

Un curioso y persistente caso de doble rasero en el análisis de conducta: Perspectivas de los especialistas en análisis de conducta sobre la fidelidad de procedimiento¹

Claire C. St. Peter¹ | Denys Brand² | Stephanie J. Jones |
Jennifer R. Wolgemuth³ | Lodi Lipien⁴

¹ Departamento de Psicología, West Virginia University, Morgantown, West Virginia, EEUU

² Departamento de Psicología, Universidad, Sacramento, California, EEUU

³ Departamento de Psicología, Salve Regina University, Newport, Rhode Island, EEUU

⁴ Departamento de Educación y Psicología Estudios, University of South Florida, Tampa, Florida, Estados Unidos

Correspondencia

Claire C. St. Peter, Department of Psychology, West Virginia University, 53 Campus Drive, Box 6040, Morgantown, WV 25606, USA.
Email: claire.stpeter@mail.wvu.edu

Editor jefe: John Borrero

Editor asociado: Linda LeBlanc

Resumen

La fidelidad de procedimiento es el grado en que las variables dependientes se aplican tal y como fueron diseñadas. A pesar de que se lleva tratando 40 años sobre la importancia de la fidelidad de procedimiento para la investigación conductual, referir los datos de fidelidad sigue siendo una práctica poco común en las revistas analítico-conductuales. Los investigadores han especulado sobre las razones de dicha infranotificación, pero aún no se han explorado las perspectivas de los estudiosos sobre cuándo es necesario reportar datos de fidelidad. Así pues, el propósito de este estudio es evaluar las posibles razones de la infrecuente comunicación de datos de fidelidad en la literatura analítico-conductual. Para abordar este objetivo, realizamos grupos de discusión con estudiosos del análisis aplicado de conducta. Surgieron cinco temas relacionados con las razones por las que habitualmente no se informa de los datos de fidelidad de procedimiento. Tratamos cómo se interrelacionan estos temas y ofrecemos sugerencias y recomendaciones para ayudar en la recopilación y comunicación de datos de fidelidad.

PALABRAS CLAVE

unknown printer took a galley of type

La fidelidad de procedimiento hace referencia al grado en que los procedimientos se aplican tal y como se han descrito o diseñado. En 1982, Peterson et al. señalaron la escasa frecuencia con la que se informaba de la fidelidad de la variable independiente en las publicaciones de *Journal of Applied Behavior Analysis (JABA)*. Incluso en ocasiones en las que se consideraba que el estudio tenía un "alto riesgo" de incurrir en errores de aplicación², rara vez se incluían datos de fidelidad. Peterson et al. identificaron varios problemas potenciales con la falta de medidas de fidelidad. Estos problemas incluían la

dificultad añadida en la identificación de relaciones funcionales, el aumento de la probabilidad de resultados irreplicables y la disminución del valor del trabajo analítico-conductual por parte de los de otras disciplinas. Los autores continuaron señalando la existencia de un "curioso doble rasero" (pág. 478) en las medidas adoptadas para garantizar la calidad de la medida de la variable dependiente en relación con la calidad de la aplicación de la variable independiente. Concluyeron advirtiendo a los analistas de conducta de "la necesidad

¹ Translated in May 2023 by Javier Virués-Ortega (The University of Auckland) with permission from SEAB and the authors. Original article: St Peter, C., Brand, D., Jones, S. H., Wolgemuth, J. R., & Lipien, L. (2023). On a persisting curious double standard in behavior analysis: Behavioral scholars' perspectives on procedural fidelity. *Journal of Applied Behavior Analysis*, 56(2), 336–351. <https://doi.org/10.1002/jaba.974>

² Peterson et al. (1982) definieron el "alto riesgo" como cualquier estudio excepto aquellos con "tratamientos definidos mecánicamente (p.ej., una máquina que entregaba una ficha cada vez que el participante pulsaba un botón) y productos permanentes (p.ej., el experimentador pintaba un cubo de basura con los colores del colegio)... intervenciones conductuales únicas (p.ej., el experimentador daba una serie de instrucciones o colocaba un cartel) o aplicación continua de la variable independiente (p.ej., cada vez que el sujeto entregaba un trozo de basura, recibía una ficha)" (pág. 484).

de evaluar la fidelidad... tanto de la variable independiente como la variable dependiente" (pág. 489).

A pesar de que Peterson et al. (1982) han sido citados casi 300 veces, muchos analistas de conducta han hecho caso omiso de su advertencia. Cuarenta años después de la publicación de Peterson et al., la mayoría de los estudios analítico-conductuales siguen sin incluir medidas de la fidelidad de la variable independiente (véase Falakfarsa et al., 2022 para una revisión). Esta omisión de los datos de fidelidad diferencia el análisis de conducta de algunas de nuestras ciencias afines (p.ej., la investigación educativa o la psicología escolar). En estas disciplinas relacionadas, las medidas de fidelidad se han convertido en estándar de calidad. Las medidas de fidelidad se especifican como necesarias en las normas de calidad para la investigación educativa (Ledford et al., 2020; National Autism Council, 2015; What Works Clearinghouse Standards Handbook Version 5.0, 2020) y pueden encontrarse en casi el 75% de los artículos publicados en psicología escolar (Sanetti et al., 2020).

La infrecuente comunicación de la fidelidad de la variable independiente persiste en la investigación analítico-conductual a pesar de las recomendaciones en sentido contrario y del conocimiento de las consecuencias potenciales de no informar la fidelidad. Dado que las primeras publicaciones sugerían la importancia de la fidelidad (p.ej., Peterson et al., 1982), ha habido repetidos llamamientos, incluso desde JABA, para que los autores informen de la fidelidad de la variable independiente (Gresham et al., 1993; McIntyre et al., 2007). Además, los analistas de conducta han demostrado empíricamente que una fidelidad reducida puede afectar negativamente los efectos y oscurecer las relaciones funcionales (véase Brand et al., 2019, para una revisión de algunas de estas demostraciones). Sin embargo, persiste la infranotificación de datos de fidelidad (véase Falakfarsa et al., 2022, y Han et al., 2022, para revisiones recientes).

Aunque los estudiosos han establecido claramente que los analistas de conducta no informan sobre los datos de fidelidad, sigue sin estar claro por qué no lo hacen. Se pueden hacer especulaciones basadas en encuestas a psicólogos clínicos o escolares, que sugieren que la terminología incoherente, la falta de normas claras sobre la comunicación de la fidelidad, la falta de conocimientos sobre cómo evaluar la fidelidad y el tiempo u otras limitaciones logísticas pueden ser barreras para la comunicación de los datos de fidelidad. Varias de estas barreras pueden aplicarse también a la investigación analítico-conductual. Por ejemplo, es concebible que existan barreras prácticas o logísticas para recopilar datos de fidelidad. Los observadores pueden ser escasos,

requerir mucho tiempo de entrenamiento o quizá se perciba como una imposición en la realización de la investigación (p.ej., cuando los estudios se realizan en el domicilio de los participantes en el contexto de servicios en curso y la presencia de otras personas o de una cámara de vídeo puede considerarse intrusiva).

Del mismo modo, las normas editoriales pueden reducir la probabilidad de que se comuniquen los datos de fidelidad, como sugieren McIntyre et al. (2007) y Sanetti et al. (2012), llegando Hagermoser et al. (2012) a identificarlas como un obstáculo. Sin embargo, JABA anima explícitamente a los autores a presentar datos de fidelidad. A partir de 2022, las directrices para autores en el sitio web de la revista señalaban que los autores debían incluir "una descripción del método utilizado para evaluar la fidelidad de la variable independiente (si procede)".

Otra posibilidad es que la recopilación de datos de fidelidad simplemente no forme parte de las expectativas del análisis aplicado de conducta como ciencia. Este sentido de las expectativas no se ha identificado como una barrera en investigaciones anteriores en psicología clínica o escolar (p.ej., Hagermoser Sanetti y DiGennaro Reed, 2012). Sin embargo, McIntyre et al. (2007) aludieron a esta posibilidad sugiriendo que los investigadores pueden no dar importancia a los datos de fidelidad, en particular cuando se observan cambios claros en la variable dependiente en la dirección esperada. La idea de que los datos de fidelidad podrían perder importancia cuando la conducta cambia en el sentido esperado no se ha evaluado en estudios anteriores. Tomar esta perspectiva constituiría un error lógico porque las amenazas a la validez interna del estudio (p.ej., los errores de fidelidad) también pueden ser responsables de que la conducta cambie como en la dirección esperada (Gresham et al., 2000; Kodak et al., 2022; Peterson et al., 1982). Por ejemplo, imaginemos un estudio que evalúa los efectos de un procedimiento de enseñanza sobre la adquisición de habilidades. El procedimiento no incluye formalmente ayudas, pero el investigador señala de forma intermitente y accidental cuál es la respuesta correcta. Cuando los participantes aprenden la habilidad, el investigador concluye que el procedimiento de enseñanza fue eficaz, pero tal conclusión es errónea porque las ayudas pueden haber facilitado el aprendizaje. Presumir que el cambio de conducta verifica la correcta aplicación de la variable independiente parecería contrario al enfoque inductivo que habitualmente adoptan los analistas de conducta y a las correspondientes estrategias de medida detalladas que se suelen adoptar (Johnston et al., 2020), pero puede

contribuir, no obstante, a la falta de informes de fidelidad en la literatura.

Así pues, existen varias explicaciones posibles para la persistente falta de datos de fidelidad en los estudios analítico-conductuales, pero ninguna de estas explicaciones ha sido identificada directamente como la responsable. Recopilar información sistemática de los académicos que publican con frecuencia estudios de análisis aplicado de conducta puede arrojar luz sobre las posibles explicaciones de las decisiones relacionadas con la comunicación de datos de fidelidad. En particular, puede ser importante explorar la tradición (es decir, la conducta gobernada por reglas) que puede estar influyendo en las decisiones de los académicos sobre cuándo y cómo recopilar datos de fidelidad y cómo informar de esos datos. Documentar las perspectivas de los académicos sobre el uso de los datos de fidelidad puede ayudar a proporcionar una imagen de las perspectivas actuales en el campo, lo que podría ser útil proporcionar un punto de partida para promover la toma y comunicación de datos de fidelidad. Por lo tanto, el propósito de la presente investigación ha sido explorar las perspectivas de los investigadores acerca de las razones por las que no se comunican datos de fidelidad en la literatura analítico conductual.

En algunos ámbitos (p.ej., la evaluación de la conducta), los analistas de conducta han reconocido la importancia de obtener respuestas a preguntas abiertas, sobre todo cuando se busca información matizada o detalles cuyos principales efectos se desconocen en gran medida (Fryling y Baires, 2016). Para proporcionar a los estudiosos flexibilidad para hablar subjetivamente sobre sus experiencias y perspectivas únicas y contextualizadas relacionadas con la medida y el informe de la fidelidad, hemos utilizado un enfoque cualitativo (p.ej., Barker & Pistrang, 2021). La investigación cualitativa pretende comprender los fenómenos naturales (p.ej., los fenómenos de medida y comunicación de la fidelidad) desde perspectivas en primera persona (p.ej., las respuestas reflexivas de los académicos analítico-conductuales a preguntas abiertas sobre la medida y comunicación de la fidelidad; Lichtman, 2013). Esperábamos que las razones de los académicos para medir e informar la fidelidad estuvieran fuertemente influidas por sus historias personales, pero que compartieran puntos en común dentro de la disciplina del análisis aplicado de conducta. Por lo tanto, utilizamos grupos de discusión, que son un método cualitativo de recopilación de datos que genera un diálogo grupal intersubjetivo para sacar a la superficie normas y prácticas culturales, sociales y educativas dinámicas (p.ej., Nerlich, 2004) en la disciplina relacionadas con la

fidelidad de procedimiento. Nuestros grupos de discusión fueron entrevistas semiestructuradas con varios académicos para permitir la interacción y, al mismo tiempo, aprovechar ideas y conversaciones (Lichtman, 2013).

