

Metodología de investigación cuantitativa-cualitativa: explorando el continuum interactivo

Isadora Newman & Carolyn Benz
Southern Illinois University Press
1998

Capítulo 3: Validez y legitimidad de la investigación (pp. 27-53)

Los resultados de una investigación no tienen valor si los métodos de los cuales se derivan no tienen legitimidad. Los métodos deben justificar nuestra confianza. Quienes leen y confían en resultados de investigación deben estar satisfechos con la validez de los estudios.

En este capítulo se presenta primero un modelo para mostrar cómo conectamos las preguntas de investigación y los métodos con el valor de verdad de los resultados. A continuación, discutimos problemas de validez en la investigación cualitativa, un tema mucho más amorfo. En contraste con los criterios clásicos de validez de la investigación cuantitativa descritos por Campbell y Stanley (1963), los criterios de validez (legitimación) de la investigación cualitativa no son objeto de consenso. Describiremos brevemente las ideas de Martin Hammersley, Margaret LeCompte, Judith Goetz, Steinar Kvale, Patti Lather and William Tierney de manera que podamos localizar nuestras propias ideas, que son más cercanas a la perspectiva postpositivista de Lincoln y Guba (1985). Por último, hacemos una lista de 13 criterios para determinar el valor de verdad que más adelante se convirtieron en nuestros propios criterios para diseñar y criticar estudios cualitativos.

Al final de este capítulo, el lector debe ser capaz de:

1. Definir validez y legitimación.
2. Definir qué es un diseño de investigación, tanto en el paradigma cualitativo como en el cuantitativo.
3. Diferenciar entre validez del diseño y validez de la medición.
4. Definir la validez interna y externa.
5. Describir las amenazas a la validez interna
6. Describir las amenazas a la validez externa
7. Describir la investigación ex post facto.
8. Comparar y contrastar las conceptualizaciones propuestas por varios metodólogos de la investigación cualitativa.
9. Describir los 13 métodos que se añaden a la legitimación de la investigación cualitativa.

La noción de *validez* tiene una larga historia y un fuerte consenso entre los investigadores más tradicionales en educación. El concepto es aplicado a los diseños de investigación en al menos dos contextos (validez interna y validez externa) y a la medición (la validez de la medición). Nos hemos esforzado por conceptualizar la validez en esos dos contextos y en concebir un nuestro propio modelo. Finalmente localizamos la validez de la medición dentro

de la categoría de Campbell y Stanley “instrumentación” (p. 1963, p. 5). Cuando hay un estimativo fuerte de la validez del diseño, del contenido y del constructo en nuestra medición, por ejemplo, las amenazas a la validez interna disminuyen.

Harry Wolcott ha definido los mismos dos conceptos de validez, pero lo hace de manera cronológica. En su ensayo en el libro de 1990 de Eisner y Peshinks, *Indagación cualitativa en educación*, Wolcott afirma estar poco convencido de que la validez necesite tanto énfasis en estudios naturalistas (1990a). Este autor sugiere que la palabra *comprensión* reemplace la palabra *validez* en investigación cualitativa. Lo que es interesante es que él elabora una crónica de la evolución del énfasis en la validez a lo largo de las últimas tres décadas como un patrón de desarrollo:

Validez del test.

Validez de los datos del test.

Validez del test y datos de medición.

Validez de los datos de investigación en tests y mediciones.

Validez de los datos de investigación.

Validez de la investigación.

El autor parece indicar que la validez de la investigación como preocupación creció a partir de una preocupación inicial por la validez de la medición.

Nuestra preocupación ha sido la validez: el valor de verdad de los resultados de investigación es más fuerte cuando tanto los datos como el diseño son válidos. Hemos utilizado deliberadamente las palabras *validez* y *legitimación* en el título de este capítulo. La validez ha significado tradicionalmente un estimativo de la medida en la cual los datos miden (o el diseño mide) lo que se intenta medir. Pero no todos los investigadores conceptualizan los vínculos entre diseño y valor de verdad de esta manera. La *validez*, como la hemos definido en nuestra propia experiencia, no encaja con las definiciones de todos los investigadores. La *legitimación* es un término reciente que hemos tomado prestado aquí para relacionarlo con una noción más amplia de valor de verdad. En esencia, puede considerarse algún paralelismo con nuestra noción de validez.

Para Denzin (1994), por ejemplo, el valor de verdad se convierte en casi circular. Sin validez, nos dice, no hay verdad y sin verdad no hay declaración de validez. Este autor ha descrito el período actual en la investigación cualitativa como una “crisis de legitimación” (p. 295). La legitimación significa que los métodos de investigación son consistentes con los presupuestos filosóficos de la cuestión. Por ejemplo, un positivista asume que existe una realidad objetiva; los posmodernos asumen que no hay realidad objetiva ni verdad objetiva (ver Apéndice A). hasta cierto punto, la noción de legitimación refleja el set de preguntas acerca de la consistencia que aparece en el capítulo 5.

Aunque un retrato perfectamente preciso de nuestras nociones sobre la validez a lo largo del continuum no es posible, podemos delinear las principales dimensiones de nuestra propuesta. Tanto como es posible en un diagrama, la figura 3 enseña nuestro razonamiento acerca de la conexión entre preguntas, métodos y valor de verdad.

Conectar las preguntas de investigación, los métodos y el valor de verdad

La figura 3 no debería ser interpretada en el sentido en que las decisiones contenidas en ella son tomadas necesariamente de esta manera lineal y jerárquica en cada punto del proceso. Por otro lado, la pregunta de investigación *siempre* inicia cualquier conjunto de decisiones que el investigador ha de tomar. Dentro de las dimensiones que siguen a la pregunta, sin embargo, las decisiones deben ser cíclicas. (Métodos cuantitativos y cualitativos pueden cruzarse en situaciones particulares, de ahí la línea punteada que los conecta en la figura 3). Lo que no es negociable es la conexión lineal general entre pregunta, métodos y valor de verdad. Aunque es necesario aceptar que las decisiones en el diseño son “emergentes” en la buena investigación cualitativa, el pensamiento del investigador no es emergente. En él, el vínculo conceptual entre pregunta-métodos-valor de verdad debe ser establecido, defendible y predeterminado.

Como se mencionó antes, la pregunta de investigación guía qué métodos seleccionar. Por esa razón, la figura comienza en la parte de arriba con la casilla “Pregunta de Investigación”, que conduce al siguiente constructo “Métodos”. Como nuestro empeño general es presentar la investigación holísticamente, no como una dicotomía, hemos querido discutir ambos paradigmas de formas conceptuales semejantes, tanto como sea posible. Después de luchar con las nociones tradicionales de validez, concluimos que, para legitimar el valor de verdad de nuestros resultados de investigación, los “Métodos” crean la dimensión que atraviesa ambos paradigmas. Los métodos son aquellos elementos que crean el contenido para las decisiones del investigador, de ahí la etiqueta. La casilla “Métodos” conduce a los dos paradigmas, el “Cualitativo” y el “Cuantitativo”.

