasta poder encontrar la distancia a la que se encuentra cada afirmación de las demás. Así se puede constituir un continuo de estímulos que representan "X" cantidades de la variable a medir. En el segundo caso, se obtienen, a partir de una muestra relativamente grande de sujetos (aproximadamente 300), los valores escalares de las aseveraciones a lo largo de un continuo, que califica a la aseveración desde extremadamente favorable o positiva hasta extremadamente desfavorable o negativa pasando por el centro (neutral) en un continuo de 11 intervalos. La escala final queda establecida por aquellas afirmaciones, aseveraciones o proposiciones, que se localizan a lo largo de todo el continuo psicológico, con los valores intercuartiles (Q) más pequeños asegurando su localización más específica dentro del continuo.

Una segunda aproximación al desarrollo de instrumentos de medición se refiere al escalamiento de personas, que parte del modelo de rangos sumariados desarrollado por Likert,¹¹ en el que se escriben afirmaciones que representan diversas cantidades del atributo que se desea medir, con cuatro a 10 opciones de respuesta, y se aplica a jueces semejantes a los que serán evaluados con el instrumento que se pretende desarrollar; deberán ser entre

cinco y 10 sujetos por reactivo que se desarrolle. Esta información recabada permite establecer cuáles de las aseveraciones o afirmaciones tienen poder discriminativo (prueba t entre el grupo alto 25% superior, y el bajo 25% inferior), que correlacionen 0.40 o más con la calificación total y no tengan valores de sesgo o kurtosis mayores a 0.50. Las afirmaciones que cumplen estas características son las que forman la escala final.

Por otro lado, la medición puede ser monodimensional o multidimensional; la primera está representada por el escalamiento de estímulos y de personas ya vistos. La versión multidimensional corresponde a constructos heterogéneos o multidimensionales (por ejemplo, la personalidad = Σ de rasgos). Para ellos se utiliza una estrategia estadística específica que es el análisis factorial, siendo el modelo más utilizado el de componentes principales con rotación ortogonal (varimax). Éste es sólo uno de los varios modelos existentes.

Características de los instrumentos

Los instrumentos deben tener ciertas cualidades, las principales son el que sean válidos y confiables. La validez se refiere a que el instrumento mida lo que se pretende medir. Ésta puede ser de diferentes tipos:

- a. Aparente. Se determina por la opinión subjetiva de personas, y no tiene fundamento científico.
- b. De contenido. Está determinada por la lectura de las afirmaciones que la constituyen y el acuerdo de un grupo de jueces expertos en el tema, utilizando jueces como en el caso de pares comparados o juicios graduados como en el de intervalos sucesivos.
- c. Concurrente. Se obtiene estableciendo la correlación de nuestro instrumento con un criterio externo (otra prueba que mide lo mismo que lo que se pretende medir), o de grupos contrastados en los que suponemos que uno de ellos tiene esa característica y el otro no y por ello se espera que haya diferencias entre ambos (por ejemplo si se pretende medir liberalismo deberá de haber diferencias en los puntajes obtenidos por un grupo de liberales en comparación con los obtenidos por un grupo de conservadores).
- d. La validez predictiva es la que se obtiene por medio de una regresión simple o múltiple que permite predecir un comportamiento o situación por una o más variables independientes.
- e. La validez de construcción se establece correlacionando el instrumento con criterios o variables externas de acuerdo a la teoría, o estableciendo la validez factorial que deberá reflejar una estructura factorial acorde a la teoría de la que se parte explicando por lo menos entre el 60% y 70% de la varianza.⁴

La confiabilidad puede ser de tres tipos:

- a. Estabilidad temporal. Se establece por medio de la correlación entre dos aplicaciones a los mismos sujetos en tiempos diferentes.
- b. Equivalencia de formas. Se establece por la correlación entre dos formas paralelas (que miden lo mismo).

