

Independence and Collaboration in Single-Case Evidence Standards: Inductive Themes from Two Expert Focus Groups

Joseph M. Lambert

joseph.m.lambert@vanderbilt.edu

Vanderbilt University

Mandy J. Rispoli

University of Virginia

Jennifer R. Ledford

Vanderbilt University

Research Article

Keywords: Single-case design, evidence-based practice, focus group research, reporting transparency, open science

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11053500/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Authoritative definitions of evidence-based practice (EBP) frequently include a “5–3–20” heuristic (i.e., at least five studies, conducted by at least three investigators or research groups, demonstrating positive outcomes for at least 20 participants). However, multi-site collaborations—often necessary for intervention development, resource sharing, and accelerated knowledge generation—can complicate “independence across groups.” We conducted two brief in-person focus groups with methodological experts at an invitation-only single-case design (SCD) conference (approximately 10–20 attendees per session; 5 and 9 active participant speakers, respectively) using a standardized prompt protocol about operationalizing “research group” and “independence,” perceived threats and advantages of coordinated collaborations, feasible safeguards, and evidence grading. Inductive thematic analysis indicated that experts framed independence as serving two primary functions—(a) bias/ownership control and (b) a portability test for tacit knowledge—and recommended a staged workflow (collaborate-to-develop, branch-to-replicate) supplemented by transparency infrastructure (e.g., decision rules/adaptation logs, structured reporting, repositories) and external checks (e.g., masking, independent assessors/monitoring). Experts also discussed “silver vs. gold” evidence grading and incentive structures (e.g., registered replications; unaffiliated teams) as pragmatic routes to unaffiliated replication. Implications are discussed for interpreting independence clauses in EBP heuristics and for designing multi-site SCD research programs that prioritize replicability without unnecessarily constraining coordinated development.

Introduction

In single-case design (SCD) research, evidence-based practice (EBP) designations typically depend on (a) the methodological adequacy of individual studies and (b) practice-level judgments based on replication across studies, participants, and contexts (Horner et al., 2005). Importantly, “practice” in this literature is broad—ranging from precise interventions to multi-component programs (Contreras et al., 2022; Ledford et al., 2021)—which raises practical questions about what kind of replication is needed to conclude that a practice is reliable and usable beyond its originators.

In SCD traditions, a commonly cited heuristic is that an intervention can be considered evidence based when positive effects are demonstrated across (a) multiple studies (e.g., 5), (b) conducted by multiple investigators or research groups (e.g., 3), and (c) a minimum number of participants (e.g., 20)—often summarized as “5–3–20” (cf. Horner et al. 2005; e.g., Clay et al., 2018). More recent definitional language has sometimes preserved this spirit while introducing ambiguity (e.g., “three different investigators or research groups”; Steinbrenner et al., 2020). Other standards sidestep this concern, requiring only different “sites” (WWC, 2022). Regardless, practical questions remain about what counts as a distinct “group”, why having different groups or sites is critical, and what degree of independence is necessary.

A core ambiguity is that independence language is consequential but often underspecified. For example, the phrase “three different investigators or research groups” (Steinbrenner et al., 2020) does not clearly specify what constitutes a distinct group. Independence could refer to institutional separation, geographic separation, authorship networks, training lineage, or independence of decision-making during implementation and analysis. This ambiguity matters because the research ecosystem increasingly incentivizes (and sometimes requires) multi-site collaboration—particularly in early-stage intervention development. Collaboration may accelerate the iterative refinement of procedures and improve feasibility, but close coordination can also concentrate conflicts of interest, amplify shared biases, and embed tacit decision-making that is not captured in formal descriptions (e.g., manuscripts). These concerns map directly onto the goals of EBP review criteria—reviewers must be able to determine whether outcomes are positive and whether threats to validity are plausible based on what is reported. They also connect to contemporary calls for improved transparency and reporting infrastructure in SCD research, including use of supplemental materials and third-party repositories when manuscripts cannot include all details, as well as practices that mitigate selective reporting and publication bias (Ledford et al., 2023; Tincani et al., 2024; Travers et al., 2026).

Notably, Ledford et al. (2023) explicitly argue that proposed methodological/reporting guidance warrants rigorous testing via approaches such as focus groups and Delphi methods to evaluate validity, establish consensus, and improve uptake. The present study responds to that recommendation by convening methodological experts to discuss how independence should be conceptualized and operationalized in the context of SCD evidence standards—particularly when high-quality evidence may be generated through coordinated multi-site collaborations.

Using two expert focus groups, we sought to inductively characterize: (a) how experts operationalize “research group” and “independence,” (b) what threats independence is intended to control (and what advantages collaboration provides), (c) what safeguards could approximate the intent of independence when complete separation is infeasible, and (d) whether evidence produced within a single coordinated collaboration should be qualified until unaffiliated replications occur.

