

EL ARTE DEL MUESTREO CUALITATIVO Y SU IMPORTANCIA PARA LA EVALUACIÓN Y LA INVESTIGACIÓN DE POLÍTICAS PÚBLICAS: UNA APROXIMACIÓN REALISTA

JUAN DAVID PARRA*

Resumen

Este artículo hace un examen crítico a aproximaciones tradicionales al muestreo cualitativo para la evaluación y la investigación de políticas públicas. En general, se trata de un debate lleno de contradicciones y de lugares comunes que limitan la capacidad de comprender mecanismos complejos que subyacen a la existencia de fenómenos o eventos sociales. Como alternativa, se exaltan las bondades del muestreo realista y su lógica de estudiar, verificar y refinar teorías preliminares sobre el cómo, el cuándo y el porqué de las políticas públicas. Se concluye que los principios del realismo deben invitar a investigadores y a evaluadores expertos, y también a instituciones contratantes de estudios y evaluaciones sociales, a replantear

prácticas y definiciones metodológicas frente a la escogencia de muestras cualitativas.

Palabras clave: muestreo cualitativo; realismo; políticas públicas; evaluación.

THE ART OF QUALITATIVE SAMPLING AND ITS IMPORTANCE FOR POLICY EVALUATION AND RESEARCH: A REALIST APPROACH

Abstract

The article engages in a critical examination of traditional approaches informing qualitative sampling techniques in policy evaluation and research. This debate is full of contradictions and commonplaces that limit

* Doctor en Ciencias Sociales, International Institute of Social Studies (iiss) de la Universidad Erasmus de Rotterdam. Docente, Universidad del Norte, Barranquilla (Colombia). [jparrad@uninorte.edu.co], [https://orcid.org/0000-0003-1902-7660].

Recibido: 18 de enero de 2019 / Modificado: 08 de abril de 2019 / Aceptado: 21 de abril de 2019

Para citar este artículo

Parra, J. D. (2019). El arte del muestreo cualitativo y su importancia para la evaluación y la investigación de políticas públicas: una aproximación realista. *OPERA*, 25, 119-136.

doi: <https://doi.org/10.18601/16578651.n25.07>

the capacity of studying causal mechanisms in society. As an alternative, the article highlights the strengths of realist sampling and its logic of verifying and refining preliminary theories about how, when and why public policies work. The conclusion indicates that the consideration of realist principles should invite researchers and professional policy evaluators, and also institutions that provide funding for social studies and evaluations, to reconsider standard practices vis-à-vis qualitative sampling.

Key words: Qualitative sampling; realism; public policy; evaluation.

INTRODUCCIÓN

¿Cuántos municipios o regiones se deben visitar? ¿Con cuántas personas es necesario hablar? ¿Cuántos grupos focales son suficientes para resolver la pregunta de investigación? Estos son algunos de los interrogantes claves al momento de emprender un análisis cualitativo. El tema también compete, en particular, al trabajo de evaluadores de políticas públicas interesados en entender cómo, y en qué contextos, ciertas intervenciones sociales podrían contribuir a resolver problemas concretos de una población. Sin embargo, persisten confusiones y malentendidos metodológicos sobre cómo elegir muestras cualitativas *idóneas*. La vigencia del debate en la literatura académica así lo demuestra (Emmel, 2015; Manzano, 2016; Gentles y Vilches, 2017). El presente texto contribuye a esa reflexión.

Dicho lo anterior, la intención no es presentar un examen exhaustivo estilo libro

de texto sobre métodos y paradigmas de investigación cualitativa. El artículo se enfoca más en *la praxis*, o, al menos, en responder a planteamientos comunes que surgen en etapas tempranas de la preparación de un diseño de investigación para evaluar una política o intervención social. Un punto de partida razonable para emprender esa tarea consiste en contrastar técnicas de muestreo comunes como, por ejemplo, la (simple) consideración de criterios de saturación de información, versiones particulares del muestreo teórico o la utilización de razonamientos estadísticos (p. ej., escogencias aleatorias), con el fin de resaltar sus fortalezas y limitaciones. En medio de estas aproximaciones al debate, se aboga por un enfoque metodológico en particular: el *muestreo realista*. Para los investigadores realistas, “no es el número de casos lo que importa, es el trabajo que se demuestra que hacen en la interpretación y explicación [del problema de investigación] lo que cuenta” (traducción propia) (Emmel, 2013, p. 160). Un objetivo del texto es presentar al lector una serie de argumentos epistemológicos que lo persuadan de que, lejos de tratarse de una apología al relativismo, el anterior es un llamado a la búsqueda de acumular conocimiento útil para delinear intervenciones sociales que respondan a retos sociales contemporáneos.

Las ideas que se discuten a continuación son relevantes para el debate sobre la enseñanza y la investigación de las políticas públicas y la administración pública en Latinoamérica. Lo anterior, en tanto existen cada vez más argumentos para tomarse en serio el análisis cualitativo, más allá de presentarlo como un simple complemento (y de menor importancia) a los

paradigmas cuantitativos de evaluación dominantes (Cartwright y Hardie, 2012; Pawson, 2013). Las denominadas *técnicas de evaluación de impacto*, por ejemplo, son relevantes para contar con panoramas generales de los posibles logros y limitaciones de programas y proyectos sociales, “pero [insuficientes] para la realización de una evaluación integral que pretenda explicar los resultados alcanzados [...] [e] identificar [...] cuál es la parte del cambio observado en un problema que puede considerarse como consecuencia de una intervención específica” (Cardozo, 2013, pp. 56-57). Una mirada cualitativa es fundamental, según se discute en el texto, para resolver dichos interrogantes y así evitar que la evaluación de intervenciones sociales se convierta en un ejercicio poco sustancial para fomentar transformaciones sociales deseables (Deaton, 2010; Deaton y Cartwright, 2018).

¿CUÁNTOS CASOS Y POR QUÉ?: APROXIMACIONES COMUNES AL MUESTREO CUALITATIVO

El debate sobre la selección de muestras cualitativas se ha convertido en un encuentro de lugares comunes, poca reflexión metodológica y, en los escenarios más extremos, en una búsqueda fútil por replicar formas de razonamiento cuantitativo. Por un lado, investigadores informados por paradigmas constructivistas tienden a esgrimir argumentos que resaltan *la naturaleza de la investigación cualitativa*, para reiterar que sus resultados “no buscan obtener leyes o principios, sino, examinar comprensivamente a los actores y contextos involucrados

en el tema de estudio” (Ventura-León y Barboza-Palomino, 2017, p. 325). A juicio de Cuenya y Ruetti (2017), la carencia de un sustento (al menos) razonable por parte de los exponentes de dicha posición epistemológica le resta claridad (p. ej., sobre cómo poner en práctica sus principios) y credibilidad a su postura relativista frente al muestreo cualitativo, la cual “no pareciera realizar ningún aporte [crítico] al paradigma de la ciencia cuantitativa” (p. 274).