- c. Consistencia interna. Se puede establecer de tres maneras: 1) calculando la correlación entre la primera y la segunda mitad de un instrumento constituido por reactivos de igual dificultad, 2) estableciendo la correlación entre los reactivos pares y nones cuando estos son de dificultad creciente, o deben ser respondidos contra reloj, 3) estableciendo la correlación de cada reactivo con la calificación total (para reactivos dicotómicos, con el coeficiente de Kuder-Richardson, y para reactivos politómicos por medio del alfa de Cronbach).⁴

Diseños de investigación

Los diseños de investigación son el plan, la estructura y estrategias que se utilizarán para obtener respuestas a las preguntas de investigación e hipótesis controlando la varianza experimental, extraña y de error. Los diseños implican partir de un marco de referencia (teoría), señalar cómo se obtendrán los datos (serán medidos, observados o se consultarán registros existentes). El diseño también señala cuántos y cuáles registros u observaciones se realizarán, cómo se analizará la información obtenida (de manera cualitativa o cuantitativa) así como el tipo de estadística, de ser el caso, que se utilizará para responder la pregunta de investigación.

Los diseños cuantitativos controlan la varianza experimental por medio de la distribución aleatoria de los sujetos a las condiciones experimentales. A las variables extrañas se les maneja de diversas maneras: a) eliminando la variable aunque se pierde un poco el poder de generalización, b) introduciéndola como variable independiente, c) apareando a los sujetos en esa variable, o d) utilizando algún método de control estadístico. Estos diseños también controlan el error de medición y muestreo. El primero se debe a situaciones no controladas, la utilización de registros incompletos, a que los evaluadores tienen diferentes criterios o que los instrumentos no son confiables ni válidos. El segundo se debe a muestras sesgadas o atípicas, accidentales, no aleatorias, o pequeñas.

Entre los diseños cuantitativos más empleados se encuentran: el de una muestra, dos muestras independientes, dos muestras relacionadas (igualadas, o antes-después), k (más de dos) muestras independientes, k muestras relacionadas (igualadas, o medidas en tres momentos diferentes). Además existen diseños que permiten establecer el efecto combinado de dos o más variables independientes sobre una dependiente: a) diseños factoriales, dos por dos (dos variables independientes con dos valores cada una), b) tres por dos (sexo: masculino y femenino, tres grupos de edad), dos variables independientes, una con tres y otra con dos valores, y c) otros más, como uno de tres por tres, dos por tres por dos (sexo: masculino y femenino, tres grupos de edad, y rural - urbano). Este tipo de diseños podrían, en teoría, contener un mayor número de variables independientes; sin embargo, en la medida que incrementan éstas constituyéndose en un mayor número de grupos de investigación, incrementa el riesgo de tener celdillas vacías, sin sujetos, debido a que la combinación de las diferentes variables puede ser inexistente o de difícil acceso, invalidándose de esta manera la

posibilidad de establecer los efectos combinados de ellas sobre la variable dependiente.⁴

Entre los diseños mixtos existentes se pueden observar cuatro categorías: triangulación, Explicativo, Integrador y Exploratorio.⁶ Algunos ejemplos de los diseños mixtos de triangulación son: triangulación, convergencia y validez cuantitativa. Entre los ejemplos de los diseños de tipo explicativo se encuentran el explicativo continuo con énfasis cuantitativo, y el de selección participante, con énfasis cualitativo. En la categoría de integrador se encuentran, entre otros, el integrador, el integrador experimental, y el integrador correlacional. En el tipo exploratorio se tienen entre otros, el exploratorio y el de desarrollo de instrumentos con énfasis cuantitativo. Se describe muy someramente un ejemplo de cada uno de los diseños señalados.

En el *diseño de validez cuantitativa* de la categoría Triangulación, se recogen datos cuantitativos por medio de encuestas, y cualitativos por medio de preguntas abiertas; se analizan ambas clases de datos, se obtienen resultados cuantitativos y cualitativos, se validan entre ellos y se ofrece una interpretación cuantitativa-cualitativa.