Methods

Design and Setting

Two in-person focus groups were conducted in a classroom at the host university during an invitation-only conference focused on single-case design. Each session was approximately 30 minutes and was attended by 10–20 conference attendees—although attendance was not formally recorded. All attendees and both moderators sat in a semi-circle around a large table in the center of the room with Powerpoint slides, which offered visual anchors for moderator prompts, displayed on a projector on one of the classroom walls. Active speakers were categorized as distinct individuals who spoke at least once, excluding moderators. The host university’s IRB approved all study activities.

Moderators

The first and second authors moderated both focus groups. These moderators conceptualized this project with the intent of seeking expert guidance on how to optimize recently acquired resources and planned empirical contributions made available through a federally funded, multi-site, research grant aimed at developing a technology capable of ameliorating training bottlenecks in higher education. Both moderators were doctoral-level board certified behavior analysts (BCBA-D) with faculty positions in research-intensive universities. Moderators were not conference organizers and did not have a direct stake in the topic (EBP heuristics or single case design guidance). Although both moderators had previous professional relationships with some conference attendees (e.g., as co-authors), all conference attendees had established mature and independent programmatic lines of inquiry and no conference attendee worked in a hierarchical relationship with either moderator. The third author was an attendee at one of the focus group sessions and was the conference organizer.

Researcher's Positionality & Reflexivity Statement

As BCBA-Ds and researchers with training in single-case experimental design, the scholarly perspective of all three authors has been shaped primarily by behavioral science and the experimental analysis of behavior. Our training emphasizes observable behavior, objective measurement, experimental control, and data-based decision making. These experiences influence how we approach research questions, interpret evidence, and evaluate claims. Although behavior analytic theory informs our understanding of human behavior, we do not assume it provides the only or most complete explanation for participants' experiences.

Our research is thus situated within a pragmatic worldview that values selecting methods best suited to answering the research question. Our training as BCBA-Ds and single-case design researchers has fostered an appreciation for rigorous experimental methods while also highlighting the limitations of relying solely on quantitative evidence to understand complex educational, clinical, and implementation contexts. To answer the questions articulated in this study, we viewed qualitative analysis as the most appropriate methodological approach, as it is capable of providing insight into participants' experiences, contextual factors, and mechanisms that quantitative data alone may not capture (Burney et al., 2023). We recognize that our behavior analytic background influences the questions we ask and the patterns we notice during analysis. To promote reflexivity, we sought alternative interpretations from Amplify® during coding, and remained attentive to findings that extended beyond or challenged our assumptions.

Participants

Conference attendees were PhD-level researchers actively working in departments of special education, applied behavior analysis, psychology, and related fields who were invited to attend because they possessed methodological expertise and because their research portfolios had been deemed impactful

by conference organizers. In a few cases, the doctoral students of these researchers were also allowed to attend. All conference attendees were awarded honorariums for conference registration, travel, lodging, and food. Conference attendees who elected to attend advertised focus group sessions could choose to listen in only, or to contribute. We did not take attendance and did not collect demographic information. However, all focus-group attendees consented to be in an audio-recorded room. Participants were active speakers—distinct individuals (excluding moderators) who spoke at least once during each focus group (Table 1).

Table 1
Participants (Active Speakers) by Session

Session	Attendance (n)	Moderators (n)	Participants (n)
1	~ 15–20	2	5
2	~ 15–20	2	9
Total	~ 30–40	2	14

Procedures

To ensure consistent experiences across sessions, focus-group content was scripted and presented via PowerPoint slides. Moderators introduced the problem context (EBP heuristics and the “three investigators or research groups” clause) and then posed the same prompts in the same order across sessions (Table 2). Sessions were audio recorded with verbal notification at the outset.

Table 2
Focus-Group Prompts

Topic Domain	Prompt
Operationalization	How should we operationalize research group ? How should we operationalize independence between groups ?
Threats	What are the threats when evidence is produced within one coordinated collaboration?
Advantages	What are the advantages of coordinated collaboration?
Safeguards	What steps could protect against threats introduced by coordinated collaboration?
Evidence grading	For threats that remain, should evidence from one group be qualified/demoted , and if so, how?

Data analysis

Audio recordings were transcribed verbatim. Transcripts were pseudonymized to protect confidentiality. We conducted fully inductive thematic analysis (cf. Creswell & Plano Clark, 2017; Patton, 2014). Analysis

was proceeded by: (a) open coding each transcript, (b) constant comparison across sessions, (c) iterative clustering of codes into themes, and (d) selecting representative quotations. Initial manuscript drafts, and pseudonymized transcripts, were then uploaded into a specialized, university-specific, subscription of ChatGPT (5.2) tailored for faculty use (i.e., Amplify®). We asked this service to review the logic and coherence of our method and results sections, solicited recommendations, and considered each in a fashion similar to a typical peer-review process: adopt, adopt with modification, reject. Then, we asked it to analyze our pseudonymized transcripts, to isolate impactful quotes, and to identify emergent themes. The first author then compared each AI-generated quote, topic, and theme against his own read of the transcripts and again made a determination to either adopt, adopt with modification, or reject each proposed quote, topic, or theme (cf. Meadan et al., 2026).