METODOLOGÍA

Diseño

Para este estudio se seleccionó un diseño cualitativo básico (p.ej., Lichtman, 2013) porque el objetivo era evocar información detallada sobre las experiencias, la comprensión y las perspectivas de los participantes en relación con la fidelidad de procedimiento. Las decisiones sobre cómo medir e informar la fidelidad son matizadas y complejas, y era importante utilizar un enfoque que nos permitiera comprender por qué y cómo los académicos deciden informar estos datos. Los diseños básicos de investigación cualitativa son aquellos que no están claramente categorizados dentro de metodologías cualitativas bien definidas (p.ej., teoría fundamentada, estudio de casos, fenomenología) y, como tales, son más adaptables y flexibles (Kahlke, 2014). El uso de un diseño cualitativo básico nos dio la flexibilidad necesaria para seleccionar los métodos y elementos de diferentes enfoques cualitativos que mejor se adaptaban a la respuesta a nuestras preguntas de investigación, incluidos los grupos de discusión, los procedimientos de la investigación cualitativa consensuada (Hill & Knox, 2021) y un enfoque de codificación para el análisis temático.

Las limitaciones de los diseños cualitativos básicos se asocian en gran medida a su potencial de ser atóricos tanto en lo que respecta a estar guiados por un marco teórico o conceptual como a alinear los métodos bajo una teoría explícita de investigación (Kahlke, 2014). Para garantizar el rigor teórico, desarrollamos nuestras preguntas de investigación, el muestreo intencionado, los grupos de discusión y el enfoque de codificación para el análisis temático basándonos en nuestros antecedentes teóricos skinnerianos y nuestros amplios historiales con la erudición analítico-conductual existente sobre la fidelidad. Los métodos también se alinearon explícitamente bajo un enfoque cualitativo interpretativista que enfatiza la experiencia subjetiva e intersubjetiva y la creación de significado (Schwartz-Shea & Yanow, 2013). Como estudio cualitativo

explícitamente interpretivista³, las consideraciones metodológicas que se asocian habitualmente a la investigación analítico-conductual, como la generalizabilidad, la objetividad y la fiabilidad de las medidas (es decir, el acuerdo entre codificadores en la investigación cualitativa), no estaban alineadas con nuestro diseño y, por lo tanto, no eran aplicables a él (p.ej., Braun y Clarke, 2013).

Participantes

Para identificar a los posibles participantes, utilizamos una lista que contenía los nombres de los académicos que habían publicado al menos un artículo en JABA como autor de correspondencia entre 2006 y 2020 ($N = 435$) y la frecuencia con la que esas personas informaban de los datos de fidelidad en sus publicaciones en JABA. La lista se creó como parte de otro proyecto que examinó la frecuencia de los informes de fidelidad en JABA (Bergmann et al., 2022). Cruzamos esta lista con los individuos que actualmente formaban parte del consejo editorial de JABA o que trabajaban o habían trabajado como editores asociados o editores jefe para JABA u otras revistas importantes del área y añadimos información sobre su cargo actual (título y ubicación geográfica) y sus principales áreas de interés (p.ej., problemas de conducta, adquisición de habilidades, alimentación). Clasificamos esta lista por la frecuencia con la que el individuo actuaba como autor de correspondencia y la frecuencia con la que informaba de los datos de fidelidad en esos estudios. Utilizamos esta lista para identificar a los individuos que variaban en función de la antigüedad en el campo, la frecuencia de publicación como autor de correspondencia en JABA, la frecuencia con la que informaban sobre fidelidad, la experiencia editorial en JABA, los temas y los entornos de investigación habituales y la ubicación geográfica. Cualquiera que apareciera en la lista podía ser incluido; no limitamos la participación a personas cuya afiliación principal fuera una universidad. Esta estrategia intencionada nos permitió generar una muestra rica en información, homogénea con respecto a haber publicado en trabajos analítico-conductuales en JABA, pero

heterogénea en factores que esperábamos que produjeran una explicación más matizada de la medida y comunicación de datos de fidelidad (Malterud et al., 2021).

Para reducir la probabilidad de que los individuos se abstuvieran de compartir sus experiencias y perspectivas en los grupos de discusión debido a las diferencias de poder percibidas, creamos grupos homogéneos en términos de antigüedad y experiencia editorial, pero diferían en su frecuencia de publicación en JABA, el grado en que mencionaban la fidelidad en sus publicaciones en JABA, sus áreas de investigación, el entorno de su lugar de trabajo (p.ej., universidad, centro médico, clínica) y su ubicación geográfica (Kruger y Casey, 2014). Esta disposición dio lugar a siete grupos de 170 posibles participantes. Inicialmente invitamos a tres o cuatro personas por grupo. Cuando un participante declinaba, invitábamos a un suplente cuyo perfil fuera similar al de la persona que declinaba siempre que fuera posible. En total, invitamos a participar a 30 académicos; de ellos, 25 (83%) dieron su consentimiento informado y participaron. No realizamos grupos de discusión adicionales más allá de los siete iniciales, ya que alcanzamos la saturación, el punto en el que los datos adicionales no aportan nueva información relevante, en los cinco primeros grupos de discusión. En una revisión sistemática reciente, Hennick y Kaiser (2022) indicaban que la saturación en estudios cualitativos se alcanza habitualmente usando entre 4 y 7 grupos de discusión. Los grupos de discusión con un tema de discusión restringido y una muestra homogénea, como sucedía en nuestro caso, tenían más probabilidades de alcanzar la saturación con 4 ó 5 grupos de discusión (Hennick y Kaiser, 2022).

La información sobre la composición de cada grupo aparece en la Tabla 1. Recordemos que los grupos se crearon con una variación intencionada en las tendencias de los individuos a mencionar la fidelidad en sus publicaciones de JABA para muestrear perspectivas potencialmente diversas. Así, cada grupo contenía al menos un académico que mencionaba con frecuencia la fidelidad en sus publicaciones (50%-100% de sus publicaciones) y un académico que mencionaba la fidelidad con menos frecuencia (0%-33% de sus

³ Los criterios de validez apropiados para la investigación cualitativa interpretivista suelen tratarse bajo términos como credibilidad, fiabilidad y confiabilidad. Tracy (2010) ofrece ocho criterios de validez para valorar la calidad de toda investigación cualitativa: adecuación del tema, rigor, sinceridad, credibilidad, resonancia, contribución significativa, ética y coherencia significativa. Como toda investigación cualitativa, estos criterios son subjetivos (o intersubjetivos). Corresponde tanto a los investigadores diseñar y describir sus estudios a la luz de estos criterios como a los lectores aplicarlos teniendo en cuenta sus propias experiencias,

conocimientos y perspectivas. Esta evaluación subjetiva puede resultar más evidente en el caso de la resonancia. Los estudios con una alta resonancia comunican los procedimientos de estudio y los resultados con suficiente riqueza como para que los lectores puedan tomar decisiones informadas sobre el grado en que los hallazgos se alinean con sus propias experiencias y perspectivas y podrían transferirse a su propio contexto. Curiosamente, y quizá irónicamente, nuestro enfoque cualitativo interpretivista significa que nuestra área temática (la medida de la fidelidad de procedimiento) es una consideración metodológica irrelevante para este estudio.

TABLA 1 Descripción de los participantes.

Grupo	Núm. de académicos	Núm. publicaciones como autor de correspondencia (2006-2020)	Núm. total y % de publicaciones que presentan datos de fidelidad	Rango del académico
Investigador sobre fidelidad	4	50	39 (78%) 6%–100%	1 titular, 3 asociados
Editor jefe	3	44	29 (70%) 62%–83%	3 titulares
Editor asociado sénior	4	22	19 (86%) 0%–100%	3 titulares, 1 otro
Editor asociado a mitad de carrera	4	22	10 (45%) 0%–100%	4 asociados
Académico a mitad de su carrera	4	18	10 (56%) 33%–100%	4 asociados
Editor asociado al comienzo de su carrera	3	16	7 (44%) 33%–50%	4 asociados
Académico al comienzo de su carrera	3	7	5 (71%) 0%–100%	3 contratados, 1 otro
Total	25	179	119 (66%) 0%–100%	7 titular 11 asociados 5 contratados 2 otros

publicaciones)⁴. Cabe destacar que aquellos académicos que efectivamente reportaban fidelidad en sus estudios, lo hacían con una frecuencia elevada (en el 65%-100% de sus publicaciones) haciendo que la frecuencia media con la que los participantes reportaban fidelidad en sus estudios fuese superior a la media general de JABA. Esto puede obedecer a tres factores: (a) la probabilidad general de incluir datos sobre la fidelidad en las publicaciones JABA sigue siendo inferior al 40% (véase Han et al., 2022 para una revisión reciente), (b) incluimos deliberadamente a personas que tenían diversos rangos de publicación de datos de fidelidad, incluidos aquellos que medían e informaban con frecuencia sobre la fidelidad, y (c) el tamaño de nuestros grupos era pequeño.

Para reducir la probabilidad de que un participante pudiera dar respuestas percibidas como socialmente deseables a nuestras preguntas y para aumentar la comodidad a la hora de compartirlas, el miembro más veterano del equipo de investigación, cuyo programa de investigación se centra en la fidelidad, sólo dirigió grupos

de discusión que incluían a sus compañeros (otros académicos veteranos, incluidos editores jefe y editores asociados veteranos). Los miembros de nuestro equipo que se encontraban en las fases intermedias de su carrera dirigieron los grupos formados por participantes que se encontraban en un estadio similar (p.ej., profesor asociado) o en la primera etapa de su carrera (p.ej., profesor adjunto).

Grupos de discusión

Realizamos los grupos de discusión a través de internet utilizando la plataforma Zoom. Cada grupo de discusión incluyó a tres o cuatro expertos, y cada experto sólo asistió a un grupo de discusión. Aunque los grupos de discusión presenciales suelen estar formados por entre seis y ocho participantes, se recomienda un número menor para los grupos de discusión en línea a fin de garantizar que todos los participantes se impliquen y tengan oportunidades de compartir sus puntos de vista

⁴ Estos puntos de corte se seleccionaron basándose en los datos de Han et al. (2022), que sugerían que una media del 38% de los estudios JABA

publicados entre 2006 y 2020 (el periodo muestreado para su inclusión en los grupos de discusión) incluían datos de fidelidad.

(Kite & Phongsavan, 2017). Los grupos de discusión se programaron con una duración de 90 minutos en días y horas en las que los participantes estaban disponibles. Los grupos de discusión duraron entre 57 y 92 minutos (Media = 77 min; Mediana = 81 min); la variabilidad fue debida en gran medida a la atención a las preguntas de discusión. Un experto no pudo asistir a su grupo de discusión programado debido a un compromiso de última hora; en su lugar, participó en una entrevista individual. La entrevista individual duró 35 min.