Otra es la postura comúnmente defendida por investigadores y consultores que argumentan trabajar bajo principios de la denominada teoría fundamentada (TF) para generar teorías causa-efecto basadas exclusivamente en apreciaciones empíricas de la sociedad. En dicho escenario, “el muestreo está dirigido por el desarrollo de [una] teoría [y] es esta la que determina la selección de los siguientes casos o incidentes que se van a recolectar” (Ardila y Rueda, 2013). Por lo general, quienes trabajan bajo esta lógica apelan al denominado muestreo intencional (o *purposive-sampling*), el cual consiste en elegir casos *ricos en información* a partir de los cuales sea posible identificar los cimientos iniciales de un ejercicio de construcción teórica (Martínez-Salgado, 2012). En términos epistemológicos, la teoría fundamentada (TF) parece ofrecer criterios más claros a los del constructivismo al momento de definir muestras cualitativas. La crítica en este caso (y que se profundizará en las siguientes secciones) se centra en el énfasis excesivo que dicha postura da al papel del investigador al momento de descifrar mecanismos sociales complejos a partir de un simple ejercicio inductivo y observacional (Downward, Finch y Ramsay, 2002; Hsieh y Shannon, 2005; Emmel, 2013).

Un tercer caso de aproximaciones comunes al tema es el de la extrapolación directa de razonamientos cuantitativos para elegir muestras cualitativas. Es tan extrema esta postura que es difícil encontrarle respaldo en la literatura académica especializada. Se menciona, sin embargo, porque entre especialistas en el campo profesional de la evaluación no es un secreto que se trata de una práctica vigente y que responde, quizás, a las presiones ejercidas por instituciones públicas y privadas en busca de obtener respuestas *generalizables* en las evaluaciones que contratan. Lo anterior, muy a pesar del amplio reconocimiento entre expertos de que las conexiones estadísticas entre variables dicen poco, o nada, de las relaciones sustantivas que permiten que se conecten entre sí para explicar el surgimiento de eventos de interés (p. ej., el alto o bajo desempeño escolar, la mortalidad infantil, la violencia perpetrada por grupos armados, etc.) (Sayer, 2000; Parra, 2016, 2018). En consecuencia, resalta Sotomayor (2008), se cae en el riesgo de reproducir diseños espurios de investigación:

... la recomendación hartamente sugerida [...] de incrementar la N y así lograr mayor representatividad es una medida que no resuelve bien ni del todo los sesgos [de análisis] y es además inadecuada para los métodos cualitativos. Por principio, una N más grande puede introducir mayor variación en el estudio, pero eso por sí solo no reduce el sesgo. [...] Esta estrategia dispuesta a seleccionar casos aleatorios de una muestra grande conlleva otro riesgo inherente importante; es decir, se pueden seleccionar casos absolutamente irrelevantes o cuyas implicaciones sean nulas para el estudio (p. 166).

Este primer trazo general frente a argumentos comunes para justificar escenarios de muestreo evidencia una especie de sesgo empirista, el cual dificulta generar criterios más concretos para identificar, por ejemplo, el volumen de información requerido para estudiar y sustentar las conclusiones de un análisis cualitativo. La postura realista, como se discute a continuación, indaga sobre elementos concretos (p. ej., el tipo de hipótesis que se intenta verificar o rebatir) que contribuyan a generar criterios para elegir muestras cualitativas idóneas.

SENTANDO LAS BASES PARA UNA ALTERNATIVA DE MUESTREO REALISTA

¿Qué se entiende por realismo? Este es un primer interrogante que puede surgir tras la lectura del título de la sección, y que resulta fundamental para construir un argumento lógico que permita exaltar las bondades de estrategias del denominado muestreo realista. Más relevante aún, ¿qué significa pensar en términos realistas frente a las políticas públicas? El trabajo conceptual de Pawson (2013) propone algunos elementos indicativos para dar respuesta a dicho interrogante. Según el autor:

Un supuesto básico de la evaluación realista es que los programas son intervenciones complejas introducidas en sistemas sociales complejos [...]. Las intervenciones en todos los dominios sustantivos de las políticas públicas ofrecen incentivos para tratar de cambiar el comportamiento (exenciones de pago para ampliar la participación

universitaria, subsidios para que empleadores acepten a desempleados, etc.). La “incentivación” es, por lo tanto, un mecanismo genérico de [las políticas públicas] y una ciencia de evaluación haría bien en construir un entendimiento común de sus fortalezas y debilidades generales (pp. 33, 83-84) (traducción propia).

En el entender realista, toda intervención pública (o privada) implica introducir recursos materiales (p. ej., subsidios, inversión en infraestructura) o culturales (p. ej., nuevas prácticas, entrenamiento docente) en un ámbito humano (p. ej., una región, un segmento poblacional) preexistente. Lo anterior, bajo la expectativa de que dichos recursos *incentivarán* un comportamiento específico, entre miembros de una población, conducente a una transformación social (p. ej., el aumento del consumo de X bienes, mejoramiento de las prácticas pedagógicas en colegios públicos). La tarea de un evaluador de intervenciones públicas consiste, por ende, en estudiar bajo qué contextos, y hasta qué punto, los programas sociales modifican conductas humanas, y cómo esto se relaciona con el cumplimiento de metas esperadas (y no esperados) de las políticas públicas (Pawson, 2013; Parra, 2017).

Tal razonamiento parte del supuesto de la existencia de circunstancias *específicas* en

donde una intervención social puede activar mecanismos, o fuerzas causales (*reales*) concretas. Se trata, sin embargo, de una postura realista *crítica*, alejada del realismo clásico (p. ej., *a la Morgenthau*) o de la pretensión del imperativo cuantitativo (Michell, 2003)¹, en tanto reconoce que para explorar incentivos al comportamiento humano es necesario trascender la visión simplista y mensurable del instrumentalismo racionalista (Deaton y Cartwright, 2018)². Dicho de otra forma, la noción de complejidad inserta en las definiciones propuestas por Pawson (2013) implica desarrollar estrategias metodológicas para estudiar las causas subyacentes detrás de las percepciones y las acciones de individuos y colectividades en diferentes escenarios sociales. La tarea es compleja, en tanto dichos mecanismos comportamentales están permeados por la incidencia de fenómenos contingentes (o espontáneos) y por la reflexividad (no siempre predecible) de los seres humanos en medio las circunstancias específicas que enfrentan (Parra, 2016).

De vuelta a los paradigmas de investigación cualitativa previamente reseñados, desde el realismo, una limitante importante de la TF es que esta tiende a negar la complejidad tal y como fue esbozada en el párrafo anterior. Según indican Downward, Finch y Ramsay

¹ Según explica este último autor, esta es una postura epistemológica según la cual la cuantificación es el único medio legítimo para estudiar de manera científica la sociedad.