Suscribimos la conceptualización existente sobre los problemas tradicionales para la validez en la investigación cuantitativa. Estas preocupaciones acerca de la validez incluyen tanto la validez externa como interna, por un lado, y la validez de la medición, por otro. Ambas categorías son generadas por la necesidad de tener confianza en que nuestro test, datos o diseño de hecho miden o reflejan o producen lo que se supone que midan, reflejen o produzcan.

Nuestra posición sobre la preocupación alrededor de la validez en investigación cualitativa nos coloca más cerca de lo que Denzin (1994) describe como la *perspectiva post-positivista*, esto es, se necesita que haya un conjunto separado de criterios para evaluar la investigación cualitativa. Los post-positivistas hacen un llamado en favor de un conjunto de reglas o procedimientos para establecer la validez. Estos criterios, desde Lincoln y Guba (1985) originalmente (y como nosotras lo definimos) sirven para un propósito similar al de los criterios desarrollados por Campbell y Stanley (1963) para la investigación cuantitativa. En la figura 3 presentamos algunos de esos criterios que apoyan metodológicamente la

legitimación de la investigación cualitativa. Hay 14 métodos, comenzando con “compromiso prolongado”. Todos serán discutidos en mayor detalle más adelante.

El concepto de validez y el lenguaje utilizado para discutir la validez no son bien establecidos o consensuales dentro de la literatura sobre investigación educativa. Algunos investigadores parecen asumir una dicotomía entre la preocupación de los investigadores cualitativos por la validez y la de los investigadores cuantitativos por la confiabilidad. Hammersley (1992) plantea que la confiabilidad es un tipo de verdad (validez). Este argumento está basado en diferentes filosofías más que en diferentes metodologías. Nuestra posición es que la validez no puede existir sin confiabilidad. Por definición, la validez estima la medida en la cual el test o el conjunto de datos o el diseño realmente mide o refleja o produce lo que se supone que mida, refleje y produzca. El propósito básico de la confiabilidad es ayudar a los investigadores a estimar la validez como un estimativo de la medida de error. Por lo tanto, si uno tiene validez, no hay necesidad de estimativos de confiabilidad.

Polkinghorne (1991) se pregunta si los hallazgos de un estudio son creíbles o no. Define entonces la validez como la correspondencia entre los hallazgos y la “realidad”. Otros, como Lincoln y Guba (1985), prefieren el término *credibilidad* al término *validez*.

Nuestro énfasis en explorar la validez de los métodos de investigación a lo largo del continuum cualitativo-cuantitativo está dado por la necesidad de fortalecer el valor de verdad de los resultados de la investigación que realizamos. En la figura 3 mostramos el razonamiento que conecta los métodos con el valor de verdad de los resultados.

Cada lista de “Métodos” conduce a un punto etiquetada como “Diseño”. Todos los métodos juntos crean el diseño.

Poner a andar el diseño en cada paradigma nos lleva al siguiente punto en la figura, “Resultados de Investigación”, que son comunes a ambos paradigmas. De los resultados, finalmente, sale el “Valor de Verdad”, nuestro último punto en la figura. La etiqueta *valor de verdad*, aunque originada en la metodología de investigación cualitativa, es sinónimo de la validez de los resultados de investigación. en alguna medida esto es lo Polkinghorne se está preguntando: ¿son los resultados del estudio creíbles o no?

Problemas de validez en investigación cuantitativa y sus relaciones con los métodos.

Hay una diferencia importante entre la validez de la medición y la validez del diseño. La primera estima qué tan bien logran los instrumentos medir lo que se proponen medir. La segunda, la validez del diseño, abarca la validez interna y la validez externa. La medición de la validez cae dentro de lo que Campbell y Stanley llaman “Instrumentación” (1963).

La validez interna es la medida en la cual cualquier diferencia causal en la variable dependiente puede ser atribuida a la variable independiente. La validez externa es la medida en la cual los resultados de la investigación pueden ser generalizados a otros diseños o a otros grupos.

La validez de la medición y la validez del diseño no son independientes. Un diseño de investigación sólo es internamente válido si tiene validez en la medición y confiabilidad. Esto también es verdad para la capacidad de generalización: es necesario estimar la estabilidad de las mediciones para poder estimar la validez externa. La validez de la medición es un subconjunto de la validez interna. Como se ha discutido en esta sección, una fuerte validez en la medición disminuye las amenazas a la validez interna incluidas en la categoría “instrumentación”.

Escogimos separar la validez de la medición de la validez del diseño para clarificar la discusión. A continuación, vamos a discutir la validez interna y la validez externa en relación con la investigación cuantitativa; después la discutiremos en relación con la investigación cualitativa.

Validez interna y externa

Necesitamos estándares para evaluar la investigación en por lo menos tres dimensiones: la probabilidad de que las variables estén causalmente relacionadas; la confianza que uno puede tener en ser capaz de generalizar los resultados; y el nivel de consistencia entre el propósito de la investigación, los presupuestos, los métodos y las conclusiones.

La validez del diseño para investigaciones cuantitativas tradicionalmente ha sido abordada a través de los conceptos de validez interna y validez externa. Nosotras, junto a Goetz y LeCompte (1984), argumentamos que una preocupación similar debería ser desarrollada por aquellos estudios que son predominantemente cualitativos al final del continuum. Dentro del contexto de nuestra visión holística, sin embargo, subsumimos todas estas preguntas por la consistencia dentro del concepto de validez de los *métodos*. Como mostramos en la figura 3, el investigador toma decisiones sobre los métodos, controlando las amenazas a la validez interna y externa, y los métodos se convierten en diseño de investigación.

La validez interna es definida como “la medida en la cual uno puede decir que la variable independiente causa los efectos de las variables dependientes. En otras palabras, se tiene la capacidad para asumir la causalidad en la medida en que el investigador tiene control” (Newman & Newman, 1994). Le Compte and Goetz (1982) plantean esta pregunta para llegar a la validez interna: “¿Los investigadores científicos realmente observan o miden lo que creen que están observando y midiendo?” (p. 43). En la medida en que ellos están observando y midiendo lo que creen que están observando y midiendo, hay validez.

La segunda área conceptual es la validez externa, definida como “la medida en la cual el estudio es generalizable a otras personas, grupos, investigaciones, etc.” (Newman & Newman, 1994, p. 119). LeCompte y Goetz (1982) formulan esta pregunta para evaluar la validez externa: “¿en qué medida los constructos abstractos y los postulados generados, refinados o puestos a prueba por los investigadores son aplicables a otros grupos?” (p. 43).

Keppel (1973) describe la gran cantidad de esfuerzo hecho por los metodólogos cuantitativos para alcanzar la validez interna. Estos investigadores emplean controles (es decir, tratan de mantener constante la mayor cantidad posible de factores que pueden influenciar el fenómeno

objeto de estudio) para medir más eficientemente la influencia del tratamiento. Se desarrollan pruebas para las mismas condiciones ambientales, utilizando los mismos instrumentos, por parte de los mismos investigadores. A los factores que no son controlables se les permite variar aleatoriamente a lo largo de las condiciones de tratamiento, lo cual opera como otro mecanismo de control.