El *diseño integrador experimental* de la categoría Integrador, inicia con un abordaje cualitativo antes de la intervención, se obtiene una medición cuantitativa previa a la intervención, volviéndose a medir después de ella y durante todo el tiempo que dure; después de la intervención también se llevan a cabo registros cualitativos que permiten establecer una interpretación final de los resultados basada en información cualitativa y cuantitativa.

Dentro de la categoría Explicativo está el modelo mixto *explicativo continuo con énfasis cuantitativo*, que consiste en la siguiente secuencia de actividades: se recogen datos cuantitativos, se analizan de manera cuantitativa, se presentan resultados cuantitativos que permiten identificar a cuáles se les dará seguimiento; durante el seguimiento, se recogen datos cualitativos, se analizan desde la perspectiva cualitativa, se presentan los resultados cualitativos, y finalmente se da una interpretación cualitativa-cuantitativa de los datos.

En la categoría Exploratorio se encuentra el *diseño desarrollo de instrumentos con énfasis cuantitativo*, que tiene la siguiente secuencia: se recogen datos cualitativos, se analizan cualitativamente, se llega a los resultados cualitativos y se desarrolla el instrumento, se recogen datos cuantitativos, se analizan cuantitativamente, se presentan los resultados cuantitativos, y se hace una interpretación final cuantitativa-cualitativa.

Análisis estadístico

Por último, cabe señalar que también se debe conocer la relación que existe entre los niveles de medición de las variables investigadas, los tipos de diseños de investigación utilizados, la naturaleza de las preguntas que uno desea responder, y los procedimientos estadísticos adecuados para cada caso. El desconocimiento de cualquiera de estos aspectos invalida las conclusiones a las que se llegue. A continuación se mencionan algunos ejemplos.

Cuando el nivel de medición es nominal, y los datos provienen de una sola muestra, se recomiendan la prueba de *ji cuadrada* y la prueba binomial; si son dos muestras

independientes, la prueba de Fisher o la *ji cuadrada*; si son dos muestras relacionadas (igualadas en alguna variable de interés, o una muestra medida en dos ocasiones), la prueba de McNemar; si son *k* (más de dos) muestras independientes, *ji cuadrada*, y si estuvieran relacionadas (igualadas en alguna variable o medidas en más de dos ocasiones), la *Q* de Cochran. Para establecer la relación entre las variables de interés, se deberá usar el coeficiente de contingencia.

Si los instrumentos que miden las variables de interés lo hacen a nivel ordinal y los datos provienen de una sola muestra, se deberán utilizar las pruebas de Kolmogorov-Smirnov o la prueba de rachas; si se tienen dos muestras independientes, las pruebas de la mediana, de Mann-Whitney, Kolmogorov-Smirnov, Wald-Wolfowitz, o la de Moses; cuando son dos muestras relacionadas, las pruebas de los signos o la de Wilcoxon; para el caso de *k* muestras independientes, las pruebas de la extensión de la mediana o la de Kruskal-Wallis; para *k* muestras relacionadas, la de Friedman. Si se desea establecer la relación entre las variables, los coeficientes de correlación de Spearman, Kendall o el coeficiente de concordancia.

Para variables medidas a nivel intervalar, en el caso de dos muestras independientes, la prueba de aleatorización o la *t* de Student; para dos muestras relacionadas, la de Walsh o la *t* de Student para muestras relacionadas. Para *k* muestras independientes, el análisis de varianza, y para las relacionadas el análisis de varianza para medidas repetidas. Si se desea establecer la relación entre las variables, se puede usar la correlación producto-momento de Pearson.¹²