² Este es un tipo de racionalidad comúnmente defendida entre economistas y que asume que las personas siempre buscan maximizar algo (p. ej., utilidad, felicidad, etc.). Deaton y Cartwright (2018) resaltan cómo, por años, en la teoría de la evaluación ha dominado la lógica metodológica de que es posible entender los mecanismos causales de las políticas públicas de la misma manera en que se conceptualiza el movimiento gravitacional entre precios y cantidades. Bajo dicha práctica, los evaluadores no se centran en estudiar el comportamiento humano en contexto, sino que lo asumen (p. ej., es racional), y a partir de ello intentar explicar el porqué de las políticas públicas.

(2002), los exponentes de las versiones más tradicionales de este enfoque³ trabajan bajo una lógica fuertemente inductiva, y asumen, por ende, la posibilidad de capturar (con cierto grado de certeza) la esencia de mecanismos sociales complejos a partir de la (simple) observación y el discernimiento (netamente) empírico. En tal esquema, más que reconocer las particularidades de un contexto social en el que entra a operar una política pública para potenciar, o no, la activación de mecanismos sociales, el análisis de causalidad se centra en la capacidad (subjética) del investigador de leer e identificar relaciones entre categorías de información (Hsieh y Shannon, 2005). La pretensión de generar teorías sociales objetivas (o producto de una lectura no sesgada de información recolectada de forma neutral), carece, por tanto, de un argumento epistemológico sólido (Downward, Finch y Ramsay, 2002), y resta poder analítico al argumento de un muestreo teórico empirista.

De otro lado, si se toman en serio los principios del realismo, habría que descartar (o, al menos, problematizar con ahínco) los extremos del constructivismo y del uso de extrapolaciones del razonamiento cuantitativo que respaldan el diseño de muestras cualitativas. En el primer caso porque, como lo resaltan Cuenya y Ruetti (2017), se parte

de la “creencia de que no existe una realidad independiente del observador, sino que esta posee muchas interpretaciones validadas por las experiencias asociadas con la explicación a través del lenguaje” (p. 274). La forma deliberada en que sus exponentes descartan la existencia de *causas reales* detrás de los resultados de una política pública lleva a preguntar si este tipo de evaluadores tienen algún interés en contar con un diseño muestral específico (Parra, 2017). El segundo caso sería uno que se asemeja al imperativo cuantitativo mencionado. Por tanto, además de cómo ello exalta una contradicción inmanente (al abordar un debate en el que, en principio, los investigadores cuantitativos ortodoxos poco creen), se cae en argumentos epistemológicos débiles que privilegian técnicas de muestreo poco alineadas con el estudio de realidades que trascienden la percepción humana⁴.

Las siguientes secciones toman elementos de estas críticas a prácticas y argumentos de soporte comunes en la investigación cualitativa a fin de proponer criterios más concretos para la delineación de estrategias de muestreo. No obstante, antes de dar ese salto conceptual vale la pena hacer un repaso muy breve sobre los posibles fundamentos epistemológicos de las diferentes estrategias de muestreo cualitativo reseñadas. Se trata del debate sobre la

³ Es relevante mencionar que existen versiones de la TF que se alinean con el constructivismo social (TF constructivista). Según la exposición de Emmel (2013), la confusión en torno a sus procedimientos lleva a que se le pueda aplicar una crítica que está en consonancia con la que se le hace en este texto al muestreo constructivista (o relativista).

⁴ Cabe recordar que los métodos cuantitativos trabajan sobre indicadores mesurables o índices de percepción, y, por tanto, reducen el estudio de las dinámicas sociales a la identificación de patrones o correlaciones entre datos empíricos.

deducción y la inducción, como artefactos analíticos comunes en ciencias sociales, y la introducción de la abducción y la retroducción como la alternativa en la que se cobija el paradigma realista de evaluación.

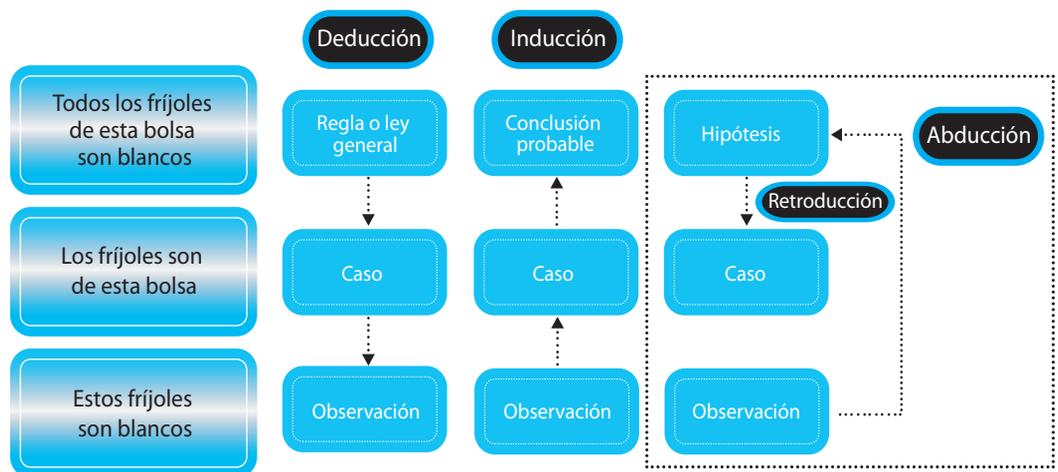
ABDUCCIÓN Y RETRODUCCIÓN: APUNTES SOBRE LA BASE EPISTEMOLÓGICA DEL REALISMO

El contraste de la sección anterior entre aproximaciones realistas y no realistas al estudio de fenómenos sociales complejos puede resultar arbitrario si no se apoya en fundamentos epistemológicos (es decir, desde la reflexión del *cómo estudiar* algo) claros. El debate sobre formas de llevar a cabo inferencias analíticas en estudios científicos permite identificar dicho sustento epistemológico. La figura 1 fue construida a partir de los trabajos de Danermark *et*

al. (2002) y Wuisman (2005) en dicha materia. El ejemplo de los frijoles es relativamente simple, pero ilustrativo de las diferencias relevantes entre formas tradicionales de fundamentar estrategias de investigación en las ciencias sociales: la deducción y la inducción. El ejemplo sirve también para introducir otras formas de razonamiento, quizás menos reconocidas (al menos explícitamente), pero que han tenido un papel fundamental en la historia de la ciencia: la abducción y la retroducción.

Por lo general, la deducción se entiende como un ejercicio de razonamiento lógico de migrar desde lo general a lo particular. En la figura 1, si se parte de una premisa fáctica de que todos los frijoles de la bolsa son blancos, es lógico suponer que cualquier muestra de frijoles que se extraigan de esa misma bolsa serán de ese color. Un claro ejemplo de este razonamiento en las ciencias sociales es la economía matemática en la cual basta proponer

FIGURA 1. MODOS DE INFERENCIA USADOS EN LAS CIENCIAS SOCIALES



Fuente: adaptado de Danermark *et al.* (2002).

una serie de premisas axiomáticas (p. ej., sobre las preferencias de consumidores y productores) para deducir la operación generalizable de mecanismos de mercado en la sociedad. El caso contrario sería el de la estadística aplicada, o la econometría; si se observa que una cantidad considerable de fríjoles son blancos, y se presume que todos provienen del mismo lugar, se puede concluir que dicha característica obedece a la totalidad de granos en la bolsa. En este segundo caso —es decir, el de la inducción— es claro que la certeza no depende de una premisa irrefutable de partida, sino de una función de probabilidad; si la muestra de fríjoles que se observa es estadísticamente significativa, se puede predecir, con cierto margen de error, el resultado para el universo de fríjoles. En ambos casos, cabe resaltar, el ejercicio de inferencia se hace a partir de los datos *presuntamente* asociados, de manera directa, al fenómeno estudiado.