Los grupos fueron moderados por uno de los investigadores con un segundo investigador de apoyo especialmente en caso de problemas técnicos inesperados (aunque estos no sucedieron). Los grupos de discusión estaban semiestructurados para permitir un diálogo dinámico y la exploración de los temas en profundidad, al tiempo que se garantizaba que los debates no se desviasen del tema de interés (Kruger & Casey, 2014). El investigador formuló un conjunto predeterminado de preguntas (véase la información suplementaria A), pero a veces cambió el orden de estas o hizo preguntas de seguimiento acorde al diálogo que estaba teniendo lugar. Grabamos los grupos de discusión utilizando Zoom e hicimos una grabación de audio paralela en caso de que hubiese algún problema con el vídeo (lo que no ocurrió).

Codificación cualitativa

Utilizamos un servicio profesional (GoTranscript) para crear transcripciones de las grabaciones. Una vez realizadas las transcripciones, las revisamos mientras escuchábamos la grabación para resolver cualquier error de transcripción, incluidos términos técnicos u otras palabras ininteligibles para el servicio de transcripción (p.ej., "JABA"). Utilizamos las transcripciones revisadas para todos los análisis de datos posteriores.

Nuestro análisis siguió los procedimientos de codificación deductiva e inductiva recomendados para el análisis temático en psicología (p.ej., Braun y Clarke, 2006; Ryan y Bernard, 2003). El análisis temático permite a los investigadores buscar temas importantes que surgen para describir o explicar un fenómeno (Daly et al., 1997). Los procedimientos deductivos incluyen la clasificación de los datos en categorías para ajustarlas a las preguntas de la investigación. Los procedimientos inductivos se utilizan para dar sentido a los datos aplicando códigos y buscando patrones para desarrollar temas (Miles et al., 2020; Sandelowski et al., 2007). Nuestro enfoque general del análisis fue consensuado, derivado de los enfoques basados en el trabajo en equipo para llegar a un consenso sobre los hallazgos y las interpretaciones que se asocian a diseños bien

establecidos en psicología, como la investigación cualitativa consensuada (Hill & Knox, 2021). Específicamente, la investigación cualitativa consensuada consiste en tener preguntas abiertas para recopilar datos a través de los individuos, contar con múltiples investigadores en el proceso de análisis para garantizar una diversidad de perspectivas, utilizar el consenso para llegar a juicios sobre el significado de los datos, incluir a un miembro externo para minimizar el pensamiento de grupo entre el equipo de investigación e incorporar dominios e ideas centrales en el análisis (Hill et al., 1997).

Antes de codificar las transcripciones, elaboramos un repertorio de códigos inicial basado en las evaluaciones existentes sobre la comunicación de la fidelidad, incluidas las especulaciones previamente establecidas sobre los motivos por los que los académicos pueden no notificar la fidelidad en sus publicaciones. Por ejemplo, nuestro repertorio de códigos inicial incluía el código "comparación/contraste con el acuerdo entre observadores", que derivamos de la comparación ya sugerida por Peterson et al. (1982) entre el acuerdo entre observadores (AEO) y la medida de la fidelidad, así como códigos para identificar las barreras propuestas por Hagermoser et al. (2012), tales como los recursos, el entrenamiento y las normas o contingencias externas. También incluimos códigos basados en nuestras experiencias como académicos que han realizado amplias investigaciones sobre la fidelidad. Por ejemplo, incluimos el código "revisión por pares" para capturar la conversación sobre el papel del proceso editorial en la inclusión de datos de fidelidad.

Una vez elaborado el repertorio de códigos inicial, nos reunimos en equipo para revisar los posibles códigos y crear definiciones de cada uno de ellos. Por ejemplo, el código "revisión por pares" se definió como cualquier mención al proceso editorial, incluidos los comentarios del revisor/editor o la normativa de la revista. Pusimos a prueba el libro de códigos evaluando individualmente partes de una transcripción para determinar si era necesario perfeccionar los códigos. Llegamos a un consenso sobre los códigos y las definiciones de los códigos y el perfeccionamiento de los códigos se produjo durante una reunión de grupo.

Cargamos las transcripciones finales en una plataforma de software de codificación cualitativa (Dedoose, 2021) para codificar las transcripciones. Los tres miembros del equipo de investigación con amplia experiencia en análisis de conducta y fidelidad de procedimiento aplicaron los códigos a la misma transcripción y, a continuación, se reunieron para revisar las posibles diferencias en la interpretación de los

códigos. En esta reunión, los códigos se refinaron y aclararon aún más y se asignaron dos miembros del equipo a cada transcripción. Utilizando el libro de códigos, codificaron individualmente las transcripciones (incluida la recodificación de la primera transcripción codificada) y luego se reunieron por parejas para comparar los códigos y llegar a un consenso sobre los códigos finales aplicados a cada transcripción. En la documentación complementaria B figura una tabla que muestra los códigos iniciales y cómo evolucionaron a lo largo del proceso de codificación.

Una vez finalizada la codificación de los siete grupos de discusión, las citas correspondientes a los códigos de las transcripciones se exportaron a Excel. Tres miembros del equipo revisaron las citas en el contexto de los códigos y redactaron notas interpretativas para establecer conexiones conceptuales entre los códigos, identificar categorías más amplias y capturar posibles significados (Birks et al., 2008). Los códigos, las citas y los memos se trataron en reuniones de equipo que permitieron llegar a un consenso sobre un conjunto final de temas.

Papel de los investigadores

Las historias y perspectivas de los investigadores se consideran parte integrante de la interpretación de los resultados cualitativos (Berger, 2015; Darwin Holmes, 2020), especialmente en los enfoques basados en la interpretación. A tal fin, los lectores deben saber que nuestro equipo de investigación incluía a tres personas que consideran las evaluaciones de la fidelidad como una de sus principales líneas de investigación. Nuestras amplias experiencias reflexionando sobre la fidelidad y trabajando sobre cuestiones de medida e información de la fidelidad moldearon sin duda las preguntas que formulamos durante los grupos de discusión e informaron nuestras interpretaciones individuales y consensuadas de los correspondientes códigos y temas. Nuestro equipo también incluía a dos expertos en métodos mixtos y metodología cualitativa que recurrieron a su experiencia para orientar y apoyar el proceso de diseño y análisis. También actuaron como amigos críticos durante el proceso de recopilación y análisis de datos que ayudaron al equipo a ir más allá de sus ideas preconcebidas formulando preguntas, señalando supuestos y ofreciendo conexiones e interpretaciones únicas desde sus perspectivas sobre la validez en la investigación social y del comportamiento (Blanco & Rossman, 2022).

En la investigación cualitativa, las historias de los investigadores suelen considerarse fortalezas necesarias que proporcionan la base de conocimiento para una

interpretación rigurosa, frente a diversos sesgos (Roulston & Shelton, 2015). Aprovechando sus historias personales, los investigadores cualitativos se implican en la reflexión individual o grupal a lo largo del proceso analítico, examinando al mismo tiempo sus supuestos y manteniendo una postura abierta con respecto a sus compromisos interpretativos (Creswell & Poth, 2018). Nuestro enfoque analítico consensuado permitió estas actividades, ya que compartimos y cuestionamos mutuamente nuestras perspectivas para llegar a un consenso sobre los temas que tratamos a continuación.

RESULTADOS

Identificamos los cinco temas siguientes en relación con la importancia (o la falta de importancia) de informar sobre los datos de fidelidad: (a) historia del campo, (b) contingencias débiles o inexistentes, (c) confianza en la capacidad de quien aplica la estrategia de fidelidad, (d) valor relativo de los datos de fidelidad y (e) falta de conocimientos o recursos. A continuación, se analiza en detalle cada uno de los cinco temas. Para cada tema, ofrecemos una visión general de los resultados y un breve tratamiento de las posibles interpretaciones antes de vincular los temas en la sección de Discusión.

Obsérvese que los académicos participantes utilizaron diversos términos para referirse a la fidelidad, como fidelidad de tratamiento, fidelidad de procedimiento, integridad de tratamiento e integridad de procedimiento. El uso de estos diversos términos se ha destacado anteriormente como un posible obstáculo para la medida y la comunicación de la fidelidad (p.ej., Hagermoser Sanetti y DiGennaro Reed, 2012). Por lo tanto, preguntamos a cada grupo de discusión cómo utilizaban los términos. En general, los especialistas consideraron que la fidelidad de procedimiento era un término con un campo semántico más amplio, pero habitualmente utilizaron los demás términos referidos como sinónimos. En las citas que figuran a continuación, mantenemos los términos que utilizaron los académicos, pero nos referimos a la *fidelidad de tratamiento* en nuestras descripciones independientemente del término seleccionado por el académico.

Tema 1: Historia del campo

El análisis aplicado de conducta surgió como una rama del análisis experimental de la conducta cuando individuos entrenados en el enfoque operante tradicional (de laboratorio) empezaron a aplicar dicho paradigma a problemas de importancia social. Así, muchos (pero no todos) de los primeros pioneros del análisis aplicado de

conducta tenían sus bases en contextos de laboratorio altamente controlados (p.ej., la investigación con animales en cajas operantes). En esos contextos, informar sobre la fidelidad de la variable independiente probablemente parecía innecesario porque los procedimientos estaban mecanizados o automatizados de otro modo.

Algunos estudiosos señalaron que esta historia de automatización de los procedimientos, y el bajo riesgo asociado de una escasa fidelidad, puede haberse trasladado una vez que los analistas de conducta empezaron a realizar estudios en los que los humanos servían como ejecutores. Por ejemplo, uno de los expertos en la fase inicial de su carrera comentó,

Pienso en el sector experimental del campo.... Tienes una caja de Skinner, y ves presiones de palanca, ¿qué hay en términos de fidelidad del tratamiento? ¿Queremos que alguien se sienta y calcule el programa de RV [razón variable]? ... Una parte de mí piensa que [la falta de interés en la fidelidad] tiene un origen histórico.

Otros estudiosos caracterizaron estas variables históricas como "tradiciones" en el campo. Por ejemplo, un experto a mitad de su carrera señaló que "la gente tradicionalmente no ha comunicado datos de fidelidad". Otro académico a mitad de su carrera sugirió: "Hay un peso de la tradición sobre decisiones que parecen arbitrarias. ... Así es como me siento yo también cuando no mido la fidelidad del tratamiento en estos estudios más complejos. Siento que en mi caso es solo por tradición". Más adelante en este mismo intercambio, otro académico a mitad de carrera señaló: "Creo que [la falta de informes de fidelidad] empezó a darse y ha permaneció; creo que es una situación que cambiará cuando algunas personas vengan y hablen de ello lo suficiente como para que consigan que un montón de gente se suba al carro."

Este énfasis en la tradición y la historia puede resultar desconcertante a la luz de la historia de tratar la importancia de las medidas de fidelidad. Como ya se ha señalado, los artículos sobre la importancia de incluir datos de fidelidad empezaron a ser ampliamente reconocidos en 1982, 40 años antes de la redacción del presente artículo. Incluso antes, científicos muy respetados habían empezado a señalar que los analistas de conducta aplicados estaban empezando a utilizar procedimientos de tratamiento complejos que podían ocultar hasta qué punto la supuesta variable independiente afectaba a los datos obtenidos (Michael, 1980). Por lo tanto, aunque los estudiosos (en particular los que se encontraban al principio o a mitad de su

carrera) a veces señalaban la historia o la tradición como motivo para no informar la fidelidad, parece probable que variables más contemporáneas también influyan en la conducta de los científicos.