La aleatorización de lo que Keppel llama “variables fastidiosas” es una forma importante de obtener validez interna, esto es, “la eliminación de los sesgos que, de estar presentes, pueden invalidar cualquier conclusión derivada de las manipulaciones hechas en el experimento” (p. 173). Esta aleatorización remueve el efecto del sesgo y la variabilidad debidos a esos factores.

Otros intentos por alcanzar la validez interna incluyen compensar a lo largo del tratamiento aplicado. Esto se llama *formas equivalentes* (Newman & Newman, 1994) e incluye tener dos tratamientos, por ejemplo, y darle al grupo 1 los tratamientos en el orden A-B y al grupo dos en el orden B-A. en este caso, el orden de los estímulos es utilizado como variable independiente, y su influencia en la variabilidad de la variable dependiente puede ser aislada y medida. Medir este ayuda al investigador a obtener validez interna porque su efecto puede ser separado de los efectos del tratamiento.

Sin validez interna, uno sólo puede concluir que el abordaje utilizado para medir la pregunta de interés es capaz de estimar la relación, pero ningún enunciado sobre la causalidad sería posible. (Más adelante en este capítulo discutiremos esto en los métodos cualitativos). Aun cuando haya entre los rangos de investigadores cualitativos quien diga que no está interesado en la validez interna, quienes desean inferir relaciones causales deben ocuparse de este aspecto de la investigación. de hecho, aun aquellos que descartan esta preocupación porque la consideran un dilema exclusivo del investigador cuantitativo, aceptarán procesos como la triangulación y el muestreo teórico, que con tentativas conceptuales y técnicas para obtener la validez interna.

La validez externa refleja la medida en la cual el diseño y los datos coinciden con el mundo. Aunque el control estrecho sobre las variables de interés aumenta la validez externa, tiende a hacerlo a expensas de la validez externa, esto es, cuanto más parecidas sean las condiciones a las del laboratorio, más precisas y válidas son las medidas. Sin embargo, el mundo no es tan controlado y las variables operan en el mundo exterior al laboratorio. El muestro de sujetos a través de diferentes estratos que reflejan el mundo al cual serán generalizados los resultados (por ejemplo, edad, estatus socioeconómico, ocupación) aumenta la validez externa. La inhabilidad para controlar las variables independientes en la investigación ex post facto (aunque tiende a disminuir la validez interna) tiende a incrementar la validez externa porque tales variables son más relevantes para la distribución en el mundo real (Newman & Newman).

Los diagramas utilizados para ilustrar los diseños de investigación ayudan al investigador a planear, interpretar y analizar. Los símbolos, definidos aquí, son usados para mostrar los componentes de la estrategia de la investigación y son aplicados en el capítulo 5. Aunque los

siguientes son ejemplos de símbolos comúnmente usados, ellos no abarcan todas las posibilidades, ni son utilizados universalmente.

X Tratamiento o tratamiento experimental; se refiere al tratamiento experimental para un grupo.

-X Tratamiento o ausencia de tratamiento (un grupo puede recibir tratamiento, el segundo grupo no; este segundo grupo es generalmente llamado *grupo de control*).

O Medición u observación, frecuentemente algún tipo de puntaje en un test, puede tener cualquier número de subíndices.

R Asignación aleatoria [*random*] de sujetos a los grupos.

M Asignación aleatoria [*random*] de sujetos a los grupos utilizando combinaciones.

Mr Asignación aleatoria [*random*] de sujetos a los grupos utilizando combinaciones primero y luego distribución aleatoria.

Según el trabajo clásico de Campbell y Stanley (1963) la validez del diseño es limitada cuando una variedad de factores permanece sin control.

La historia es un factor que puede contar para el cambio en un grupo. *Historia* significa cualquier cosa ajena al tratamiento, un evento externo que interviene en medio de las mediciones pre y post test. Por ejemplo, un agrupamiento de escuelas estudia el efecto de nuevo texto de matemáticas en el desempeño escolar midiendo los resultados al comienzo y al inicio del año escolar. Durante ese año, sin embargo, los estudiantes deben ser desplazados a un nuevo edificio. Este es un ejemplo de un factor de la historia que puede afectar los resultados de la investigación. El investigador, en este caso, no podría saber si los nuevos libros o la nueva escuela causaron cualquier diferencia en los resultados de logro escolar medidos por el estudio.

Otro factor que puede contar para el cambio en un grupo es la *maduración*, que significa cualquier crecimiento o desarrollo que normalmente tiene lugar con independencia de cualquier tratamiento experimental. Un ejemplo podría ser el análisis del impacto de un programa de alimentación escolar en el crecimiento de los estudiantes. Normalmente ocurriría un crecimiento continuo como parte de la maduración de los estudiantes. En este caso, la tarea del investigador es identificar qué cambio se debe a la alimentación escolar y qué cambio es producto del proceso de crecimiento.

Historia y maduración se confunden con frecuencia, porque se relacionan con algo que ocurre en el tiempo entre el pre y el post-test. El punto que permite la diferenciación es que el efecto histórico es externo, en cuanto la maduración es interna. Otros ejemplos de cambios internos podrían ser psicológicos, como el aburrimiento, o psicológicas, como la fatiga.

Los factores asociados con los dispositivos de medición, factores de la aplicación de las pruebas, también pueden causar la ocurrencia de cambios. Esto puede pasar con un pre-test

sensibiliza a las personas respecto a un tratamiento experimental y causa que de hecho se comporten de forma diferente durante el tratamiento. Supongamos que un profesor está evaluando el impacto de un programa en el desarrollo moral de los estudiantes. Si el profesor aplica un pre-test que plantea preguntas sobre la moralidad, el pre-test puede dejar a los estudiantes preocupados respecto al tema y por lo tanto influenciarlos, haciendo que sean más receptivos al programa que lo podrían ser sin haber tomado el pre-test. Este mismo fenómeno puede tener lugar con otros factores, como el logro académico.

Por otro lado, el término *instrumentación* se refiere a los efectos que se deben a instrumentos de medición poco confiables. Si se tiene un pre-test poco confiable, cualquier cambio que notemos en las mediciones post-test pueden ser debidas a la inestabilidad del dispositivo de medición y no al tratamiento. La instrumentación es una amenaza para la validez interna y está relacionada con la validez y la confiabilidad de los datos. Los datos que el investigador utiliza como evidencia se recogen a través de varios instrumentos que miden las variables de interés. La siguiente sección describe los problemas de validez y confiabilidad de dichas mediciones.

Validez y confiabilidad de las mediciones: instrumentación

Dos preocupaciones tiene el investigador cuando recoge datos a través de instrumentos de medición: la validez y la confiabilidad del instrumento. Las debilidades en cualquiera de las dos amenazan la validez interna y externa. En esta sección discutiremos la validez de la medición y haremos una discusión sobre la confiabilidad.