Visto desde un paradigma realista de las ciencias sociales, los modos de inferencia lógica descritos arriba reducen la ciencia a una actividad demasiado estrecha (p. ej., verificar cosas que ya se saben) o que confía en exceso en la uniformidad del mundo para sustraer normas generales de datos empíricos *probablemente* relacionados entre sí (Parra, 2016). En este punto es relevante mencionar que, para efectos de las estrategias de análisis cualitativo esbozadas hasta el momento, la crítica que merece mayor consideración es la que se hace a la lógica de la inducción. Por ejemplo, en la TF, el reto consiste en sistematizar un método que permita moverse desde la observación empírica a la propuesta de teorías explicativas (Emmel, 2013). El argumento inductivo es que dicha explicación puede agotarse por medio de intentos

de generalización basados en datos y, por tanto, desprovistos de sesgos analíticos (p. ej., ideas *a priori* sobre cuál podría ser esa explicación). Para Wuisman (2005), sin embargo, lo anterior emana de una contradicción lógica:

La validez de su conclusión no puede ser probada con absoluta certeza. Para que una inferencia inductiva sea válida, es lógicamente necesario que la regla o ley general se conozca de antemano como una hipótesis. La inducción solo se puede usar para validar esa regla o ley general previamente hipodimensionada. [...] [El problema es que] el origen de la hipótesis sigue siendo un completo misterio (p. 382) (traducción propia).

La figura 1 introduce también los conceptos de la abducción y la retroducción. El primero supone la idea de poner sobre la mesa diferentes teorías o marcos de referencia para proponer posibles explicaciones a datos observados en contextos específicos. Es decir, la “abducción consiste en moverse de una concepción de algo a una diferente, posiblemente más desarrollada o más profunda” (Danermark *et al.*, 2002, p. 91) (traducción propia). La retroducción, por su parte, consiste en “un modo de inferencia en la que los eventos son explicados por medio de la postulación [...] de los mecanismos capaces de producirlos” (Sayer, 1992, citado en Parra, 2016, p. 223). Nótese cómo estas dos últimas definiciones contribuyen a resolver, al menos desde un plano abstracto, el problema lógico de la inducción resaltado por Wuisman (2005). Hacer abducción implica consultar diferentes marcos de referencia (p. ej., teorías sociales, visores conceptuales) desde los cuales interpretar y proponer factores o hipótesis que dan sentido

a la existencia de eventos empíricos particulares (Olmos, 2018). La retroducción, por su parte, es un complemento a la abducción, en tanto plantea estrategias para descartar hipótesis explicativas. Para pensadores realistas, la combinación de esfuerzos abductivos y retroductivos representa una mejor descripción, frente a la de la deducción e inducción puras, de la lógica del descubrimiento científico:

Lo revolucionario de las recontextualizaciones es que dan un nuevo significado a fenómenos ya conocidos. Los descubrimientos de las ciencias sociales están en gran medida asociados con la recontextualización. Los científicos sociales no descubren nuevos eventos que nadie conocía antes. Lo que se descubre son conexiones y relaciones, no directamente observables, por las cuales [se vuelve posible] entender y explicar sucesos ya conocidos de una manera novedosa (Danermark *et al.*, 2002, p. 91) (traducción propia).

La noción de lo *no directamente observable* es clave para exaltar la distinción entre formas tradicionales de inferencia y aquellas cobijadas por el pensamiento realista. Nuevamente, la esencia del pensamiento inductivo es inferir una serie de patrones comunes en el comportamiento de unos datos, a partir de un conjunto de datos similares. En la figura 1 se induce que los frijoles de la bolsa podrían ser blancos, dado que los frijoles que se observan, y que son, aparentemente, de esa bolsa, son de ese mismo

color. En la alternativa abductiva y retroductiva, la inferencia de la realidad (o la explicación) de un conjunto de datos incluye la posibilidad de examinar datos que, al menos desde una primera mirada empírica, no guardan una relación aparente con el fenómeno observado. En el ejemplo, de la información disponible se sabe que en una bolsa particular todos los frijoles son blancos, lo cual permitiría sugerir que la razón por la cual los frijoles que se observan son blancos, y no rojos, es porque provienen de *esa* bolsa de frijoles blancos. Sin embargo, dependiendo del contexto (p. ej., si se está en la cocina de una casa o en un supermercado surtido de granos) podría existir información adicional (o no directamente observable) que lleve a que la explicación inicial sea tan solo una *posible* teoría. Por ejemplo, es posible que un tercero (p. ej., un vendedor) haya seleccionado intencionalmente unos frijoles blancos de una bolsa surtida de frijoles blancos y rojos y los puso sobre el mostrador porque le interesa vender más frijoles blancos. En dicho caso, y cuando se obtenga esa información, habría que descartar que la explicación de que los frijoles que se observan son blancos es debido a que provienen de la bolsa de frijoles blancos originalmente enunciada.

En síntesis, la justificación para migrar hacia estrategias abductivas y retroductivas, más allá de aquellas informadas exclusivamente por la inducción y la deducción⁵, es que el mundo

⁵ La literatura reseñada es también enfática en señalar que estos argumentos no demeritan la importancia de los modos tradicionales de inferencia. La abducción, en sí, representa una forma de inducir posibles explicaciones a patrones de datos que se observan. La retroducción, por su parte, se asemeja a la deducción en el sentido que propone teorías generales a partir de las cuales se podría dar significativo a dichos datos. La diferencia, nuevamente, recae en las fuentes (directas o indirectas) que se consultan en cada caso.

no es tan transparente o uniforme como lo asumen visiones tradicionales de inferencia lógica. El salto lineal de datos a conclusiones no solo acarrea cierta pretensión ingenua de que el investigador cuenta, de antemano, con herramientas, y con niveles de información, suficientes para identificar, por medios empíricos, explicaciones a fenómenos observables. De paso, este ignora elementos no lineales (p. ej., intenciones de individuos, relaciones de poder entre actores sociales, etc.) esenciales para entender por qué el mundo es exactamente como es. Como se discutirá en las siguientes secciones, este es precisamente el pilar epistemológico del muestreo realista; explorar teorías sobre por qué diferentes eventos sociales (p. ej., el crimen en una sociedad, las tasas de mortalidad infantil en una población, etc.) muestran cierto comportamiento, y no otro, en contextos geográficos y temporales específicos.