Tema 2: Contingencias débiles o inexistentes

Una variable contemporánea señalada por varios de los estudiosos fue la relativa falta de contingencias para informar sobre la fidelidad en los manuscritos que se envían a las revistas analítico-conductuales. Se han identificado como barreras contingencias débiles similares para incluir datos de fidelidad en la investigación en psicología escolar (Hagermoser Sanetti y DiGennaro Reed, 2012), pero esas barreras se centraban principalmente en contingencias editoriales. Por el contrario, los estudiosos del presente estudio señalaron tres fuentes potenciales de contingencias: contingencias editoriales, contingencias de los proyectos financiados y contingencias de inclusión en revisiones o metaanálisis. En general, los académicos parecían percibir que la fuerza de las contingencias para informar sobre la fidelidad aumentaba en esas tres categorías, siendo las contingencias editoriales las más débiles y las contingencias metaanalíticas las más fuertes.

Los académicos de todos los niveles señalaron haber recibido comentarios poco frecuentes a través del proceso editorial sobre la necesidad de datos de fidelidad, a pesar de la impresión de que se requerirían datos paralelos para la variable dependiente (es decir, el AEO) (véase Essig et al., 2023, para un comentario reciente sobre la relación entre el AEO y los datos de fidelidad). Por ejemplo, un académico a mitad de carrera señaló: "Son las contingencias existentes las que reforzarán mi conducta de recopilar datos de AEO, con más regularidad que intentar hacer malabarismos y tener una medida de fidelidad decente al mismo tiempo".

La mayoría de los editores asociados y editores jefe confirmaron que los datos de fidelidad no eran fundamentales para las decisiones que tomaban sobre los manuscritos. La mayoría de los editores declararon que no rechazarían un artículo basándose únicamente en la falta de datos sobre la fidelidad, pero muchos sí señalaron que había circunstancias en las que imponían requisitos editoriales sobre la fidelidad (p.ej., cuando personas que no tienen un entrenamiento sólido en análisis de conducta actúan como encargados de la aplicación). Por ejemplo, un editor comentó,

Para mí, creo que realmente dependería del estudio. ... No veo que lo rechazara sólo por eso, pero si hubiera algún otro aspecto extraño que

realmente requiriera conocer la fidelidad para entenderlo, entonces probablemente lo haría.

Del mismo modo, un académico a mitad de carrera señaló,

Si se trata de una persona no experta, novata, si no dispone de datos sobre la fidelidad del tratamiento, si no puede hacérmelos llegar [los datos sobre la fidelidad] o si no tiene forma de hacerlo, entonces rechazaría, independientemente de lo que muestren los datos.

Las contingencias editoriales pueden variar en función del ámbito de investigación, con diferentes contingencias adoptadas por diferentes comunidades dentro del paraguas más amplio de un enfoque analítico-conductual. Los estudiosos que realizaban principalmente diseños de grupo que se publican en revistas médicas informaron de que recogían datos de fidelidad con poca frecuencia. Por el contrario, los que trabajaban en educación especial sugirieron que se esperaban datos de fidelidad para su publicación.

La adopción de estándares de fidelidad por parte de otros campos (p.ej., la educación especial) puede estar afectando a los analistas de conducta que solicitan subvenciones o que están preocupados por la inclusión de su investigación en futuros meta-análisis. Esto puede haber tenido como consecuencia que algunos académicos adopten las normas de fidelidad que especifican las instituciones que conceden proyectos (e.g., Institute of Education Science). Un editor sugirió: "Creo que mi conducta ha cambiado en lo que respecta a la fidelidad de procedimiento, sobre todo por los comentarios de los revisores y... ciertamente para la investigación financiada con fondos públicos y privados. Tienes que tener esos [datos de fidelidad] para las subvenciones". Sin embargo, no estaba claro si la adopción de estas normas se extendía más allá de la investigación financiada con subvenciones.

Los académicos también parecían sensibles a la medida en que los datos de fidelidad eran necesarios para que su investigación llegara a un público amplio o se incluyera en metaanálisis o revisiones. Por ejemplo, un editor asociado senior señaló,

Yo trabajo en educación especial, además de en análisis de conducta, y estamos muy atentos a que nuestro material se incluya en las revisiones sistemáticas y a que se cumplan los nuevos estándares para la investigación de diseño de caso único... como los estándares del What Works Clearinghouse, que incluyen la fidelidad

de procedimiento. Si no cumples esas normas, tu material no se incluye en las revisiones sistemáticas. ... Si quiere que su material se incluya, debes cumplir las normas.

Un editor se hizo eco de la misma percepción: "Ahora, cuando hay metaanálisis y revisiones de diseños de caso único, con mucha frecuencia nos ponen pegajos por no haber informado sobre la fidelidad de procedimiento... con lo que tus estudios pasan de clasificarse como empíricamente sólidos a cuestionables, básicamente".

Así, los estudiosos identificaron contingencias diferenciadas: las revistas analítico-conductuales no suelen exigir datos de fidelidad, pero sí lo hacen las instituciones que conceden subvenciones y otras disciplinas. Los individuos que valoran influir en campos ajenos a las revistas analítico-conductuales pueden ser más propensos a comunicar datos de fidelidad. Esta posible desconexión merece ser explorada. Hasta la fecha, no conocemos ningún estudio que compare directamente la comunicación de la fidelidad entre disciplinas. Sin embargo, estimaciones recientes de la prevalencia de los datos de fidelidad en los artículos publicados sugieren que dichos datos se reportaron en aproximadamente el 70% de los estudios de investigación sobre intervención escolar publicados entre 2005 y 2012 (p.ej., Barnett et al., 2014), pero sólo en el 46% de los estudios analítico-conductuales aplicados en el periodo 2008 a 2019 (Falakfarsa et al., 2022).

Una opinión común fue que fomentar la presentación de informes en revistas analítico-conductuales daría lugar a que más investigadores incluyeran medidas de fidelidad. Algunos estudiosos sugirieron que tales contingencias no sólo serían útiles, sino necesarias. Por ejemplo, un investigador experto en fidelidad comentó: "sólo tenemos que exigirlo y entonces la conducta [de los investigadores] igualará las contingencias si no pueden publicar". De este sentimiento se hizo eco un académico en el inicio de su carrera que sugirió (con respecto a la pregunta de qué podríamos hacer como disciplina para fomentar la presentación de datos de fidelidad):

[Deberíamos ofrecer] más informes, creo que todo tiene que ver con las contingencias. Todo gira en torno a lo que se requiere para publicar o cuáles son las normas y luego hacer que los autores se atengan a esas normas y, si ello fuera así, estoy seguro de que nos alinearemos a tales normas muy rápidamente.

Si el análisis de conducta va a la zaga de otras disciplinas en la comunicación de la fidelidad, lo que

puede tener consecuencias negativas para la ciencia (p.ej., una menor probabilidad de subvenciones y de inclusión en metaanálisis), cabe preguntarse por qué no hemos instituido ya contingencias para la comunicación de los datos de fidelidad. En ocasiones, los académicos se mostraron reticentes a la obligatoriedad de los datos de fidelidad para su publicación, señalando que los mandatos pueden limitar la publicación de estudios novedosos, dar lugar a la inclusión de datos de fidelidad de baja calidad o alargar innecesariamente los manuscritos. Por ejemplo, un editor asociado a mitad de carrera señaló: "Si se imponen requisitos, sólo se conseguirá superficialidad. Sería una casa de locos, para ser sincero, y odio ser un AE [editor asociado] o revisor y ver estas páginas horriblemente largas llenas de información innecesaria." Otros sugirieron que revistas como JABA podrían adoptar normas más rigurosas sin cusar una respuesta negativa. Por ejemplo, un editor asociado senior señaló,

No creo que fuera un obstáculo si la información [sobre la obligatoriedad de los datos de fidelidad] se comunicara con claridad. Siempre habrá otras revistas. Se basa en las tres principales revistas que deberían ser un estándar. Siempre habrá otros lugares donde difundir esa información.

Quizá debido a estas diferencias de opinión, parece poco probable que las normas editoriales se adopten pronto de manera uniforme en las revistas analítico-conductuales. Como dijo un editor:

No creo que vaya a conseguir que publiquen su estudio en JABA si no tiene ningún AEO, y podría conseguir que lo publicaran si no tiene datos sobre la fidelidad de procedimiento. Diría que creo que está mal de esa manera intelectual en la que pensamos que las cosas están mal pero no necesariamente tomamos medidas a fondo para aplicar consecuencias para que no vuelva a ocurrir.

Por lo tanto, las contingencias sobre la comunicación de la fidelidad pueden producirse principalmente a partir de fuerzas externas (p.ej., subvenciones, revistas de otras disciplinas) en lugar de a partir de revistas analítico-conductuales. Puede que sean necesarias fuentes de motivación más allá de las actuales contingencias editoriales para informar sobre los datos de fidelidad.

Tema 3: Confianza en las capacidades del que aplica la estrategia de fidelidad

Algunos académicos señalaron que la medida de la fidelidad pierde valor si quienes la aplican carecen de un entrenamiento sofisticado o una amplia experiencia. Un académico al principio de su carrera señaló la amplia confianza histórica en el trabajo de eruditos conocidos, diciendo,

En nuestro pequeño campo, en gran parte de lo que hacemos, simplemente confiamos los unos en los otros y damos por sentado que todo el mundo hizo lo que dijo que iba a hacer. Hasta hace poco, todo ha sido a ciegas. Ves quién escribió el artículo y fue [nombre omitido] y uno de sus estudiantes, simplemente confías, lo hicieron bien, [pero] esa no es la forma en que operan la mayoría de las ciencias.

Un experto a mitad de carrera se hizo eco: "Si me dices que la medida de fidelidad fue aplicada por un analista de conducta certificado, no me preocuparé tanto como si me dices que lo aplicó un estudiante universitario". En algunos casos, esta confianza se extendía al propio trabajo del individuo. Por ejemplo, otro experto a mitad de carrera señaló: "No quiero que suene así, pero yo llevaba ocho años haciendo la intervención y me había entrenado con la persona que la creó."

La confianza en los encargados de aplicar los programas se extendía a veces a las instituciones en las que trabajaban. A veces se consideraba que los estudios realizados en centros conocidos, o que en general se consideraba que disponían de recursos considerables, no necesitaban tantos datos sobre la fidelidad. Por ejemplo, un experto señaló: "trabajando en los laboratorios de [nombres omitidos], no creía que la fidelidad de procedimiento fuera tan importante. Sigo pensando que un lugar realmente bueno va a cuidar siempre de su variable independiente".

Esta confianza no siempre se extendía a los estudiantes de postgrado o a los expertos que iniciaban su carrera. Por ejemplo, un experto a mitad de carrera señaló: "[la intervención] puede ser un poco más compleja, pero la dirige un estudiante de doctorado. Esos son los momentos en los que tengo que pensar [si medir la fidelidad]". Un experto a principios de su carrera señaló la falta de confianza en la aplicación de los procedimientos por parte de sus compañeros, afirmando: "Cuando tienes investigadores prometedores, puede ser importante asegurarse de que hay un cierto nivel de

supervisión que puede no ser necesario para la gente que lleva tiempo haciendo esto."

Casi uniformemente, los estudiosos creían que era probable que la medida de la fidelidad fuera importante cuando individuos sin entrenamiento formal en el análisis de conducta aplicaban la estrategia de fidelidad. Para algunos participantes, la necesidad de datos de fidelidad cuando los padres eran los encargados de aplicar las medidas parecía una conclusión inevitable: "Obviamente, si estás investigando con padres, quieres saber lo que están haciendo". Un editor asociado señaló: "Si los padres estuvieran aplicando [la intervención], tengo que ver esos [datos de fidelidad]. Tengo que ver los datos de los padres, o no aceptaré el estudio".