Por último, debe reconocerse que la mayor parte de los temas relacionados con la educación (evaluación de conocimientos, predicción del rendimiento escolar, factores de riesgo y protección relacionados con el aprendizaje, evaluación de competencias, enseñanza basada en competencias, evaluación del desempeño docente, razonamiento clínico, estrategias de enseñanza y aprendizaje, percepción de autoeficacia académica, estrategias de estudio, etc.), hacen referencia a variables complejas, y la manera de abordarlas es utilizando análisis estadísticos multivariados. Entre ellos destacan la regresión múltiple (una variable dependiente, varias independientes), la regresión logística (variable dependiente dicotómica o polinómica a nivel nominal), comparación de perfiles de variables mediante el análisis de correspondencia (variables nominales) o análisis discriminante (variables intervalares), escalamiento multidimensional, reducción de variables nominales por medio del análisis de conglomerados, y de variables intervalares y de razón por medio del análisis factorial en sus diversas modalidades. La tecnología y desarrollo de la estadística permite que en la actualidad se pongan a prueba modelos complejos por medio del modelamiento de ecuaciones estructurales.¹³

Conclusión

De manera muy breve y concisa se ha tratado de dar una visión panorámica de lo que constituye el proceso de investigación científica, que tiene como propósito responder las preguntas de interés dentro de un campo de estudio dado. La investigación científica es una y la misma

para casi todos los campos del conocimiento; los objetivos, los objetos o sujetos, y los campos de conocimiento de los que se parte son los diferentes. La diversidad de las posibles preguntas que se pretende responder obedece a las necesidades que planteadas a los investigadores en esos diversos campos, entre los cuales se encuentra la educación.

En nuestro país, el problema de la educación se ha visto exacerbado, en virtud de los bajos puntajes obtenidos en los diferentes "rankings" a los que se someten las instituciones educativas a nivel nacional e internacional, y al incremento de la población en edad escolar, entre otros. Ello conlleva a la necesidad de investigar sobre el tema, asegurando hacerlo de manera adecuada.

La complejidad del concepto educación, las diversas visiones de la misma y las nuevas estrategias de enseñanza (basada en competencias, el uso de las tecnologías de la información y la comunicación, etc.), obligan a realizar investigación que permita responder a esta necesidad. Para ello, se deben conocer los diferentes paradigmas de investigación y sus nuevas perspectivas, los diferentes modelos de medición de los que se parte para la recolección de datos, sus características (validez y confiabilidad), los diferentes tipos de diseños de investigación a los que se puede acceder, y las técnicas de análisis de datos existentes, para ayudar a dar una mejor respuesta a los esfuerzos que realizan las instituciones educativas.

Referencias

1. Lynd-Stevenson RM. Concerns regarding the traditional paradigm for causal research: The unified paradigm and causal research in scientific psychology. *Rev Gen Psychol* 2007;11:286-304.
2. Bronfenbrenner U. Toward an experimental ecology of human development. *Am Psychol* 1977;513-531.
3. Cook TD, Campbell DT. *Quasi experimentation: Design and analytical issues for field settings*. Chicago. Rand McNally. 1979.
4. Kerlinger FN, Lee HB. *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. México. McGraw-Hill/Interamericana Editores. 2002.
5. Mook DG. In defense of external invalidity. *Am Psychol* 1983;38:379-387.
6. Creswell JW, Plano Clark VL. *Designing and conducting mixed methods research*. Thousand Oaks, CA. Sage Publications. 2007.
7. Greene J. *Mixed methods in social inquiry*. San Francisco, CA. Wiley and Sons Inc. 2007.
8. Tashakkori A, Teddlie C. *Mixed methods in social and behavioral science*. Los Angeles, CA. Sage. 2010.
9. Thurstone LL. A law of comparative judgment. *Psychol Rev* 1927;34:273-286.
10. Thurstone LL. Psychophysical analysis. *Am J Psychol* 1927;38:368-389.
11. Likert RA. A technique for the measurement of attitudes. *Arch Psychol* 1932;140:1-55.
12. Siegel S. *Nonparametric statistics for the behavioral sciences*. New York. McGraw-Hill Book Company, Inc. 1956.
13. Hair JF, Anderson RE, Tatham RL, Black WC. *Análisis multivariante*. Madrid. Prentice Hall. 1999.