EL MUESTREO REALISTA Y LA POSIBILIDAD DE ESTUDIAR TEORÍAS SOCIALES EN ACCIÓN

El trabajo metodológico de Emmel (2013) es crucial para informar la siguiente fase del análisis. Con la base epistemológica de la abducción y la inducción, el objetivo es ahora profundizar en criterios (más) concretos para sustentar la definición de muestras cualitativas. Según explica el académico de la escuela de Sociología y Política Social de la Universidad de Leeds:

Uno elige el paquete más adecuado para probar una teoría de cómo eso que se está investigando puede ser

explicado. El truco para elegir los casos es ensamblar la configuración óptima de los ingredientes para refinar una teoría y luego repetir el truco dentro de las limitaciones de recursos del proyecto de investigación. Dentro de cualquier investigación, son las teorías las que informan la elección de la muestra y las estrategias de muestreo que se utilizan para probar la teoría. Estas teorías no son el resultado de datos sensoriales (empirismo). Son suposiciones sostenidas como conjeturas audaces que siempre se refieren a fenómenos reales (Emmel, 2013, p. 69) (traducción propia).

Una lectura inicial de este fragmento puede llevar al lector a pensar en estrategias de muestro teórico similares a las suscritas por representantes de la visión tradicional de la TF. Sin embargo, en este caso no se hace referencia a una posición epistemológica empirista que soporta una construcción teórica a partir de una lectura desprevenida (o insesgada) de información cualitativa, o alejada de postulados teóricos previos. La visión encarnada en este razonamiento coincide más con posiciones como la de Maxwell (2012), quien habla de la importancia de trascender la descripción e interpretación de eventos empíricos para enfocarse en el estudio de construcciones teóricas que el investigador *trae a, y desarrolla durante*, su análisis. Es decir, los llamados sesgos de los que busca escapar la TF pueden representar, en sí mismos, puntos de partida válidos y legítimos para orientar la evaluación de una política pública. La pregunta es si dicho procedimiento garantiza sesgar (o restarle rigor a) los resultados del estudio. Para expertas en metodología realista, como Archer (2011), este no es necesariamente el caso. El siguiente pasaje tiene como contexto una crítica que hace la sociología a un

autor que sugiere que un análisis causa-efecto, como el que interesa a un evaluador, debe partir de visiones más *objetivas y estructuradas* del objeto de investigación:

Creo que él podría estar equivocado *a menos que* insista en un “análisis causal completo” de los poderes [o causas] reales [...] de cualquier unidad [o fenómeno social] emergente en el orden que él propone. Este implica comenzar por una lista las partes que la caracterizan y una explicación sobre cómo estas se relacionan entre sí [...] [Pero la] sociología es mucho más caótica que eso; [el estudio de la sociedad] inicia con el planteamiento de corazonadas [o hipótesis preliminares], no de nombrar las partes, y probablemente no llegaría a ningún lado si iniciara de esto último (p. 61) (traducción propia).

Lo anterior reitera el mensaje asociado a la complejidad tácita en el estudio de las políticas públicas (p. ej., por cuenta de la multicausalidad, por la esencia relacional, y no atomista, de la sociedad)⁶. Ante ello, pierde sentido epistemológico argumentar que un investigador, o un equipo de investigación, cuenta con la capacidad de identificar mecanismos sociales reales a partir de una simple ponderación *objetiva* de información cualitativa o de datos estadísticos. Como lo sugiere la cita de Archer (2011), existen, por tanto, pocas rutas diferentes a examinar en profundidad, y reinterpretar

eventos (o un proceso de abducción), para luego preguntarse cuáles son esos mecanismos que explican el surgimiento y la forma de los mismos (un salto retroductivo) (Belfrage y Hauf, 2017). Dicha tarea, sugieren los realistas, puede (y debe) beneficiarse de prejuicios iniciales (p. ej., preferencias teóricas del investigador), de la lectura de fuentes secundarias, de referencias a la literatura especializada, etc. A partir de ello, la tarea del analista o evaluador de una política pública consistirá en iniciar un ejercicio de contraste entre abstracciones conceptuales e información empírica “para refinar diferentes teorías sobre qué funciona, para quién, en qué aspectos, hasta qué punto, en qué contexto y cómo” (Parra, 2017, p. 11).

Por cuestiones de espacio y de enfoque, el presente artículo no ahondará en estrategias específicas sobre cómo generar esas *corazonadas* o (posibles) teorías preliminares (o las hipótesis abductivas enunciadas en la sección anterior). El enfoque de evaluación realista (ER) promulgado por Pawson (2013) –y examinado, con cierto grado de exhaustividad por Parra (2017)– ofrece un cuerpo conceptual riguroso para emprender ejercicios de abducción y retroducción en el contexto de la evaluación de intervenciones sociales. Brousselle y Buregeya (2018) invitan también a considerar otras propuestas conceptuales pertenecientes a la familia

⁶ En el entender del realismo, los fenómenos que existen en el mundo son resultado de relaciones complejas (p. ej., no lineales) entre elementos materiales o culturales del mundo. Un colegio que aglomera estudiantes con bajos resultados académicos, por ejemplo, es el resultado de una combinación específica, en un contexto determinado (p. ej., un momento del tiempo, una región de Colombia) de recursos (o falta de los estos) escolares, docentes (des) interesados, concepciones específicas de la calidad educativa, etc. Esta es una visión diferente a la de un enfoque atomista, desde el cual se prevé que el mundo es el resultado de la sumatoria lineal de hechos individuales (p. ej., piénsese en un modelo econométrico de la forma $Y = \beta_1 X + \beta_2 Z + \beta_3 W + \varepsilon$).

de las *evaluaciones basadas en teoría*. Se resaltan, de momento, las implicaciones de dicha forma de concebir el estudio de las políticas públicas sobre la definición de muestras cualitativas. En línea con la cita de Emmel (2013) al inicio de esta sección, el muestreo cualitativo tiene como fin último el permitirle a investigadores y analistas estudiar teorías causales sobre el cómo y por qué se dan ciertas transformaciones sociales, y no otras, a raíz de intervenciones de política en la sociedad.

La siguiente sección profundiza en algunos elementos, quizás de orden más práctico, relevantes para poner en marcha un muestreo realista al servicio de un proyecto concreto de evaluación. Al final se discute un ejemplo a fin de ilustrar con mayor detalle algunos de los retos y las bondades de la investigación cualitativa realista de las políticas públicas.