Hasta donde sabemos, no existen datos que respalden la confianza (o falta de confianza) en las conclusiones extraídas de estudios analítico-conductuales basados en el título obtenido, el entrenamiento previo o la institución del investigador. Está claro que las personas que han recibido formación basada en el desempeño en un procedimiento dado pueden aplicarlo con mayor fidelidad que los individuos que no han recibido dicha formación (p.ej., Brand et al., 2018; Breeman et al., 2020; Carroll et al., 2013); una gran cantidad de literatura sobre el entrenamiento en habilidades conductuales respalda esta afirmación (DiGennaro Reed et al., 2018; Parsons et al., 2012, 2013). Sin embargo, las personas que han recibido entrenamiento, tienen experiencia o cuentan con credenciales avanzadas a veces aplican intervenciones con una fidelidad relativamente baja. Por ejemplo, Brand et al. (2018) evaluaron la fidelidad de los procedimientos de entrenamiento con ensayos discretos realizados en el ambiente natural. Tres terapeutas participaron en el estudio, incluido un analista de conducta certificado con 9,5 años de experiencia (otro terapeuta tenía 9 años de experiencia, pero no era analista de conducta certificado). Los terapeutas (a pesar de su experiencia y credenciales) cometieron errores relacionados con las indicaciones de respuesta, la provisión de reforzamiento tras las respuestas incorrectas de los alumnos y los procedimientos de corrección de errores, aplicando estos últimos con una fidelidad del 18% al 55%. Claramente, la experiencia por sí sola parece insuficiente para garantizar una alta fidelidad.

En resumen, aunque las personas con un entrenamiento previo en un procedimiento pueden aplicarlo con niveles más altos de fidelidad que las que carecen de dicho entrenamiento, aquellos seguirán cometiendo errores. Por lo tanto, independientemente del entrenamiento, el historial o el lugar de empleo de quien aplique la intervención, los investigadores deben supervisar la fidelidad para identificar los casos que

podrían desviarse de la aplicación canónica de la intervención (Hansford et al., 2010).

Tema 4: Devaluación de los datos de fidelidad

Hubo un desacuerdo considerable, con firmes opiniones contrapuestas, sobre el valor de recopilar datos sobre la fidelidad. Estos resultados contrastan fuertemente con los de una encuesta realizada a psicólogos escolares (Hagermoser Sanetti & DiGennaro Reed, 2012). En el estudio anterior, los cuatro ítems relacionados con la "falta de apreciación de la fidelidad del tratamiento" (pág. 198) se clasificaron entre los que menos obstáculos suponían para medir o informar sobre la fidelidad. La diferencia podría reflejar la selección intencionada de académicos que era probable o improbable que mencionaran la fidelidad en su propio trabajo en el estudio actual, pero el entusiasmo de los académicos por la recopilación de datos sobre fidelidad no se correspondía claramente con la frecuencia con la que mencionaban la fidelidad en sus propias investigaciones. Por otra parte, podría sugerir factores culturales (p.ej., prácticas aceptadas, historiales de entrenamiento, modelos que se han recibido) que pueden contribuir a una menor comunicación de la fidelidad en el análisis de conducta.

Las opiniones iban desde que los datos de fidelidad eran una necesidad para todas las investigaciones hasta que los datos de fidelidad rara vez eran necesarios. Los académicos que pensaban que las medidas de fidelidad eran necesarias a menudo expresaron su preocupación por la menor probabilidad de obtener resultados reproducibles en ausencia de datos de fidelidad. Por ejemplo, un editor jefe señaló: "No entiendo cómo se pueden interpretar los hallazgos sin una medida de fidelidad de procedimiento". Este académico continuó señalando: "Creo que es independiente de los datos y que lo estás haciendo [recopilar datos de fidelidad] desde el principio... Antes incluso de saber cómo van a ser los datos, deberías estar recopilando datos de fidelidad de procedimiento."

Cuando los estudiosos devaluaron la inclusión de datos sobre la fidelidad, hubo tres razones principales: la habilidad del ejecutor, la complejidad o novedad de los procedimientos y los resultados obtenidos del estudio. Tratamos la confianza en las habilidades del aplicador en la sección anterior. La complejidad de los procedimientos y los datos obtenidos se describen a continuación.

Ya desde Peterson et al. (1982), la complejidad procedimental y el correspondiente riesgo de errores han sido un punto de discusión en torno a la medida de la fidelidad en el análisis de conducta. El papel de la

complejidad procedimental persistió en los comentarios de los participantes. Por ejemplo, un académico a mitad de carrera señaló,

En cuanto a la pregunta de investigación, también depende del tratamiento que se esté aplicando. Al principio de mi carrera, trabajé mucho con la música como tratamiento de la estereotipia. No hay mucha fidelidad que tomar ahí. Se trata de si se activa con música o no. La gente podría pedirme que midiera la fidelidad del tratamiento, pero hasta qué punto sería útil frente a si hago una intervención más compleja como reforzamiento diferencial de otras conductas, en tal caso se hace importante, creo, porque podrías perderte información importante.

Un aspecto interesante de la consideración de la complejidad es que no hubo mucho consenso sobre lo que constituye una intervención compleja o simple. Por ejemplo, el académico a mitad de carrera citado anteriormente categorizó el reforzamiento diferencial de otras conductas como una intervención compleja que requeriría la medida de los datos de fidelidad. Ese mismo erudito continuó señalando que las intervenciones podían considerarse complejas cuando eran "multicomponente" o "realmente novedosas". Otros estudiosos consideraron que los procedimientos eran complejos sólo cuando había pequeñas diferencias entre las condiciones. Por ejemplo, un editor asociado a mitad de carrera señaló: "Si la variable independiente es, digamos, la demora de la ayuda de uno frente a la de tres segundos, entonces me parece absolutamente crucial tener información sobre cuáles fueron esas demoras reales, porque somos humanos."

Otros estudiosos evaluaron el riesgo en función de si un ser humano aplicaba los procedimientos, considerando todos los estudios aplicados por un ser humano de alto riesgo y sólo los estudios en los que los procedimientos estaban mecanizados (p.ej., aplicados por un ordenador) de un riesgo lo suficientemente bajo como para justificar la falta de datos sobre la fidelidad. Un académico a mitad de carrera señaló,

Si se realiza un estudio en el que todo lo que hace el ser humano es leer algunas instrucciones de un guion o algo así al principio de la sesión y, después, los estímulos [son] presentados por un ordenador o algún otro dispositivo, entonces no veo por qué se necesitarían datos de fidelidad de

procedimiento. Aparte de eso, creo que probablemente siempre deberían recopilarse.

Si los analistas de conducta deben determinar cuándo recopilar datos de fidelidad en función del riesgo de error o de la complejidad del análisis, puede ser beneficioso llegar a un consenso sobre qué aspectos de un procedimiento lo harían de alto o bajo riesgo. Como ya se ha señalado, actualmente no existe mucho consenso sobre esta cuestión.

Los académicos también parecían valorar a veces los datos de fidelidad principal o exclusivamente cuando los datos de la variable dependiente eran variables o no cambiaban en las direcciones esperadas cuando cambiaba la variable independiente. Por ejemplo, un editor señaló: "particularmente en los fallos de replicación, si se quiere publicar eso, creo que va a ser realmente importante tener su fidelidad de procedimiento."

Medir o notificar los datos de fidelidad de los fallos de replicación o de los hallazgos inesperados puede, de hecho, ayudar a identificar supuestas relaciones funcionales que se debieron a errores de fidelidad (no medidos). Imagine una situación en la que un estudio inicial sobre un tema concreto demostró que la variable independiente Y parecía aumentar la variable dependiente X , pero no se midió la fidelidad. En un estudio posterior realizado por un grupo de investigación diferente, X no tuvo ningún efecto sobre Y . Este último grupo grabó en vídeo las sesiones y analizó los datos de fidelidad post hoc para demostrar que X se aplicó tal y como se había descrito. Esta medida de la fidelidad refuerza considerablemente la afirmación del segundo grupo de investigación de que la relación funcional identificada inicialmente puede haber sido espuria o deberse a una menor fidelidad.

Sin embargo, la falta de medida de la fidelidad en el primer estudio hace que no esté claro si la falta de replicación se debió a (a) la identificación de los límites de generalidad del hallazgo (es decir, X afecta a Y en las circunstancias del primer estudio, pero no del segundo) o (b) a errores de fidelidad en el primer estudio (para el que no se midió la fidelidad). La inclusión de una medida de la fidelidad en ambos estudios dilucidaría con mayor claridad las razones de la falta de replicación, con lo que avanzaríamos en nuestra comprensión.

Tema 5: Falta de conocimientos o recursos

En consonancia con evaluaciones anteriores (p.ej., Hagermoser Sanetti y DiGennaro Reed, 2012), los académicos señalaron con frecuencia la falta de conocimientos o de recursos como motivo para no medir

o presentar datos de fidelidad. Hubo varios factores generales que los académicos identificaron en relación con la falta de entrenamiento. En primer lugar, algunos académicos señalaron que la instrucción sobre la medida de la fidelidad no estaba incluida en los cursos de posgrado. Por ejemplo, un experto de nivel medio señaló: "Nuestros estudiantes no están formados para medir la fidelidad del tratamiento" y sugirió que la falta de profesionales formados era un obstáculo considerable para la medida.

Varios académicos relacionaron el entrenamiento inadecuado de los estudiantes con una segunda barrera de conocimiento: la falta de recursos publicados sobre cómo recopilar y analizar datos. Un editor asociado a mitad de carrera señaló,

[T]enemos métodos bien establecidos para obtener acuerdo entre observadores en diferentes contextos. Puedes consultar un libro de métodos de investigación y decir: "Estoy recopilando datos de frecuencia. Estas son [sic] mis opciones. Estoy haciendo intervalos parciales. Aquí están [sic] mis opciones. No tenemos nada parecido para la fidelidad del tratamiento.

Este académico continuó señalando su preocupación por la falta de recursos en lo que respecta no sólo a la recopilación sino también al análisis de los datos de fidelidad:

Entonces, hagamos un porcentaje, bueno, ¿son todos iguales? Si no son iguales, ¿cómo sopesamos la importancia de este componente frente a ese otro? En mi opinión, parte de la preocupación de tener un estándar es que produciría lo que hemos hablado. Sería como hacer churros, pero puede que no nos aporte nada más de lo que tenemos. Creo que, si los revisores están bien formados en la importancia de esto, los revisores van a saber preguntar bajo las condiciones apropiadas, "¿dónde están los datos de fidelidad?" y ser capaces de rechazar investigaciones que no los recogieron y no pueden explicar por qué ocurrió algo en relación con la intervención que hicieron.

Otros estudiosos se hicieron eco de la falta de entrenamiento suficiente para promover una medida flexible y adecuada de la fidelidad. Un editor señaló,

En mi clase de métodos de investigación, dedicamos toda una clase a cómo medir y elaborar su sistema [de fidelidad]. Es un reto para los estudiantes. Pueden hacer el AEO, es bastante sencillo, pero es la fidelidad de procedimiento y, en particular, aplicarla a modelos ligeramente nuevos, es mucho más complejo. Si la variable independiente es diferente, el estudio es ligeramente distinto, a veces tienen problemas. He intentado dar algunas pautas generales con muchos ejemplos de esos sistemas.