Un test o instrumento de medición tiene lo que se conoce como *validez aparente* en la medida en que se presenta a los individuos participantes en el estudio midiendo lo que se propone medir. Esto generalmente se considera una estimativa deficiente de la validez. Puede ser importante porque, si un test no tiene *validez aparente* (credibilidad), puede disminuir la probabilidad de que la gente participe o se postulen como voluntarios.

Cuando los expertos en cada área de investigación elaboran juicios subjetivos sobre la validez del instrumento, se dice que hubo validación por partes, que también ha sido llamada *validez lógica* y, algunas veces, *validez definicional*. Es similar a la validez aparente, pero es generalmente estimado utilizando una tabla de especificaciones en la cual el contenido que se espera medir con el test es evaluado a partir de qué tan bien o qué tan suficientemente están representados los ítems a estudiar.

La *validez simultánea* es consiste en determinar qué tan bien un instrumento se relaciona con un instrumento ya establecido y validado. La *validez por grupo conocido* es un tipo de validez simultánea. Se estima a partir de qué tan bien logra el instrumento diferenciar entre dos grupos conocidos. Si se espera que el instrumento mida el éxito en el matrimonio, el instrumento debería ser capaz de diferenciar entre dos grupos de personas que han sido previamente identificadas como exitosamente casadas o no. En la medida en que el instrumento logre hacer eso, tendrá validez por grupo conocido.

La estimación de qué tan bien un instrumento predice un futuro resultado de investigación es conocida como *validez predictiva*. La principal diferencia entre la validez predictiva y la validez simultánea es que esta última tiende a ocurrir dentro del período actual, mientras que la validez predictiva es orientada hacia el futuro. Algunas veces ambas se combinan y entonces hablamos de *validez estadística, empírica o de criterio*.

Lo más importante y al mismo tiempo lo más difícil de estimar es la *validez del constructo*. Se usa para determinar qué tan bien mide el instrumento un constructo subyacente que intenta medir, y es generalmente estimada utilizando una combinación de los otros tipos de validez ya mencionadas. La técnica estadística llamada *análisis factorial* también es utilizada para estimar la validación del constructo.

La investigación cuantitativa que carece de documentación sobre la validez de la medición tiene un valor limitado; los investigadores necesitan recordar que esa validez, cuando se afirma, es siempre un estimativo, nunca una medición actual.

Si la validez es confirmada, implícitamente se cuenta también con la confiabilidad; sin embargo, es posible tener confiabilidad sin validez. El encuadramiento básico que utilizamos es que el propósito fundamental de la confiabilidad es, o bien apoyar, o bien mejorar la validez. La confiabilidad refleja consistencia. Mientras que la validez estima qué tan bien un estudio o un conjunto de instrumentos logra medir lo que se propone medir, la confiabilidad estima si los resultados permanecerán estables a lo largo del tiempo (o sea, si serán “repetibles”) o si son consistentes entre varios observadores independientes (o sea, si diferentes observadores reportarán el mismo resultado).

En investigación cuantitativa, la confiabilidad en la recolección de los datos se puede asegurar de tres maneras: midiendo la consistencia interna, aplicando coeficientes de correlación test-retest y usando formas equivalentes del instrumento. Si la confiabilidad no es asegurada, entonces el presupuesto científico de exactitud es violado. Los hechos no son repetibles.

Así como la confiabilidad es estimada calculando la consistencia interna del test, una medición similar puede derivarse de una *agenda de entrevista estructurada*. Es posible controlar el tiempo, el ambiente y el orden de las preguntas cuando no se puede ejercer dicho control con cuestionarios. En la medida en que estos controles fortalecen la validez, satisfacen los requerimientos de confiabilidad. Para entrevistas no estructuradas no son posibles estimativos de confiabilidad.

No se puede tener validez sin confiabilidad y, concomitantemente, en la medida en que se tiene validez no es necesario estimar la confiabilidad. Por esa razón el esfuerzo predominante en este libro tiene que ver con la validez.

Cuando los grupos son seleccionados con base en puntajes extremos, y sólo por esa razón, ocurre un fenómeno llamado *regresión estadística*. Si los sujetos son seleccionados para un estudio únicamente por su puntaje es extremadamente bajo, en el pos-test ellos van a tender a tener puntajes más altos, sin importar el tratamiento. Lo opuesto también es verdad. Ambos

son ejemplos de casos extremos de regresión hacia la media de la población. Por lo tanto, pueden ocurrir diferencias significativas entre pre-test y post-test porque la elección inicial fue hecha con base en extremos.

La pérdida de sujetos entre tests se llama *mortalidad experimental*. Si una muestra de 100 sujetos tiene un puntaje promedio de 95 en un pre-test de coeficiente intelectual (IQ) y, por alguna razón, 50 sujetos desertan antes del post-test, entonces el IQ promedio de los restantes 50 sería 150. Por lo tanto, las diferencias entre pre-test y post-test podrían, muy posiblemente, no ser debidas al tratamiento sino a la pérdida de sujetos entre pre y post test (o sea, la mortalidad).

Cuando los sujetos son asignados a dos o más grupos de comparación y no todos los grupos reciben tratamiento, esto se llama *sesgo de selección*. Si estos grupos son diferentes antes del tratamiento, entonces cualquier diferente entre los puntajes pre y post-test pueden ser debidas a las diferencias iniciales más que al tratamiento. Un ejemplo podría ser asignar instrucción personalizada (tratamiento) a niños altamente motivados e instrucción tradicional a niños con baja motivación. Si los grupos fueran examinados en busca de ganancias al final de la unidad, cualquier diferencia encontrada puede, de hecho, deberse a las diferencias en la motivación inicial en lugar de a las diferencias en el tratamiento.

Estos factores contaminantes pueden ser considerados al evaluar la investigación. son criterios útiles para juzgar la calidad de la investigación.

Investigación ex post facto

Hay dos tipos de variables independientes: la *activa* y el *atributo*. Las variables activas están bajo el control del investigador y pueden por lo tanto ser manipuladas. Las variables de atributo, como el género y la raza, no pueden ser manipuladas. Si todas las variables independientes fueran no manipulables, entonces la investigación se define como *ex post facto*. La investigación ex post facto es a veces relegada a una posición inferior entre los otros tipos de diseños de investigación. Los términos *investigación ex post facto* e *investigación correlacional* son a veces utilizados de forma intercambiable.

Borg y Gall (1989) se refieren a esta condición como “comparativo causal” (p. 53). En investigación correlacional, la causalidad no puede ser inferida. Muchos metodólogos nos advierten sobre las posibles malas interpretaciones de las investigaciones en las cuales el investigador no tiene control sobre las variables independientes. Algunos consideran que la investigación ex post facto es exploratoria (Newman & Newman, 1994).

En la investigación ex post facto, la causalidad a veces se infiere de manera inapropiada por parte de aquellos que asumen que una variable es probablemente la causa de otra porque la precede o porque la variable tiende a estar altamente correlacionada con aquella (por ejemplo, fumar como variable independiente asumida como causa del cáncer, la variable dependiente).