MÁS ALLÁ DE LA UNIDAD DE ANÁLISIS. LA NECESIDAD DE ESTUDIAR TEORÍAS

Desde una postura realista de la evaluación (causa-efecto) de las políticas públicas, uno de los malentendidos más comunes en el debate sobre el muestreo cualitativo gira en torno a la escogencia de la denominada *unidad de análisis*. Por ejemplo, un argumento común en la literatura dicta que “en la indagación cualitativa las unidades de observación y análisis pueden ser [...] personas, familias, grupos, instituciones, áreas geográficas o culturales, períodos [...] programas, documentos [...], entre otras” (Martínez-Salgado, 2012, p. 614). Esto ayuda a entender, a lo mejor, la obsesión que se percibe en algunos círculos de evaluadores

(y, en particular, de quienes contratan evaluaciones de política) por contar con un número *idóneo* de personas con las cuales hablar, de grupos focales por llevar a cabo o de estudios de caso por realizar, para que la historia sea, de alguna manera, representativa (de algo). De hecho, existen cálculos (ciertamente estadísticos) que sugieren el número 12 como una especie de óptimo de cantidad de entrevistas para encontrar elementos de saturación en trabajos cualitativos (Guest, Bunce y Johnson, 2006; Emmel, 2013).

La crítica a la búsqueda de un *punto de saturación*, o “el momento en el cual se ha escuchado ya una cierta diversidad de ideas y con cada entrevista u observación adicional no aparecen ya otros elementos [relevantes para el análisis]” (Martínez-Salgado, 2012, p. 617), no debe centrarse, sin embargo, en desconocer la importancia de valorar *la suficiencia* en la evidencia cualitativa para respaldar un hallazgo de investigación. Incluso, desde el realismo se hace un llamado a contar con cantidades sustanciales de información (Manzano, 2016). Más bien, el debate gira en torno al punto de partida desde el cual se justifican esos niveles *óptimos*. En la TF, por ejemplo, tiende a apelarse a argumentos de orden teórico —es decir, cuando se han logrado explicar la mayor cantidad posible de variaciones dentro de teoría en construcción—, de orden práctico —es decir, según la disponibilidad de información— y de orden técnico —es decir, según la sensibilidad (o capacidad) analítica del investigador— (Ardila y Rueda, 2013). Desde el realismo, el problema radica, nuevamente, en la epistemología empirista en que sustentan esos argumentos. Ello implica, como sugiere Pawson (1996),

un cambio en la forma de entender el proceso mismo de recolectar información cualitativa:

El punto de partida para este esfuerzo es repensar la “tarea” de la entrevista [y la indagación cualitativa, en general], así como el “posicionamiento” del entrevistado. Quizás la diferencia crucial en lo que defiende es un cambio en la forma de pensar sobre el tema de la entrevista [...]. Los [...] modelos “convencionales” tienden a suponer que el foco de la entrevista es el sujeto entrevistado. Por lo tanto, la tarea es determinar (de acuerdo con el método favorito) información que sea fiel a los pensamientos y acciones del sujeto. *En el modelo basado en la teoría, la teoría del investigador es el foco de la entrevista, y el sujeto está allí para confirmar o falsear y, sobre todo, para refinar esa teoría* (p. 299) (traducción propia).

En el paradigma realista, la unidad de análisis de un evaluador o estudioso de las políticas públicas es, por ende, la teoría propuesta para ser verificada o refinada durante el proceso de evaluación/investigación. Bajo dicha óptica, la definición de los individuos, los grupos, las poblaciones o las locaciones geográficas sobre las que se trabaje debe hacerse en función de la teoría subyacente (sobre relaciones entre contextos y mecanismos sociales) que se propone estudiar. Un corolario que se deriva de ello, valga reiterarlo, es que “el muestreo realista debe diseñarse para probar los contextos que importan dentro de la hipótesis. Esto podría no ser solo sobre los sitios. También podría tratarse de grupos de población, o de barreras de implementación y (factores) facilitadores [de un proceso de transformación social]” (Manzano, 2016, p. 348) (traducción propia).

El trabajo de Manzano (2011) sobre sistemas de multas diseñados en 2003 para reducir los retrasos en descargas (o dadas de alta) hospitalarios en Inglaterra sirve para ilustrar algunas de las implicaciones prácticas de estos principios. En primer lugar, para abordar la investigación la autora no parte de un vacío teórico, ni se propuso inferir explicaciones directamente de información empírica. Por el contrario, desde el diseño de su investigación se planteó la relevancia de discutir hipótesis que podrían dilucidar la operación de mecanismos sociales causales en el contexto particular que le interesaba explorar. A juicio de la autora, la frecuente demora en instituciones médicas en dar salida a pacientes puede explicarse por: 1) la falla en la planeación de dichos procesos por cuenta de ineficiencias administrativas, o 2) la ausencia de alternativas de cuidado (p. ej., en hogares, en otras instituciones) para personas que, pese a ser dadas de alta, requieren aún de este.

Para llevar a cabo dicha exploración, precisa Manzano (2011), se descartó todo intento de muestreo probabilístico, en tanto este nunca podría ser representativo ni útil para entender, y explicar, la complejidad del mecanismo operante entre pacientes con ciertas necesidades insatisfechas (p. 24). Según los postulados teóricos de entrada, y el uso de herramientas metodológicas de escuelas de evaluación basadas en teoría, se optó, finalmente, por entrevistar a 39 personas, entre ellas 13 pacientes, 12 funcionarios de un hospital del norte de Inglaterra y 14 miembros de su servicio social (p. ej., trabajadores sociales). Dicho muestreo cualitativo estuvo acompañado de actividades de observación directa de procesos administrativos

de descargas hospitalarias. Como indica el siguiente pasaje (proveniente de un texto más reciente de la autora, pero que reflexiona sobre su trabajo de 2011), desde luego fue necesario tomar decisiones *prácticas*; el punto, sin embargo, es que estas se hicieron luego de considerar las teorías de soporte, y no como resultado de un diseño muestral en abstracto:

... en un estudio que evaluó un programa para acelerar las altas hospitalarias (Manzano-Santaella, 2011), se llevaron a cabo varias reuniones formales e informales con las partes interesadas [...] para discutir cómo y dónde reclutar pacientes con diferentes enfermedades que podrían representar distintos retos para las teorías por ser probadas y refinadas. En consecuencia, las entrevistas cara a cara con los pacientes permitieron al investigador cubrir un rango diferente de experiencias de alta en 14 pacientes muy diversos, desde un hombre de 40 años que necesitaba realojarse después de un derrame cerebral que lo dejó incapaz de usar las escaleras en su apartamento, a una mujer de 87 años con antecedentes de largo plazo de dolores múltiples y que ingresó en el hospital debido a dificultades respiratorias causadas por una enfermedad respiratoria crónica. Aunque el muestreo se basó en consideraciones teóricas (es decir, probando teorías sobre los motivos por los cuales algunas personas permanecen en el hospital más tiempo que otras) tuvo en cuenta también dificultades prácticas. Varios de los posibles participantes se perdieron en las *contingencias* médicas y en la organización social del trabajo médico: [...] las personas se transfirieron a otras salas, a unidades de cuidados intensivos, se les dio de alta sin previo aviso o murieron (Manzano, 2016, p. 350). (Traducción propia).