En conjunto, estas preocupaciones sobre la medida llevaron a algunos estudiosos a preguntarse sobre el valor de los datos de fidelidad y si todas las medidas de fidelidad tenían el mismo valor. A los académicos les preocupaban especialmente las medidas de fidelidad que a veces calificaban de "memorísticas" o "superficiales". Por ejemplo, un editor señaló haber observado una sesión en la que el observador de la fidelidad parecía registrar la secuencia de ayudas enumeradas en el protocolo independientemente de si la conducta del ejecutor era la prevista. Los expertos consideraron que este tipo de medidas de fidelidad no capturaban "cómo responde la variable dependiente a las variables independientes".

La falta de entrenamiento de los estudiantes (y, tal vez, de los propios investigadores principales) llevó a algunos académicos a sugerir que la recopilación de datos de fidelidad era insostenible, al menos en situaciones en las que los estudiantes actuaban como encargados de aplicar o recopilar los datos. Por ejemplo, un editor asociado al comienzo de su carrera dijo: "Muchas veces, los estudiantes de maestría tienen dificultades si les das más de dos o tres cosas de las que deben de estar pendientes; por ello, con frecuencia, otros aspectos de mayor importancia, como el AEO, desplazan el registro de la fidelidad".

Aunque hubo un reconocimiento generalizado de la falta de entrenamiento y de recursos, varios académicos consideraron que se trataba de un área emergente del campo. Sugirieron estrategias asequibles que podrían adoptarse para mejorar el acceso a los recursos y formación adecuados. Estas estrategias incluían realizar más investigaciones en las que la fidelidad es la variable dependiente a fin de informar sobre la frecuencia necesaria de la recogida de datos de fidelidad, así como sobre los tipos de errores que deberían medirse. Un editor asociado senior señaló que las normas actuales aplicadas

en relación con la frecuencia de la medida de la fidelidad⁵ son "una engañifa". Otro editor asociado senior señaló que las normas actuales para la medida de la fidelidad son "completamente arbitrarias".

Para alejar el campo de las normas arbitrarias y mejorar la probabilidad de obtener datos de fidelidad de alta calidad, varios académicos sugirieron que sería valioso disponer de más documentos de orientación y materiales de entrenamiento. Un editor asociado senior señaló su lucha histórica por identificar recursos didácticos relacionados con la fidelidad. Este académico se lamentó diciendo: "Si piensas en los pocos libros de texto que tenemos sobre metodología de caso único, ¿cuántos de ellos tienen un capítulo dedicado a la fidelidad de procedimiento? El número es muy bajo e incluso podría ser cero". Los capítulos de libros fácilmente consumibles o los tutoriales sobre la medida de la fidelidad y el cálculo de los coeficientes de fidelidad pueden ser beneficiosos para utilizarlos en los cursos o en la supervisión.

Independientemente de las contingencias para informar sobre los datos de fidelidad (como se ha descrito anteriormente), varios estudiosos señalaron que animar a los autores a incluir detalles de los datos de fidelidad recopilados puede ser valioso. Varios académicos recomendaron que dichos detalles se incluyeran como información suplementaria en la publicación correspondiente para aumentar los recursos y modelos a disposición de otros estudiantes y profesionales. Un editor asociado senior señaló,

Sería súper fácil para el editor de una revista decir simplemente: "Se requiere que todas sus hojas de datos, fidelidad de procedimiento, AEO, recopilación de datos, todas esas cosas, se depositen y se hagan públicas para que los autores puedan verlas". ... ¿De dónde saco la mayoría de mis métodos? Obtengo los métodos de las secciones de métodos de otros artículos que la gente ha hecho antes. Si esos detalles adicionales estuvieran a mi disposición, sin duda buscaría allí y eso me daría una buena base desde la que avanzar.

Un académico a mitad de carrera coincidió en estos sentimientos y sugirió que incluir detalles de las medidas de fidelidad en la información suplementaria podría ser

un trampolín útil para otros interesados en desarrollar o validar medidas de fidelidad. Este académico opinó que "estaría bien saber si tienen un repositorio donde podamos depositar ejemplares exactos de esa [medida de fidelidad]".

Los investigadores consideraron que la inclusión de medidas detalladas de fidelidad también permitiría una evaluación más sólida de los resultados de la investigación. Un investigador de la fidelidad sugirió,

Me gusta mucho mirar esas listas de comprobación de la fidelidad, porque... a veces ves componentes que no parecen críticos, pero que probablemente aumentan el porcentaje. Por ejemplo, tener papel y el lápiz presentes para tomar los datos, y es como, vale, pues claro [risas]. Sin embargo, ese tipo de componentes no me importa tanto en una medida de fidelidad de procedimiento, pero me importan los antecedentes, me importan las consecuencias, me importan todos los componentes de la propia intervención.

A menudo veo lo que yo consideraría variables de relleno, un montón de variables que realmente no importan en una lista de control de fidelidad, lo que me preocupa entonces cuando veo porcentajes que son un poco más bajos de lo que sería preferible, y entonces me pregunto, ¿cuáles son los componentes que no se aplicaron correctamente? Si eran los componentes críticos de la intervención, entonces me preocupa que las cosas que no parecen importar tanto sean las que están haciendo subir ese porcentaje, y que las cosas importantes no estén ocurriendo tan a menudo. Me gusta poder ver qué utilizaron como medida para poder hacer este tipo de juicios.

DISCUSIÓN

Durante los últimos 40 años, los estudios han documentado repetidamente la escasez con la que la fidelidad de procedimiento es presentada en la literatura analítico-conductual (p.ej., Falakfarsa et al., 2022; Gresham et al., 1993; Han et al., 2022; McIntyre et al., 2007; Peterson et al., 1982). Sin embargo, las razones de ello han seguido siendo en gran medida especulativas.

⁵ En el laboratorio de este académico, "La regla del 25: 25% de evaluación de AEO, 25% de sesiones de tratamiento con datos de fidelidad de procedimiento, 25% de esas sesiones con AEO". Esta "regla" es similar a las métricas de calidad para la investigación de caso único, que a menudo sugieren que la fidelidad debe medirse en al menos el 20% de las sesiones y el nivel de fidelidad deben ser de al menos el 80% (p.ej., Ledford et al.,

2020; National Autism Council, 2015). El carácter arbitrario de estas normas se ha establecido empíricamente posteriormente en estudios que demuestran que los efectos de la intervención pueden perderse fácilmente cuando ésta se aplica con una fidelidad inferior al 80% (Jones y St. Peter, 2022).

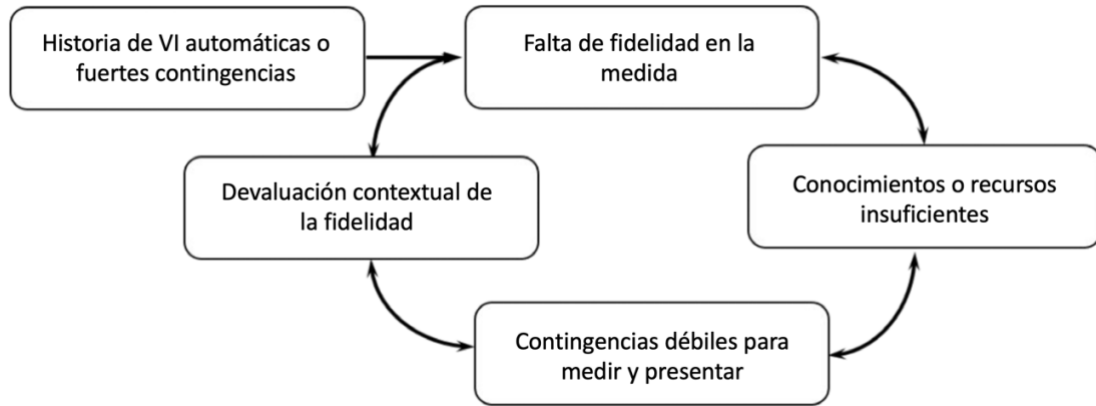


FIGURA 1 Interacción entre los temas identificados.

Los estudios realizados con académicos de ciencias afines (p.ej., psicología escolar) han sugerido que la falta de directrices y las consideraciones prácticas pueden ser barreras significativas para incluir datos sobre fidelidad. Estos hallazgos son limitados porque no abordaron las consideraciones culturales que pueden diferir entre el análisis de conducta y otras ciencias afines.

Para abordar esta laguna en la literatura, entrevistamos a 25 académicos que actuaron como autores de correspondencia de al menos un estudio publicado en JABA durante 2006-2020. El análisis de las discusiones de los académicos reveló cinco temas principales que contribuyen a la falta de presentación de datos de fidelidad. Estos cinco temas fueron: variables históricas, contingencias débiles, confianza en las capacidades del aplicador, devaluación de los datos de fidelidad y falta de conocimientos o recursos.

Los cinco temas están claramente interrelacionados (véase la figura 1). El análisis aplicado de conducta contemporáneo surgió de raíces históricas que habitualmente implicaban variables dependientes automatizadas. Aunque un humano diseñaba los parámetros del programa de reforzamiento y programaba el aparato, los primeros experimentos de análisis de conducta no dependían de un humano para aplicar cada componente. Así, de forma muy parecida a la investigación contemporánea automatizada por ordenador, los experimentadores necesitaban comprobar rigurosamente el programa (y quizá verificar el funcionamiento correcto antes de cada sesión) pero no necesitaban cuantificar más la entrega de la variable

independiente para identificar relaciones funcionales claras⁶.

Quizá debido a esta historia de medida automatizada, es posible que los primeros estudiosos del análisis aplicado de conducta no desarrollaran sistemas para medir o informar del grado en que los procedimientos se aplicaban tal y como se habían diseñado. También es posible que la necesidad de este tipo de medidas se considerara menos crítica porque las variables dependientes eran, por lo general, muy potentes. Se consideraba necesaria una descripción completa de los procedimientos (p.ej., la dimensión tecnológica descrita por Baer et al., 1968), pero se omitía la medida en que esos procedimientos se aplicaban tal y como se habían descrito.

Así, las primeras publicaciones sobre análisis aplicado de conducta carecían de medidas de fidelidad. Peterson et al. (1982) informaron de que sólo el 23% de los artículos publicados en JABA en 1968 incluían la medida de la fidelidad de algún componente de la variable independiente; cada uno de la serie de siete años posteriores tuvo porcentajes aún más bajos. La falta de medidas tempranas significaba que los estudiosos no estaban acumulando los conocimientos necesarios para desarrollar sistemas de medida de la fidelidad eficaces y eficientes. Los sistemas de medida de la fidelidad quedaron subordinados a los de medida del AEO, que ya estaban bien establecidos a principios de los años ochenta.

Los sistemas poco desarrollados pueden haber contribuido a que los miembros del profesorado no se

⁶ Sin embargo, cabe destacar que las comprobaciones informales de la fidelidad de la variable independiente han sido, y son, frecuentes en los estudios del análisis experimental de la conducta. Por ejemplo, los autores interesados en los efectos de la tasa de reforzamiento en estudios de

laboratorio informarán no sólo de las tasas de respuesta obtenidas, sino también de las tasas de reforzamiento obtenidas. Esto último sirve para que los lectores comprueben que la manipulación afectó realmente a la variable independiente prevista.

sintieran preparados para enseñar a sus alumnos a medir e informar sobre la fidelidad. La mayoría de nuestros académicos participantes podían recordar claramente la instrucción sobre el AEO, pero su recuerdo de la instrucción sobre las medidas de fidelidad era vago en el mejor de los casos. Los expertos que ejercieron como miembros del profesorado también lamentaron la falta de recursos disponibles en la actualidad con los que apoyar a sus estudiantes. Esto es especialmente desafortunado porque sin un entrenamiento graduado apropiado en la medida de la fidelidad, los profesionales pueden ser menos propensos a recoger esos datos en apoyo de una supervisión apropiada.