Mientras la precedencia y la correlación son necesarias, no son suficientes para inferir relaciones causales. Para asumir una relación causal, debemos tener validez interna; esto es,

todas las otras explicaciones para el efecto (variable dependiente) deben ser controladas y la única posible explicación para los cambios en la variable dependiente es que sean debidos a la variable independiente bajo investigación.

Sólo con un diseño verdaderamente experimental se tiene control experimental suficiente para alcanzar esa validez interna. La investigación ex post facto carece de ese control por una variedad de razones. No es posible asignar aleatoriamente y manipular la variable independiente porque ya ha tenido lugar y no está bajo el control del investigador. Otra debilidad común es que el diseño no es capaz de controlar los efectos de confusión derivados de la auto-selección. Por ejemplo, supongamos que se lleva a cabo una investigación para ver qué efecto tiene la formación temprana en la infancia sobre la motivación. Supongamos también que se encuentra una relación significativa entre la temprana formación para la independencia y la posterior motivación. Uno podría, entonces, concluir incorrectamente que la formación para la independencia causa esta motivación adulta. Otra explicación podría ser que los sujetos voluntarios que tuvieron esa formación son más proclives a demostrar un mayor grado de motivación adulta. Lo que causa esta motivación podría estar más relacionado con lo que causa que los sujetos se postulen voluntariamente para el estudio que con la formación para ser independientes.

Si la pregunta tiene que ver con causalidad, claramente la investigación ex post facto no es apropiada. Sin embargo, si la pregunta tiene que ver con relaciones, es apropiada. Algunas veces una pregunta de investigación tiene variables independientes que no pueden ser manipuladas. En esos casos el investigador puede o bien decidir hacer investigación ex post facto o decidir no hacer investigación en absoluto.

Uno de los usos más efectivos de la investigación ex post facto es ayudar a identificar un pequeño conjunto de variables dentro de un conjunto mayor, relacionadas con las variables dependientes, para futura manipulación experimental.

Hemos discutido amenazas a la validez interna. Nuestra discusión, derivada del clásico de Campbell y Stanley (1963) ahora se mueve hacia la validez externa.

La validez externa, de acuerdo con Campbell y Stanley (1963), plantea preocupaciones sobre la posibilidad de generalización. Hay cuatro amenazas: primero, los efectos reactivos de procesos de test en los cuales el pre-test pueda causar que los participantes en el estudio sean más o menos sensibles al tratamiento. Segundo, los efectos reactivos de la situación misma de la investigación. En estos dos casos, la muestra de participantes y sus experiencias se hacen poco probable la generalización en otra población. Tercero, la posibilidad de generalizar puede disminuir cuando hay interacción entre sesgos en la selección de los participantes y el tratamiento. Finalmente, la interferencia de múltiples tratamientos puede también limitar la validez externa. Que los participantes en la investigación sean sujetos a más de un tratamiento puede resultar en que tratamientos previos tenga un efecto en tratamientos posteriores. Los resultados de estos últimos no podrán ser generalizados a grupos que no han experimentado los tratamientos previos.

Nuestra discusión sobre la validez del diseño en investigación cuantitativa ha enfatizado en dos preocupaciones de los investigadores que conducen estudios experimentales en el paradigma cuantitativo: la medida en que cualquier diferencia detectada en un estudio pueda estar relacionada con la variable independiente (validez interna) y la medida en que los resultados del estudio sean representativos de la población, o sean generalizables (validez externa).

Antes de avanzar de estos problemas en el paradigma cuantitativo a problemas similares en el cualitativo, mencionemos brevemente la investigación multivariada. Esta vuelta sobre la investigación cualitativa es apropiada porque, en lugar de investigar una variable a la vez, la investigación multivariada está basada en las complejidades de la investigación científica humana y social. Casi siempre opera más de una variable en las preguntas de investigación que formulamos. En el caso de una investigación cualitativa, la complejidad humana, en algún nivel, subyace a la pregunta de investigación que conduce a métodos cualitativos.

Análisis univariados y multivariados.

Los investigadores cuantitativos utilizan variables dependientes unitarias (diseños univariados) o múltiples variables dependientes (diseños multivariados). Intentar representar las complejidades de los fenómenos sociales conduce a preguntas de investigación con muchas variables.

Los diseños factoriales son una manera de lograr la validez externa. Diseños factoriales de mayor orden (de dos vías, de tres vías, de cuatro vías, etc.) se aproximan de forma más cercana al mundo exterior, que diseños de un factor.

En esencia, cada variable independiente adicional incrementa la relevancia del diseño para la realidad, en la medida en que dichas variables den cuenta de alguna variación única en la variable dependiente. Por ejemplo, los efectos en el logro en matemáticas no son sólo explicados por el IQ sino, lo que es más increíble, por el IQ, antecedentes familiares, educación de los padres y estatus socioeconómico. Sin embargo, con cada variable adicional, el número de casos en cada celda tiende a decrecer. Como el número de variables aumenta, debería considerarse la necesidad de aumentar el tamaño de la muestra. Por lo tanto, algunos investigadores utilizan análisis multivariado como opuesto al análisis univariado para aumentar la validez externa. En otras palabras, los fenómenos comportamentales casi siempre tienen un impacto sobre más de una variable dependiente; y medir una al tiempo que se da cuenta de la varianza de las otras resulta más seguramente en conclusiones que reflejan la relevancia del mundo exterior.

De acuerdo con Michael Patton (1980, 1987, 1990), la metodología cualitativa está basada en el presupuesto de que el estudio del comportamiento humano debe ser diferente del estudio de los fenómenos no humanos. Strike (1972, citado por Patton, 1980, p. 44) quien afirma que, como los humanos tenemos “propósitos y emociones”, hacemos planes, construimos culturas y mantenemos ciertos valores y propósitos. En resumen, un ser humano vive en un

mundo que tiene significado; y, dado que el comportamiento de una persona tiene significado, ese significado puede ser descubierto y explicado.

Patton infiere acertadamente que el comportamiento humano es mucho más complejo que el comportamiento no humano. Aunque Patton argumenta en favor del paradigma cualitativo, esta visión no es muy diferente de la postura asumida por los investigadores cuantitativos. Estos últimos indican que, en una tentativa por tratar de entender hasta el más simple de los comportamientos humanos, deben examinarse muchas variables y las situaciones en que ellas ocurren. En otras palabras, las variables pueden tener efectos diferenciales, dependiendo de las especificidades de la situación. Apuntan a diseñar estrategias que controlen los múltiples efectos en mediciones independientes.

Siguiendo a Patton, uno puede someter a prueba los efectos interactivos de la emoción y el propósito en un cierto comportamiento. Más aún, “la habilidad percibida”, la “probabilidad de fracaso” y “los estadios de necesidad” pueden ser medidos, y sus relaciones con el comportamiento también. Mientras Patton sugiere que la ciencia puede relacionar sólo preguntas medios-efectos, uno puede, de hecho, usar diseños científicos avanzados que consideren el ambiente comprensivo en el cual el comportamiento humano tiene lugar.