Se deriva, por tanto, y siguiendo el espíritu del análisis cualitativo, que más que generar

una fórmula para elegir cantidades óptimas, los principios del realismo dan un piso conceptual (más) sólido a la noción de muestreo teórico. Lo anterior, es relevante anotar, permite además resignificar la noción de *puntos de saturación* en fases de análisis de información. El debate no consiste en cuántas entrevistas, grupos focales o categorías para segmentar fragmentos de un documento *saturan* conexiones *relevantes* (desde una construcción teórica basada en datos), sino, nuevamente, en la creación de “una situación en la cual los postulados/conceptos teóricos bajo investigación estén abiertos a la inspección” (Parra, 2017, p. 33).

Luego de presentar una síntesis de los argumentos esgrimidos hasta el momento para defender el muestreo cualitativo realista, la última sección del texto amplía reflexiones sobre las implicaciones de estas ideas en la práctica. En particular se consideran aspectos inherentes a la relación entre consultores e instituciones contratantes al momento de llegar a acuerdos metodológicos que sirvan de sustento para hacer una serie de recomendaciones sobre cómo reorientar algunas prácticas comunes en el análisis e investigación de las políticas públicas e intervenciones sociales.

REFLEXIÓN FINAL: (ALGUNOS) RETOS DEL MUESTREO CUALITATIVO EN LA PRÁCTICA

Este artículo se ha propuesto la tarea de examinar críticamente prácticas comunes en el muestreo cualitativo. Lo anterior, a partir de postulados del paradigma realista aplicado al estudio y la evaluación de políticas públicas e

intervenciones sociales. Este parte de la premisa de la complejidad inherente a cualquier proceso de transformación social, el cual es resultado de las acciones y los comportamientos de seres humanos en contextos específicos (Pawson, 2013). La identificación de mecanismos sociales comportamentales consiste, por tanto, en estudiar procesos causales reales que, debido a dicha complejidad, requieren de herramientas metodológicas (como la abducción y la retroducción) que trascienden el simple empirismo. El problema con los argumentos comunes para definir muestras cualitativas provenientes de paradigmas como el constructivismo, el imperativo cuantitativo o la TE, es que no permiten dar ese salto (hacia lo real).

El muestreo realista cuenta con el potencial de ofrecer a investigadores y evaluadores de políticas públicas criterios (más) concretos, y mejor fundamentados (en comparación con el de los otros paradigmas en mención) al momento de plantear su diseño de investigación. Ello, en tanto asume la complejidad del mundo social como punto de partida, y fomenta esfuerzos explícitos por estudiar construcciones teóricas preliminares sobre el tipo de causas complejas que podrían estar explicando un fenómeno o un evento social. Por motivos de enfoque y de espacio el artículo no se adentra en reflexiones detalladas sobre cómo formular dichas hipótesis, pero sugiere la seria consideración de herramientas metodológicas

de tradiciones de la evaluación basada en teoría (Parra, 2017; Brousselle y Buregeya, 2018) para informar dicha tarea. De lo anterior se deriva el hecho de que una estrategia de muestreo teórico encaminada a estudiar procesos complejos reales no puede ir desligada del planteamiento de hipótesis preliminares (o corazonadas) desde la fase de diseño de una investigación cualitativa. Ese procedimiento no cierra el interrogante de cuánta información es suficiente, pero sí robustece los criterios a fin de elegir el tipo de información necesaria para avanzar en la investigación.

¿Por qué es importante esta reflexión? Como se anticipó desde la introducción del texto, existen voces que reconocen la creciente importancia del análisis cualitativo como eje fundamental para poder avanzar en la acumulación de conocimientos prácticos que potencien el poder transformador de las políticas públicas en la sociedad⁷. Al respecto, vale la pena citar intelectuales influyentes como Deaton (2010), para reiterar por qué, si bien el conocimiento cuantitativo representa una pieza fundamental para estudiar y evaluar intervenciones en la sociedad, la adherencia exclusiva al mismo (por más sofisticado que sea)⁸ resulta insuficiente para incidir en el cumplimiento de metas de transformación social:

... bajo circunstancias ideales, las evaluaciones aleatorias de los proyectos son útiles para obtener una estimación

⁷ Por ejemplo, en un reconocido manual de evaluación de impacto del Banco Mundial, los autores reconocen que “el trabajo cualitativo puede contribuir a que los responsables de políticas comprendan lo que está ocurriendo en el programa” (Gertler, Martínez, Premand, Rawlings y Vermeersch, 2011, p. 17).

⁸ El comentario cobija, por tanto, a todo tipo de innovación en el uso de técnicas cuantitativas modernas, como, por ejemplo, las asociadas al denominado *Big Data*. Para una discusión más profunda sobre esta crítica en particular, consultar Fuchs (2017).

convinciente del efecto promedio de un programa o proyecto. El precio de este éxito es un enfoque demasiado estrecho y demasiado local para decirnos *lo que funciona* en [políticas de] desarrollo, [y] para diseñar políticas, o para avanzar en el conocimiento sobre los procesos de desarrollo. Es improbable que las evaluaciones de proyectos, ya sea mediante ensayos controlados aleatorios o métodos no experimentales, revelen los secretos del desarrollo y, a menos que se guíen por una teoría que esté en sí misma abierta a revisión, es improbable que sean la base de un programa de investigación acumulativo que podría conducir a una mejor comprensión del desarrollo (Deaton, 2010, p. 426) (traducción propia).

El interés acá no es lanzar un ataque contra formas de análisis cuantitativo que, como se reconoce, son importantes para la evaluación de políticas e intervenciones sociales. Se considera más relevante reivindicar el *estatus* del análisis cualitativo como herramienta fundamental (tanto práctica como científica) (Parra, 2018) para mejorar el conocimiento sobre los mecanismos que, según diferentes contextos, puedan efectivamente contribuir al mejoramiento de vida de la gente, transformar la calidad educativa, reducir niveles de contaminación o reducir el embarazo adolescente,

entre otras metas. Para generar ese tipo de evidencia es fundamental contar, es importante reiterarlo, con muestreos cualitativos que contribuyan en esa tarea.

Llegados a este punto, es posible que algunos lectores (quizás evaluadores expertos, o académicos interesados en el estudio cualitativo de las políticas públicas) se planteen una reflexión del tipo: ¿cómo vamos a persuadir a instituciones que invierten recursos millonarios para una evaluación que una muestra grande (como se tienden a solicitar) puede no ser necesaria? La recomendación a estos lectores es concreta: es necesario insistir que, apelando a argumentos del realismo (sobre la complejidad, por ejemplo), una muestra grande no es garante de una buena explicación y que, por el contrario, el insistir en ella sin argumentos diferentes al *más es mejor que menos* puede implicar perder tiempo (p. ej., de análisis) y dinero (p. ej., por recoger información que en nada aporta al estudio). Estrategias complementarias como, por ejemplo, proponer reuniones en fases del diseño de la evaluación para construir, con el apoyo la parte contratante⁹, preguntas relevantes (o hipótesis de trabajo) para la evaluación, puede ayudar

⁹ A pesar de que, como se ha indicado, no se va a entrar a discutir en detalle estos preceptos metodológicos, resulta importante matizar esta afirmación para evitar malentendidos. Según Blamey y Mackenzie (2007) “el evaluador realista articula la teoría a través de conversaciones generales y entrevistas con [...] partes interesadas. [...] Esta decisión se basa en la evidencia existente y en el conocimiento y la experiencia del evaluador, más que en la importancia relativa que los ejecutores ponen en las teorías. En este sentido, aunque las teorías generadas emergen, en parte, de las discusiones con las partes interesadas, estas son especificadas [...] por los evaluadores en lugar de ser aprobadas y *suscritas* por las partes interesadas. Del mismo modo, es el evaluador quien prioriza las [teorías] que son interesantes para hacer una mayor investigación y que se convierten en el foco de la evaluación” (p. 447) (traducción propia). Dicho de otra manera, aunque discutir teorías iniciales con partes contratantes resulta extremadamente útil para enfocar el trabajo del analista, es este segundo el que, haciendo uso de su papel de experto independiente, debe definir las teorías finales que harán parte del estudio.

a reducir márgenes de incertidumbre frente a la posibilidad de hacer un estudio cualitativo riguroso con muestras no probabilísticas de una población.