Uno de los recursos que, según los estudiosos, faltaba especialmente era una orientación clara sobre los detalles de la medida, incluida la frecuencia necesaria de la medida y los coeficientes de fidelidad mínimos necesarios (recordemos que los estudiosos describieron este aspecto de la práctica científica contemporánea como "completamente arbitrario" y sus propias normas al respecto como "engañifa"). La falta de orientaciones claras publicadas sobre cómo medir la fidelidad, con qué frecuencia medir la fidelidad y los coeficientes mínimos de fidelidad probablemente contribuyó a las débiles contingencias actuales para incluir datos de fidelidad en las publicaciones de los estudiosos del analítico-conductual.

La falta de contingencias para medir e informar sobre la fidelidad permite a los académicos seguir confiando ciegamente en el conjunto de habilidades de sus propios aplicadores, en particular cuando esas personas tienen experiencia previa en análisis de conducta. Los académicos pueden sentirse especialmente confiados cuando la conducta cambia en la dirección esperada. En estas circunstancias, el experto puede creer que no es necesario medir o informar sobre la fidelidad. Cuando los estudiosos no miden la fidelidad, siguen ignorando felizmente los posibles errores de aplicación, lo que puede reforzar aún más la idea de que los datos sobre la fidelidad no son necesarios. Esta falta de medida reinicia así toda la cascada de variables que interactúan, como se muestra en la figura 1.

Aunque nos hemos centrado en una visión prospectiva sobre la relación entre variables, las relaciones son probablemente bidireccionales. Cuando los académicos no miden la fidelidad, tampoco identifican necesariamente cuándo los ejecutores han cometido errores y cómo esos errores pueden haber facilitado el cambio de conducta en la dirección esperada. Cuando esos estudiosos actúan como responsables editoriales (p.ej., editores asociados, editores jefe), es posible que no soliciten ni exijan datos sobre la fidelidad

en los trabajos que procesan para su publicación. Por lo tanto, esos trabajos no contribuyen a una base de datos de medidas de fidelidad que puedan utilizar otros autores. Los investigadores posteriores, al carecer de recursos suficientes, tampoco miden la fidelidad.

Varios de los temas apuntaban a cuestiones relacionadas con el esfuerzo que rodea a la medida de los datos de fidelidad. El esfuerzo es un constructo difícil de definir, pero parece estar correlacionado negativamente con la fluidez de la respuesta (Chung, 1965). Varios estudiosos mencionaron obstáculos para la recopilación de datos de fidelidad relacionados con el tiempo y los recursos necesarios, así como con la falta de entrenamiento. En combinación, estas perspectivas sugieren que mejorar la fluidez de los estudiosos con el desarrollo, el uso y el análisis de sistemas para obtener datos de fidelidad puede ser un paso útil para aumentar la frecuencia y la calidad de los datos de fidelidad en la investigación analítico-conductual.

Por supuesto, nuestros datos se limitan a las opiniones de sólo unos pocos eruditos establecidos en el análisis aplicado de conducta que publicaron su trabajo en JABA como autor de correspondencia en los últimos años. Creemos que archivar las perspectivas de este pequeño, pero presumiblemente bien informado, grupo de individuos es una importante contribución a la literatura. Sin embargo, sin duda habríamos recibido aportaciones diferentes si nos hubiéramos enfocado a otros estudiantes y profesionales que realizan investigación, incluidos los investigadores que se centran en los principios conductuales, los estudiantes de posgrado o las personas que ejercen como clínicos a tiempo completo. Esta posibilidad se ve subrayada por las diferencias en nuestros hallazgos en relación con los de autores que han encuestado a estudiosos de la psicología clínica o escolar (p.ej., Hagermoser Sanetti & DiGennaro Reed, 2012). Al igual que en estudios anteriores, los factores relacionados con la falta de directrices, recursos y entrenamiento se percibieron como barreras. Sin embargo, nuestro enfoque cualitativo nos permitió identificar algunas barreras potencialmente novedosas que no se habían evaluado o identificado previamente, como la historia de la medida dentro del análisis de conducta y la confianza en las capacidades del aplicador. Estas diferencias en los efectos ponen de relieve las variables culturales potencialmente únicas dentro del análisis de conducta como disciplina y proporcionan puntos claros para la intervención si los estudiosos desean aumentar los informes de fidelidad en la investigación analítico-conductual.

Al igual que nuestros participantes, nuestro equipo incluía personas con un historial de publicaciones en

JABA (así como personas con un sólido historial en métodos cualitativos que eran relativamente nuevos en las perspectivas analítico-conductuales). Nuestro equipo entró en este proyecto de investigación con nuestra propia opinión establecida y unánime: que la medida de la fidelidad es crítica para cualquier procedimiento aplicado por humanos. Sin embargo, el proceso de investigación afinó aún más nuestras opiniones y cambió nuestras perspectivas en direcciones inesperadas.

En primer lugar, adoptamos un lenguaje diferente en relación con la medida de la fidelidad de la variable independiente. Aunque los tres autores analítico-conductuales utilizaron por defecto el término "fidelidad de tratamiento", después de los grupos de discusión nos inclinamos a favor de "fidelidad de procedimiento". Coincidimos con las opiniones de algunos estudiosos participantes que señalaron que el término "fidelidad del tratamiento" puede ser a veces problemático. Varios estudiosos señalaron que "tratamiento" sugiere que la medida sólo debe producirse en el contexto de la intervención en lugar de en todas las fases experimentales (p.ej., línea base, mantenimiento) y debe limitarse a los contextos en los que se está "tratando" un problema." El término "fidelidad" también es potencialmente problemático debido a la connotación de que una fidelidad baja o deficiente es necesariamente un efecto negativo o tiene que ver con una característica estática del que lo aplica. Coincidimos con algunos de los participantes de nuestros grupos de discusión en que la adopción del término "fidelidad de procedimiento" elude algunos de estos problemas potenciales.

En segundo lugar, hemos cambiado nuestras ideas sobre los puntos de intervención para aumentar la comunicación de la fidelidad de procedimiento sobre el terreno. Algunos de los temas que identificamos en el estudio actual serán difíciles de cambiar. No podemos reescribir la historia, obligar a los individuos a empezar a medir la fidelidad ni cambiar la normativa editorial por decreto. Sin embargo, la falta de recursos parece un punto de intervención razonable para las personas (como nosotros) que creen que el campo se beneficiaría de una comunicación más frecuente de los datos de fidelidad. Un editor asociado senior se mostró de acuerdo al afirmar: "Creo que vamos a necesitar capítulos y artículos sobre esto para que las cosas avancen. Quizá incluso como precursores de los requisitos editoriales". Esperamos que la publicación de nuestros datos cualitativos sirva de trampolín para el desarrollo de estos recursos adicionales.

Existen varias áreas potenciales para desarrollar una base de conocimientos sobre la fidelidad. La medida del AEO puede haber mejorado tras las comparaciones sistemáticas de varios tipos de cálculos de concordancia,

en el contexto de procedimientos variados de recogida de datos. Estas comparaciones incluyeron evaluaciones con "valores verdaderos" conocidos de los datos y réplicas con datos recogidos en contextos de investigación más habituales. Hasta la fecha, se han publicado pocas evaluaciones similares centradas en la fidelidad de procedimiento (para excepciones notables, véase Collier-Meek et al., 2020, 2021; Cook et al., 2015). La identificación de las ventajas y desventajas relativas de los diversos métodos de medida de la fidelidad y de los métodos correspondientes para calcular los coeficientes de fidelidad puede ser un área importante para futuras investigaciones. Se anima a los lectores deseosos de seguir un entrenamiento adicional sobre la medida de la fidelidad a que revisen varios artículos existentes que proporcionan instrucciones útiles y tratan sobre la medida y el cálculo de la fidelidad (Gresham et al., 2000; Hagermoser Sanetti & Fallon, 2011; Kodak et al., 2022; Vollmer et al., 2008), incluyendo hojas de datos de ejemplo para la medida de dichos datos (Han et al.)

Mientras los estudiosos siguen elaborando directrices adicionales, los investigadores que ya miden sistemáticamente la fidelidad pueden contribuir al desarrollo de la bibliografía describiendo formas de medir la fidelidad que requieran poco esfuerzo. Los estudiosos que participaron en el presente estudio mencionaron con frecuencia la incorporación de medidas de fidelidad en el sistema primario de recogida de datos o el uso de la grabación en vídeo para recopilar datos de fidelidad a posteriori. Disponer de diversos métodos publicados para medir la fidelidad puede ayudar a identificar las condiciones en las que determinadas estrategias resultan más útiles. Los investigadores también podrían compartir sus medidas de fidelidad (p.ej., la lista de comprobación, la rúbrica o la hoja de datos reales), así como los coeficientes de fidelidad detallados (p.ej., los tipos y frecuencias de errores) como parte de la información suplementaria de la publicación (p.ej., Han et al., 2022; Kodak et al., 2022). De este modo, estos materiales estarían fácilmente disponibles para los lectores que estén trabajando en áreas similares o que deseen reproducir los hallazgos. La inclusión de estos datos y otra información de apoyo puede minimizar los posibles efectos negativos de las largas secciones sobre fidelidad en la legibilidad del artículo y, al mismo tiempo, reducir la probabilidad de que se pida al autor que disminuya o elimine la información durante el proceso editorial.

Los temas identificados en la presente evaluación proporcionan orientación para informar sobre la fidelidad en el trabajo analítico-conductual. Basándonos en nuestros hallazgos, ofrecemos tres recomendaciones

principales. En primer lugar, animamos a los investigadores a medir e informar sobre la fidelidad de procedimiento en cualquier estudio en el que la variable independiente sea aplicada por personas. Los investigadores no deben suponer que la fidelidad es adecuada simplemente porque el aplicador tenga una credencial avanzada o trabaje en una institución concreta o porque la variable dependiente haya cambiado en la dirección esperada. Informar de manera más amplia acerca de la fidelidad también proporcionaría una base para los análisis correlacionales con el fin de determinar las posibles "condiciones límite" en las que las intervenciones son eficaces o los tipos o niveles de fidelidad que podrían manipularse posteriormente en estudios controlados. En segundo lugar, más investigadores deberían evaluar en qué medida los procedimientos se ven afectados por los cambios en la fidelidad. Estas evaluaciones ayudarían a establecer normas basadas en datos con respecto a las cuales podrían evaluarse la frecuencia y los valores de fidelidad obtenidos. En tercer lugar, y como ya se ha señalado, los estudiosos deberían producir (y consumir) recursos relacionados con formas eficientes y eficaces de medir la fidelidad en diversos ámbitos de la investigación sobre la conducta.

Como mínimo, los datos actuales proporcionan apoyo a varios factores relacionados con la presentación de informes que anteriormente sólo habían sido especulativos. Esperamos que estos datos animen a los investigadores contemporáneos a adoptar las partes de nuestra historia que promueven la inclusión de datos de fidelidad, lo que conllevará más recursos y, en última instancia, unos estándares de calidad más elevados en las publicaciones.

CONFLICTO DE INTERESES

Los autores declaran no tener ningún conflicto de intereses.