Problemas de validez en investigación cualitativa y sus relaciones con los métodos

Yvonna Lincoln (1995), en un artículo en la revista de *Indagación Cualitativa*, presenta el desarrollo histórico del diálogo sobre la validez y el dilema de los criterios en el contexto de la investigación naturalista. Ella localiza los cinco criterios que ella y Egon Guba describieron por primera vez en 1985 dentro de un contexto histórico.

En nuestra discusión de los presupuestos filosóficos de la investigación cualitativa y cuantitativa al inicio de este libro, notábamos que la noción de *verdad* es problemática porque la verdad deviene una construcción social, idiosincrática y situacionalmente específica. ¿La validez es un criterio posible dada la naturaleza de la indagación cualitativa? En el artículo de Lincoln, ella delinea las respuestas a lo que Denzin (1994) llamara la “crisis de legitimación” en investigación cualitativa.

Muchos metodólogos cualitativos y pensadores han respondido a esta pregunta por la validez (veracidad o legitimación), y se presenta un breve resumen para proponer nuestra propia respuesta, la perspectiva post-positivista, en contexto.

Hammersley (1992) describe cuatro filosofías o estadios que reclaman diferentes maneras de pensar sobre la evaluación de la investigación cualitativa. Primero, describe el “metodismo” como sinónimo de positivismo. Dentro de esa filosofía, el método científico es un camino a la verdad. Hammersley sostiene que en esta perspectiva no podemos conocer nada que existe por fuera de nuestra experiencia. La validez no puede ser conectada con la realidad exterior, así que se conecta con lo que usualmente entendemos por confiabilidad “(el acuerdo entre los hallazgos de diferentes observadores o entre los hallazgos del mismo investigador en diferentes ocasiones) y/o la validez predictiva (el acuerdo entre los resultados de la investigación y mediciones establecidas de la propiedad más relevante)” (p. 196).

El autor contrasta este metodismo con el “realismo”. En esta segunda filosofía, el investigador establece la validez porque se acerca al sujeto de estudio. Hay una fuerte correspondencia entre el conocimiento ganado en la investigación y la realidad que él representa. Esta filosofía asume que la realidad exterior puede ser conocida.

Tercero, dado que los etnógrafos construyen la realidad de los pueblos o las personas que estudian, esta investigación requiere la validez que el autor llama “relativismo”. El relativismo acepta que hay múltiples maneras válidas de dar cuenta de la realidad, basadas en esta epistemología constructivista. El consenso dentro de una comunidad acerca de la verdad de los hallazgos constituye la validez en estos casos.

El “instrumentalismo”, la cuarta filosofía, abandona las preocupaciones acerca de la validez y opta, en lugar de eso, por investigación que hace algún bien. Esta noción sugiere que los planteamientos filosóficos de los teóricos críticos y el feminismo.

Como Hammersley se mueve dentro de varias nociones de validez, el constructo de verdad y la validez son cuestionados en el mundo posmoderno. Se reconocen múltiples realidades; y en un contexto posmoderno, Tierney (1993) utiliza la ficción etnográfica para explorar la vida organizacional. Este autor cita a Kitzinger para definir los propósitos de esta estrategia de investigación. “El intento de introducir en el lector una suspensión voluntaria de la incredulidad, mientras simultáneamente se reconoce que el argumento de cada cual es una “versión”, una “construcción, más que una verdad objetiva, puede llevar la escritura dentro del constructivismo social al reino de las artes (Kitzinger, 1990, p. 189).

Tierney comenta: “A través de la ficción, entonces, reacomodamos los hechos, eventos e identidades con el fin de llevar al lector a una historia de una manera tal que posibilita una comprensión más profunda de los individuos, las organizaciones o los eventos en sí mismos” (p. 313).

Teniendo en cuenta este propósito y este método, la validez asume una perspectiva bien diferente. Tierney propone entonces las siguientes preguntas sobre la validez: ¿los personajes son creíbles? ¿la situación es plausible? ¿el texto me ha hecho reflexionar sobre mi propia vida?

Otra investigadora cualitativa, Patti Lather (1993) señala “cuatro tipos de validez postmoderna” como maneras alternativas de pensar sobre la verdad y rechazar la idea (positivista) de verdad como correspondencia entre las teorías y la verdad (p. 54). La “validez irónica”, la “validez paralógica”, la “validez rizomática” y la “validez incorporada” cuestionan la noción de verdad según los posmodernos. La legitimación de la investigación cualitativa exige formas de asegurar que la investigación tiene valor de verdad y que podemos confiar en ella.

La *validez irónica* de Lather coincide con la epistemología posmoderna porque ve la verdad como un problema. El valor de verdad de la investigación reside en su habilidad para mostrarnos binarios que coexisten, opuestos que coexisten (p. 57). La *validez paralógica* es aquella cualidad de la investigación que la legitima en la medida en que revela paradojas,

cosas “indecidibles”, porciones de significado que son imposibles de categorizar. Lather se refiere a la legitimación de la investigación que incluye fragmentos de entrevistas que no son mediadas por el investigador, que no son interpretadas o que el investigador decide no interpretar para no disminuirlas (p. 57). La *validez rizomática*, a diferencia de los significados profundamente arraigados que se solidifican en las categorías post-positivistas a través del análisis del contenido, se acoge al rechazo posmoderno de la verdad estable. Esta validez llega a través “de cruces, superposiciones, significados sin raíces profundas”. El significado llega a través de mapeos y no sólo a través de la descripción (p. 58). (Se nos recuerda el camino de análisis del paradigma cuantitativo: yendo hacia donde nos lleva el patrón de los datos para construir una teoría). La *validez incorporada* resulta de la naturaleza idiosincrática del estudio, de lo que para Lather es la investigadora sabiendo más de lo que sabe, escribiendo más de lo que es capaz de comprender (p. 59). En esencia, esta es la naturaleza de la interpretación, crear una especie de cierre para el trabajo intelectual y emocional de clasificar y filtrar los datos.

Como el conocimiento generado por la investigación cualitativa es el resultado de una construcción social, Steinar Kvale (1995) afirma que este fenómeno está relacionado con la validez del constructo y es el objetivo de la investigación cualitativa. El autor presenta tres aproximaciones a la validez que resultan de este razonamiento. Adoptando la perspectiva posmoderna, no aplica la correspondencia entre teoría y verdad como definición de validez. La construcción social de la realidad se valida sólo a través de la práctica.