En dicha línea argumentativa es importante también enviar, desde este escrito, algunos mensajes a quienes están parados en el lado de los contratantes de evaluaciones o investigaciones. Por ejemplo, es fundamental, si se quiere realmente avanzar en la comprensión a profundidad de las políticas públicas, abandonar pretensiones universalistas de tratar de entenderlo todo en una sola evaluación. La condición de complejidad que se ha resalta-do reiteradamente en el artículo supone una limitación, si se quiere filosófica, que así lo reivindica. Desde ese punto de vista puede ser más fructífero, antes de emprender una gran evaluación, contratar estudios más pequeños que permitan ir explorando, con suficiente profundidad, diferentes mecanismos sociales en acción. Dentro de esa lógica, se vuelve fundamental que en cada nueva convocatoria de una evaluación que se haga se motive al consultor o investigador a estudiar teorías preliminares que han surgido de estudios exploratorios. En las primeras aproximaciones (p. ej., estudios intermedios, pilotajes) se puede ser más flexible con los argumentos de muestreo. Ya, en el momento de plantear evaluaciones más comprensivas se podrá contar con teorías intermedias específicas a partir de las cuales aplicar estrategias realistas de muestreo cualitativo.

REFERENCIAS

- Archer, M. (2011). Morphogenesis: realism's explanatory framework. En A. Maccarini, E. Morandi y R. Prandini, *Sociological realism. Ontological Explorations* (pp. 59-94). Oxford: Routledge.
- Ardila, E. y Rueda, J. (2013). La saturación teórica en la teoría fundamentada: su de-limitación en el análisis de trayectorias de vida de víctimas del desplazamiento forzado en Colombia. *Revista Colombiana de Sociología*, 36(2), 93-114.
- Belfrage, C. y Hauf, F. (2017). The gentle art of retro-duction: Critical realism, cultural political economy and critical grounded theory. *Organization Studies*, 38(2), 251-271.
- Blamey, A. y Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, 13(4), 439-455.
- Brousselle, A. y Buregeya, J. (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 4(2), 153-168.
- Cardozo, M. (2013). Políticas públicas: los debates de su análisis y evaluación. *Andamios*, 39-59.
- Cartwright, N. y Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
- Cuenya, L. y Ruetti, E. (2017). Controversias epistemológicas y metodológicas entre el paradigma cualitativo y cuantitativo en psicología. *Revista colombiana de psicología*, 19(2), 271-277.
- Danermark, B., Ekström, M., Jakobsen, L. y Karlsson, J. (2002). *Explaining society. Critical Realism in the Social Sciences*. London-New York: Routledge.
- Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424-455.
- Deaton, A. y Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.
- Downward, P., Finch, J. y Ramsay, J. (2002). Critical realism, empirical methods and inference: A cri-

- tical discussion. *Cambridge Journal of Economics*, 26(4), 481-500.
- Emmel, N. (2013). *Sampling and choosing Cases in Qualitative Research*. London: Sage.
- Emmel, N. (2015). Themes, variables, and the limits to calculating sample size in qualitative research. *International Journal of Social Research Methodology*, 18(6), 685-686.
- Fuchs, C. (2017). From digital positivism and administrative big data analytics towards critical digital and social media research! *European Journal of Communication*, 32(1), 37-49.
- Gentles, S. J. y Vilches, S. (2017). Calling for a shared understanding of sampling terminology in qualitative research: Proposed clarifications derived from critical analysis of a methods overview by McCrae and Pursell. *International Journal of Qualitative Methods*, 16, 1-7.
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B. y Vermeersch, C. M. (2011). *Evaluación de impacto en la práctica*. Washington DC: Banco Mundial.
- Guest, G., Bunce, A. y Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59-82.
- Hsieh, H. y Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Manzano, A. (2011). A realistic evaluation of fines for hospital discharges: Incorporating the history of programme evaluations in the analysis. *Evaluation*, 17(1), 21-36.
- Manzano, A. (2016). The craft of interviewing in realist evaluation. *Evaluation*, 22(3), 342-360.
- Martínez-Salgado, C. (2012). El muestreo en investigación cualitativa. Principios básicos y algunas controversias. *Ciência & Saúde Coletiva*, 17(3), 613-619.
- Maxwell, J. (2012). *A Realist Approach for Qualitative Research*. Los Angeles, London, Nueva Delhi, Singapur y Washington DC: Sage.
- Michell, J. (2003). The quantitative imperative: Positivism, naive realism and the place of qualitative methods in psychology. *Theory & Psychology*, 13(1), 5-31.
- Olmos, P. (2018). La justificación de la abducción en el contexto del debate sobre el realismo científico: una aproximación argumentativa. *ArtefactoS. Revista de estudios de la ciencia y la tecnología*, 7(2).
- Parra, J. (2016). Realismo crítico: una alternativa en el análisis social. *Sociedad y Economía* (31), 215-238.
- Parra, J. (2017). ¿Qué funciona, para quién, en qué aspectos, hasta qué punto, en qué contexto y cómo? Una introducción a la evaluación realista y sus métodos. *Economía & Región*, 11(2), 11-44.
- Parra, J. (2018). Critical realism and school effectiveness research in Colombia: The difference it should make. *The British Journal of Sociology of Education*, 39(1), 107-125.
- Pawson, R. (1996). Theorizing the Interview. *The British Journal of Sociology*, 47(2), 295-314.
- Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. London: Sage.
- Sayer, A. (2000). *Realism and Social Science*. London: Thousand Oaks y Nueva Delhi: Sage.
- Sotomayor, A. (2008). Los métodos cualitativos en la ciencia política contemporánea: avances, agendas y retos. *Política y gobierno*, 15(1), 159-179.
- Ventura-León, J. L. y Barboza-Palomino, M. (2017). ¿Es posible generalizar en estudios cualitativos? *Ciência & Saúde Coletiva*, 22(1), 325.
- Wuisman, J. (2005). The logic of scientific discovery in critical realist social scientific research. *Journal of Critical Realism*, 4(2), 366-394.