NORMAS ÉTICAS

Esta investigación fue aprobada por el comité ético de West Virginia University. Todos los participantes dieron su consentimiento informado.

REFERENCIAS

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>
- Barker, C., & Pistrang, N. (2021). Choosing a qualitative method: A pragmatic, pluralistic perspective. In P. M. Camic (Ed.), *Qualitative research in psychology: Expanding perspectives in methodology and design* (2nd ed., pp. 27–50). American Psychological Association. <https://doi.org/10.1037/0000252-002>
- Barnett, D., Hawkins, R., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education*, *23*(1), 89–107. <https://doi.org/10.1007/s10864-013-9188-y>
- Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234. <https://doi.org/10.1177/1468794112468475>
- Bergmann, S., Long, B. P., St. Peter, C. C., Brand, D., Strum, M., Han, J. B., & Wallace, M. (2022). *A detailed examination of reporting procedural fidelity in the Journal of Applied Behavior Analysis*. [Unpublished manuscript]. Department of Behavior Analysis, University of North Texas.
- Birks, M., Chapman, Y., & Francis, K. (2008). Memoing in qualitative research: Probing data and processes. *Journal of Research in Nursing*, *13*(1), 68–75. <https://doi.org/10.1177/1744987107081254>
- Blanco, G. L., & Rossman, G. B. (2022). As a qualitative study unfolds: Shifts in design and analysis. In C. Vanover, P. Mihás, & J. Saldana (Eds.), *Analyzing and interpreting qualitative research: After the interview* (pp. 1–6). SAGE.
- Brand, D., Elliffe, D., & DiGennaro Reed, F. D. (2018). Using sequential analysis to assess component integrity of discrete-trial teaching programs. *European Journal of Behavior Analysis*, *19*(1), 30–47. <https://doi.org/10.1080/15021149.2017.1404392>
- Brand, D., Henley, A. J., DiGennaro Reed, F. D., Gray, E., & Crabbs, B. (2019). A review of published studies involving parametric manipulations of treatment integrity. *Journal of Behavioral Education*, *28*(1), 1–26. <https://doi.org/10.1007/s10864-018-09311-8>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE.
- Breeman, S., Vladescu, J. C., DeBar, R. M., Grow, L. L., & Marano, K. E. (2020). The effects of procedural integrity errors during auditory–visual conditional discrimination training: A preliminary investigation. *Behavioral Interventions*, *35*(2), 203–216. <https://doi.org/10.1002/bin.1710>
- Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An examination of treatment integrity errors on skill acquisition during discrete trial instruction. *Journal of Applied Behavior Analysis*, *46*(2), 379–394. <https://doi.org/10.1002/jaba.49>
- Chung, S. (1965). Effects of effort on response rate. *Journal of the Experimental Analysis of Behavior*, *8*(1), 1–8. <https://doi.org/10.1901/jeab.1965.8-1>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). SAGE.
- Collier-Meek, M. A., Sanetti, L. M., Fallon, L., & Chafouleas, S. (2020). Exploring the influences of assessment method, intervention steps, intervention sessions, and observation

- timing on treatment fidelity estimates. *Assessment for Effective Intervention*, 46(1), 3–13. <https://doi.org/10.1177/1534508419857228>
- Collier-Meek, M. A., Sanetti, L. M. H., Gould, K., & Pereira, B. (2021). An exploratory comparison of three treatment fidelity assessment methods: Time sampling, event recording, and postobservation checklist. *Journal of Educational & Psychological Consultation*, 31(3), 334–359. <https://doi.org/10.1080/10474412.2020.1777874>
- Cook, J. E., Subramaniam, S. S., Brunson, L. Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior Analysis in Practice*, 8(1), 37–47. <https://doi.org/10.1007/s40617-014-0039-7>
- Daly, J., Kellehear, A., & Gliksman, M. (1997). *The public health researcher: A methodological approach*. Oxford University Press.
- Darwin Holmes, A. G. (2020). Researcher positionality: A consideration of its influence and place in qualitative research: A new researcher guide. *International Journal of Education*, 8(4), 1–10. <https://doi.org/10.34293/education.v8i4.3232>
- Dedoose. (2021). Dedoose (Version 9.0.17) [Computer software]. *Socio-Cultural Research Consultants, LLC*. www.dedoose.com
- DiGennaro Reed, F. D., Blackman, A. L., Erath, T. G., Brand, D., & Novak, M. D. (2018). Guidelines for using behavioral skills training to provide teacher support. *Teaching Exceptional Children*, 50(6), 373–380. <https://doi.org/10.1177/0040059918777241>
- Essig, L., Rotta, K., & Poling, A. (2023). Interobserver agreement and procedural fidelity: An odd asymmetry. *Journal of Applied Behavior Analysis*, 56(1), 78–85. <https://doi.org/10.1002/jaba.961>
- Falakfarsa, G., Brand, D., Jones, L., Godinez, E. S., Richardson, D. C., Hanson, R. J., Velazquez, S. D., & Wills, C. (2022). Treatment integrity reporting in Behavior Analysis in Practice 2008–2019. *Behavior Analysis in Practice*, 15(2), 443–453. <https://doi.org/10.1007/s40617-021-00573-9>
- Fryling, M. J., & Baires, N. A. (2016). The practical importance of the distinction between open and closed indirect assessments. *Behavior Analysis in Practice*, 9(2), 146–151. <https://doi.org/10.1007/s40617-016-0115-2>
- Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, 26(2), 257–263. <https://doi.org/10.1901/jaba.1993.26-257>
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15(4), 198–205. https://doi.org/10.1207/SLDRP1504_4
- Hagermoser Sanetti, L. M., & DiGennaro Reed, F. D. (2012). Barriers to implementing treatment integrity procedures in school psychology research: Survey of treatment outcome researchers. *Assessment for Effective Intervention*, 37(4), 195–202. <https://doi.org/10.1177/1534508411432466>
- Hagermoser Sanetti, L. M., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational & Psychological Consultation*, 21(3), 209–232. <https://doi.org/10.1080/10474412.2011.595163>
- Han, J. B., Bergmann, S., Brand, D., Wallace, M., St. Peter, C. C., Feng, J., & Long, B. P. (2022). Trends in reporting procedural fidelity: A comparison. *Behavior Analysis in Practice*, 16, 388–398. <https://doi.org/10.1007/s40617-022-00741-5>
- Hansford, A., Zilber, I., LaRue, R. H., & Weiss, M. J. (2010). Practical issues in effective behavioral intervention development. *Behavior Analyst Today*, 11(1), 59–76. <https://doi.org/10.1037/h0100689>
- Hennick, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
- Hill, C. E., & Knox, S. (2021). *Essentials of consensual qualitative research*. American Psychological Association. <https://doi.org/10.1037/0000215-000>
- Hill, C. E., Thompson, B. J., & Williams, E. N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist*, 25(4), 517–572.
- Johnston, J. M., Pennypacker, H. S., & Green, G. (2020). *Strategies and tactics of behavioral research and practice* (4th ed.). Routledge/Taylor & Francis Group.
- Jones, S. H., & St. Peter, C. C. (2022). Nominally acceptable integrity failures negatively affect interventions involving intermittent reinforcement. *Journal of Applied Behavior Analysis*, 55(4), 1109–1123. <https://doi.org/10.1002/jaba.944>
- Journal of Applied Behavior Analysis. (2022). *Journal of Applied Behavior Analysis author guidelines*. Wiley. <https://onlinelibrary.wiley.com/page/journal/19383703/homepage/ForAuthors.html>
- Kahlke, R. M. (2014). Generic qualitative approaches: Pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods*, 13(1), 37–52. <https://doi.org/10.1177/160940691401300119>
- Kite, J., & Phongsavan, P. (2017). Insights for conducting real-time focus groups online using a web conferencing service. *F1000 Research*, 6, 122. <https://doi.org/10.12688/f1000research.10427.1>
- Kodak, T., Bergmann, S., & Waite, M. (2022). Strengthening the procedural fidelity research-to-practice loop in animal behavior. *Journal of the Experimental Analysis of Behavior*, 118(2), 215–236. <https://doi.org/10.1002/jeab.780>
- Kruger, R. A., & Casey, M. A. (2014). *Focus groups: A practical guide for applied research* (5th ed.). SAGE.
- Ledford, J. R., Chazin, K. T., Lane, J. D., Zimmerman, K. N., & Ayres, K. A. (2020, September). *Single case analysis and review framework (SCARF)*. <http://ebip.vkcsites.org/scarfv2>
- Lichtman, M. (2013). *Qualitative research in education: A user's guide* (3rd ed.). SAGE.
- Malterud, K., Siersma, V., & Guassora, A. D. (2021). Information power: Sample content and size in qualitative studies. In P. M. Camic (Ed.), *Qualitative research in psychology: Expanding perspectives in methodology and design* (2nd ed., pp. 67–81). American Psychological Association. <https://doi.org/10.1037/0000252-004>
- McIntyre, L. L., Gresham, F. M., DiGennaro Reed, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interven-

- tions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659–672. <https://doi.org/10.1901/jaba.2007.659-672>
- Michael, J. (1980). Flight from behavior analysis. *The Behavior Analyst*, 3(2), 1–22. <https://doi.org/10.1007/BF03391838>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed. ed.). SAGE.
- National Autism Council. (2015). *Findings and conclusions: National standards project, phase 2*. <https://nationalautismcenter.org/national-standards/phase-2-2015/>
- Nerlich, B. (2004). Coming full (hermeneutic) circle. In Z. Todd, B. Nerlich, S. McKeown, & D. D. Clarke (Eds.), *Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice*. Psychology Press.
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice*, 5(2), 2–11. <https://doi.org/10.1007/BF03391819>
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A pyramidal approach for training multiple human service staff. *Behavior Analysis in Practice*, 6(2), 4–16. <https://doi.org/10.1007/BF03391798>
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15(4), 477–492. <https://doi.org/10.1901/jaba.1982.15-477>
- Roulston, K., & Shelton, S. A. (2015). Reconceptualizing bias in teaching qualitative research methods. *Qualitative Inquiry*, 21(4), 332–342. <https://doi.org/10.1177/1077800414563803>
- Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, 15(1), 85–109. <https://doi.org/10.1177/1525822X02239569>
- Sandelowski, M., Barroso, J., & Voils, C. I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, 30, 99–111. <https://doi.org/10.1002/nur.20176>
- Sanetti, L. M. H., Charbonneau, S., Knight, A., Cochrane, W. S., Kulcyk, M. C. M., & Kraus, K. E. (2020). Treatment fidelity reporting in intervention outcome studies in the school psychology literature from 2009 to 2016. *Psychology in the Schools*, 57(6), 901–922. <https://doi.org/10.1002/pits.22364>
- Sanetti, L. M. H., Dobey, L. M., & Gritter, K. L. (2012). Treatment integrity of interventions with children in the Journal of Positive Behavior Interventions from 1999 to 2009. *Journal of Positive Behavior Interventions*, 14(1), 29–46. <https://doi.org/10.1177/1098300711405853>
- Schwartz-Shea, P., & Yanow, D. (2013). *Interpretive research design: Concepts and processes*. Routledge.
- Tracy, S. J. (2010). Qualitative quality: Eight “Big-Tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851. Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice*, 1(2), 4–11. <https://doi.org/10.1007/BF03391722>
- What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook (Version 5.0)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/handbooks>

INFORMACIÓN SUPLEMENTARIA

Puede obtener información adicional en internet. Ver sección de Información Suplementaria al final de este artículo.