Kvale asume que la validez para los posmodernos debe ser diferente a la validez de los positivistas. La validez es un concepto que parece no implicar una línea divisoria entre la verdad y la no-verdad, de acuerdo con el autor. Para los posmodernos, no puede haber verdad universal, pero puede haber “formas de verdad específicas, locales, personales y comunitarias, con foco en la vida cotidiana y en la narrativa local” (1995, p. 21). Y explica la perspectiva posmoderna más claramente que otros escritores:

La condición posmoderna está caracterizada por una pérdida de la creencia en un mundo objetivo y una incredulidad respecto de las metanarrativas de legitimación (Lyotard, 1984). Con la deslegitimación de los sistemas globales de pensamiento no hay fundamentos para asegurar una verdad universal y objetiva. La dicotomía moderna entre una verdad objetiva y unas imágenes subjetivas se está desmoronando y está siendo reemplazada por una hiperrealidad de señales autorreferenciales. Hay una crítica a la búsqueda moderna de formas fundacionales y a su creencia en el progreso lineal a través de más conocimiento. La dicotomía de las leyes sociales universales y los “yo” individuales y únicos es reemplazada por la interacción de redes locales, donde el yo se convierte en un conjunto de relaciones. El foco está en el contexto local y en la construcción social y lingüística de una realidad perspectivística donde el conocimiento es validado a través de la práctica (p. 24).

Desde la perspectiva de Kvale, nuestro continuum interactivo cualitativo-cuantitativo encaja en la filosofía posmoderna. Asumimos que no hay una verdad objetiva y unitaria sobre los métodos de investigación ni sobre la verdad. Rechazamos la dicotomía de las leyes universales de investigación. sin embargo, no nos acogemos a sus tres nociones de validez. Kvale se refiere a las nociones de validez de Polkinghorne: “La validación se convierte en la cuestión de escoger entre interpretaciones “falsables” que compiten entre sí, o examinar y proveer argumentos en favor de la credibilidad relativa de declaraciones alternativas de conocimiento” (Polkinghorne, 1983, p. 26). Kvale etiqueta estos constructos de validez como *validez de la investigación*, *validez comunicativa* y *validez de la acción*.

La *validez de la investigación* de Kvale es la cualidad de la artesanía. Es el control de calidad por parte del investigador. Hace una lista de cosas que el investigador debe revisar antes de la entrevista, por ejemplo. La consistencia de lo que el sujeto dice en una entrevista es contrastada con otros planteamientos que él o ella hace. Se interroga a otras personas, por ejemplo, en una especie de triangulación. La validación de la investigación incluye cómo las teorías se derivan de los datos: cómo el investigador debería comenzar a conceptualizar el tema. Por ejemplo, Kvale reflexiona sobre la investigación acerca de la calificación. ¿La teoría que emerge de los datos se relaciona con el conocimiento que tenemos sobre la evaluación, o se relaciona con la dinámica política de separar en un sentido discriminatorio grupos de estudiantes? En la medida en que la teoría se alinea con los datos y con el objetivo de la investigación, asume validez.

La segunda validez señalada por Kvale es la que él llama *validez comunicativa*. La describe así:

La validez comunicativa involucra el testeo de la validez de las declaraciones de conocimiento que están en diálogo. El conocimiento válido no es simplemente obtenido por aproximaciones a una realidad social dada; involucra una conversación sobre la realidad social: qué es una observación válida es algo que se decide a través de la argumentación de los participantes en el discurso (p. 30).

El tercer tipo de validez de este autor es la *validez de la acción*, en la cual la justificación de la verdad de la investigación se basa en si esta funciona o no. Kvale se refiere a Patton (1990) y a su noción de credibilidad y la vincula con su propia noción de validez de la acción. De acuerdo con Patton, la prueba de credibilidad de un reporte de evaluación es si es utilizado o no por los decisores políticos. La verdad, de acuerdo con Kvale, es cualquier cosa que sea útil para tomar acciones y llegar a un resultado deseado. El autor hace una comparación con Polkinghorne (1991), quien declara que la validez de los estudios de caso, las narrativas y demás puede ser probada si se los contrasta con sus efectos en la práctica. Kvale hace una referencia final a Lincoln y Guba (1985), que dicen que la indagación ayuda a la comprensión y, a través de la comprensión, los participantes pueden aumentar el control que tienen sobre sus vidas.

Resumen: criterios para la validez de los diseños (predominantemente cualitativos)

Lo que sigue es una lista de criterios que pueden ayudar a analizar, sobre todo, investigación cualitativa, en busca de la validez de su diseño. Se incluyen preguntas que pueden ser utilizadas para someter a prueba la validez de los métodos en estudios predominantemente cualitativos. Estas estrategias se sugieren como un punto de partida y no como una lista exhaustiva y son tomadas de Guba y Lincoln (1982), así como de Goetz y LeCompte (1984) y McMillan y James (1992).

1. *Neutralidad*

¿Qué tan objetivos son los datos? Dado que ninguna recolección de datos es totalmente objetiva, el lector deberá ver en dónde esta investigación entra en el continuum objetivo-subjetivo. El lector deberá observar la naturaleza sesgada o no de sus indagaciones. ¿Los datos están disponibles para que el público los vea? ¿Los juicios están documentados con evidencia, a diferencia de las puras opiniones del autor? ¿Hay más de un observador? Si es así, ¿hay consistencia entre los observadores? ¿hay consistencia en la interpretación de los datos?

2. *Compromiso prolongado*

¿El autor observó durante el tiempo suficiente para lograr una representación adecuada de la cultura o de la historia? Si sólo se tomó una observación, ¿reflejará una pequeña porción o no capturaré la esencia de la cultura o de la situación? Un ejemplo de esto sería un visitante del espacio exterior que llegara a Ohio durante el invierno y reportara que hay árboles sin hojas. Otro visitante del espacio llegaría en verano y reportaría que hay árboles con hojas. Las observaciones de ambos serían acertadas, pero ninguno de los dos reflejaría con precisión la situación real.

3. *Observación persistente (consistente)*

¿Se pasó tiempo suficiente en el lugar de la investigación como para lograr una imagen adecuada de la consistencia en el comportamiento de los actores? ¿Lo que se observó fue lo típico, o se detectó algo que no ocurre normalmente? El informe final de investigación debe reflejar esto. Aunque esto suena similar al compromiso prolongado, es diferente en el sentido en que el propósito del compromiso prolongado es ser capaz de detectar tendencias culturales o idiosincrasias, mientras que el propósito de la observación persistente es identificar o estimar si un comportamiento particular (o si un conjunto de comportamientos) es frecuente o infrecuente.

4. *Comunicación con pares*

¿El investigador habló con otro profesional o buscó otra perspectiva sobre lo que estaba experimentando? Las reuniones informativas con pares de alguna manera ayudan a controlar lo que en términos psicoanalíticos se llama contratransferencia. Esto es, uno puede apegarse mucho al ambiente en el cual está tratando de ser un observador objetivo. Los investigadores pueden empezar a interpretar las cosas desde su propia base de necesidades. Necesitan a otro profesional y sus interpretaciones como experto, para darle retroalimentación.

5. *Triangulación*

¿El investigador trató de obtener una variedad de fuentes de datos (por ejemplo, diferentes observadores o diferentes historias escritas?) si así fue, ¿había en ellas una realidad compartida? Aunque la triangulación es un concepto importante en gran parte de la literatura de metodología cualitativa, puede ser considerada también cuantitativa en el sentido en que el investigador cuantitativo busca la consistencia. En alguna medida, la triangulación puede ser vista como una revisión de la confiabilidad, pero no siempre. Es posible que una fuente de datos sea más importante que otras fuentes para la comprensión de un fenómeno en particular. Generalmente, sin embargo, cuantas más fuentes consideremos es más probable que tengamos una percepción más completa de los fenómenos.

6. *Revisión de los miembros*

La revisión de los miembros se refiere a qué tan precisos son los datos. ¿Los datos e interpretaciones se revisan constantemente? Una forma de estimar la exactitud de las observaciones personales sobre los miembros del grupo experimental es observando. Esto es, cuando el investigador vuelve sobre aquellas personas a las que entrevistó y revisar para asegurarse de que él o ella lo hizo bien, el investigador está haciendo revisión de los miembros.

7. *Materiales de referencia*

¿El investigador utilizó suficiente material de soporte (por ejemplo, grabaciones documentadas, lecturas, archivos, u otros materiales que están disponibles para los demás)? Es importante para el investigador documentar las referencias, las grabaciones y las entrevistas utilizadas y dar a conocer a los evaluadores qué tan consciente está de estos materiales. También es importante indicar qué fuentes son utilizadas en qué formas, si alguna fuente disponible no fue usada y por qué.

8. *Relaciones estructurales*

¿Hay consistencia lógica entre diferentes conjuntos de datos? Cuando se intenta interpretar los datos y formalizar las conclusiones, el investigador debería apoyar estas consideraciones, en la medida de lo posible, entretejiendo diferentes conjuntos de datos que vengan de diferentes perspectivas y que soportan un significado subyacente y emergente común.

9. *Muestreo teórico*

¿El investigador siguió los datos a donde ellos conducían? Lo típico es que el investigador entre en el campo y comience inmediatamente a recoger datos. Mientras recoge los datos, el investigador comienza a formarse explicaciones sobre su significado. Estas investigaciones tentativas sugieren otras fuentes de datos. En otras palabras, el muestreo de datos en investigación cualitativa es determinado por los datos existentes (Goetz & LeCompte, 1984). El investigador trata de capturar la mejor teoría para explicar los datos. Un investigador cualitativo podría llamar a esto “testeo de hipótesis blandas”. El investigador puede cambiar

la dirección o recoger datos diferentes y/o adicionales. Por otro lado, ese muestreo de datos puede proveer interpretaciones que apoyen o corroboren la teoría emergente.

10. Trazabilidad

¿El investigador tiene buena documentación, de forma que otro investigador pueda fácilmente replicar la investigación? Esto no sólo significa que alguien sea capaz de replicar el estudio actual, sino que también pueda confirmar o contradecir la interpretación, con base en los mismos datos.

11. Generalización

La posibilidad de generalizar subyace a la ciencia en general. Sin embargo, no estamos dispuestas a aceptar del todo que la posibilidad de generalizar sea consistente con el paradigma cualitativo. Hemos declarado a lo largo de este libro que, en principio, la posibilidad de generalizar es el propósito de la investigación cuantitativa, no de la cualitativa. De hecho, hemos asumido que, si el propósito de la investigación es generalizar, uno debería emplear una metodología cuantitativa. Parece haber un interés creciente entre investigadores cualitativos en la capacidad de generalizar, aun cuando esto parece una violación de los presupuestos básicos de la filosofía naturalista. Polkinghorne (1991), un metodólogo cualitativo, distingue entre dos tipos de generalizaciones: la *estadística* y la *agregada*. El modelo estadístico es más consistente con los presupuestos cuantitativos; y el modelo agregado, basado en descriptores profundos, es más consistente con los presupuestos cualitativos. Los descriptores profundos son suficientemente comprensivos para permitir al investigador cualitativo generalizar a cada uno de los miembros de la población. Donmoyer (1990) también plantea que la posibilidad de generalizar puede ser más amplia que la definición tradicional. Este autor acepta la generalización tradicional para investigación estadística (cuantitativa) y desarrolla el esquema de una teoría basada en la noción piagetiana de asimilación, acomodación, integración y diferenciación (investigación cualitativa) (p. 197). El cuerpo creciente de investigación cualitativa publicada ha ejercido presión sobre los metodólogos de investigación para que creen formas de que los resultados de estos estudios puedan ser aplicados a audiencias mayores, o generalizados. Los siguientes conceptos, aplicabilidad, transferibilidad (limitada al contexto) y replicabilidad, son ejemplos de los tipos de preguntas que podemos formular para contribuir a estos esfuerzos.

Aplicabilidad: ¿Esta investigación puede ser aplicada a otras muestras? El lector debería mirar el tamaño de la muestra, así como sus características. Los presupuestos importantes son que la hipótesis es emergente y no un enunciado del tipo “de la muestra a la población”; y que el propósito de los descriptores profundos es describir en detalle las características de la muestra que está siendo investigada, de forma que otros puedan hacer juicios lógicos sobre si la muestra es comparable con otras muestras. En la medida en que las muestras sean similares, los resultados se pueden aplicar cómodamente.

Transferibilidad: ¿Los resultados de investigación se sostienen en otras situaciones o contextos? En la medida en que puede argumentarse, desde un punto de vista lógico o basado

en los datos, que lo que se está observando y construyendo no depende del contexto en el cual es observado, que no es limitado al contexto, podemos decir que es transferible a otros contextos, y que es, por lo tanto, generalizable. Por ejemplo, los efectos los elogios sobre los estudiantes pueden ser más independientes del contexto que específicos de un contexto; esto es, los efectos del elogio pueden afectar lo académico, así como el desempeño en deportes y situaciones sociales.

Replicabilidad (consistencia): ¿Cuál es la posibilidad de que un resultado o evento vuelva a pasar si se dan las mismas circunstancias? La replicabilidad es difícil de lograr con cualquier nivel de confianza, especialmente en un entorno natural. Debemos identificar los cambios que se deben a efectos identificados y la frecuencia de esas ocurrencias comunes en diferentes puntos del tiempo, en diferentes contextos y por diferentes observadores. Cuando estos datos están disponibles, son valiosos para proveer reflexiones importantes.

Análisis negativo de casos: ¿El investigador ha tenido en cuenta todos los casos conocidos? Revisar continuamente las hipótesis emergentes hasta que todos los datos sean explicados por la hipótesis es la definición de análisis negativo de casos. En esencia, es una secuencia para expandir y volver a dar forma a la interpretación inicial, hasta que todos los elementos estén incluidos. La predisposición de un investigador cuantitativo a incluir lo que aparece una posibilidad de variaciones casuales en los datos.

12. Valor de verdad (credibilidad)

¿Qué confianza tiene el lector en los resultados de la investigación? previamente describimos el concepto de validez del constructo, declarando que un instrumento tiene validez de constructo en la medida en que tenga todos los otros tipos de validez (o sea, validez superficial, contenido, juicio experto, simultánea y predictiva). De forma semejante, un estudio tiene valor de verdad en la medida en que los 12 componentes existan. Es poco probable que cualquier estudio tenga todos los componentes; y, para todos los estudios, algunos de esos componentes son más importantes que otros pero, por lo general, cuantos más componentes presentes, mayor el valor de verdad.