Given that this study involved a single convening at a conference, it was not possible to add additional focus groups which precluded a traditional examination of saturation in the data. Instead, we examined convergence within the available data through constant comparison. The codes generated from the second group substantially overlapped with those from the first, with few novel codes introduced in the final stages of analysis. We report this as evidence of within-sample code convergence rather than saturation in the strict sense, given the bounded nature of the sample.

Interpretive constraints

Moderators sometimes summarized apparent agreement aloud (e.g., “unanimous across both groups...”). Accordingly, results emphasize (a) verbatim spoken content and (b) moderator summaries as moderator observations, while avoiding claims of unanimity unless explicitly stated by participants. Nonvocal agreement/disagreement was not captured in transcripts.

Results & Discussion

Full (pseudonymized) transcripts of both focus groups, as well as tables with a more comprehensive list of quotes relevant to initially identified themes, are offered in supporting materials. Representative quotes of themes which informed our conceptual framework are displayed in Table 3, and our conceptual framework is displayed in Fig. 1.

Table 3

Theme	Group #	Representative Quotes
Theme 1 Operationalizing “research group” and “independence”	1	S1-P1: “The emphasis is on independence because the verification should be unbiased... disconnected... not connected with the people who originated the intervention...” S1-P2: “I don’t think you can remove your learning history... you had the same advisor... it would be hard to say that you could achieve [independence]...” S1-MOD2: “What is independent? How independent is independent...?”
	2	S2-P2: “How often do you have shared lab meetings across Zoom or whatever? Are you having standing meetings? That to me, a lot of it has to do with if you establish a protocol and maybe you resolve protocol issues together but you’re trying to do your data collection and your analysis separately, that seems like independent enough because the protocol is intervention in a lot of ways.” S2-P4: “I think you are the definition of a research group. Honestly based on what you described, you’re collaborating and developing it and evaluating it. To me that sounds like a research group, but I do wonder if maybe this is what REDACTED were kind of referring to. If you did develop it, evaluate it together and then evaluate it separately and independently at your own sites, if that could then the decision making might be similar, but you’re doing replication independent of one another.” S2-P7: “And back to the question on teams, I also think that teams are students of folks, but then a lot of the different sites that replicate that, they’re in my opinion still part of that original team.” S2-P2: “That part document that. That’s to me that’s reasonably independent, means you need to publish your papers instead of one, don’t put all your names on everybody.”
Theme 2 Threats addressed by independence: COI/allegiance and bias	1	S1-P3: “We’re a small field, we all have conflicts of interest... it’s so hard to get around that issue.” S1-P4: “When I... try to replicate, I’ll reach out to the authors... so there is some influence... that may not be reported...” S1-P2: “Just thinking about conflict... [and]... a level of third party verifier...”

Theme	Group #	Representative Quotes
	2	S2-P8: “Two clear functions... [including]... could there be some kind of an unconscious bias...?” S2-P8: “When it’s my kid [intervention]... I see ‘em differently... but differently.” S2-P7: “There’s also bias... the added value of independent studies... [but] you’ve already introduced a lot of bias as part of the initial collaboration...”
Theme 3 Independence as a portability test (“secret sauce” and undocumented decision rules)	1	S1-MOD1: “The shorthand that exists amongst collaborators who make one decision with seven assumptions behind it... maybe just speaking the unspoken... every decision making rule is clearly articulated...” S1-P2: “A [noncontact] site... is probably [missing]... decisions that have been a critical feature... maybe the public doesn’t have access to...” S1-P2: “If there was really, really clear description [of] the decision making processes... that would give me a lot of confidence that somebody independent could [replicate].” S1-P5: “Documenting... what questions came up and how decisions were made... those are interesting data... [and]... help guide me how to approach it...”
	2	S2-P6: “So I start conceptualize, well the original person who ran it was this tenured professor who had this experience speaking and maybe had great examples and detailed examples. Whereas this graduate student, even though they’re doing the same steps, they’re providing an example from her to fidelity. Maybe their example’s not as good, they haven’t had the 30 years of experience, but if that grad student had done the zoom call with the professor, ‘Oh yeah, the example I gave was this,’ that’s now influencing it. And that was something that wasn’t in the original written out necessarily.” S2-P9: “I think the true marker is whether you just take a protocol, hand it to somebody, they go and then it works. So you could go all of those things and it still wouldn’t be effective in the classroom, or you could do none of it, hand it to them and then it’s a whole problem. Grand slam.”
Theme 4 Advantages of coordinated collaboration (pace, feasibility, development constraints)	1	S1-MOD2: “Sometimes [multi-site collaborations] are required by funders... Sometimes you’re developing a technology that only your team has access to...” S1-MOD1: “And if we’re a research group, well [expletive], what a waste of resources. What a unique opportunity we have to come together and really bang out a lot of potentially high quality research on a technique or an approach that didn’t exist.” (profanity redacted)
	2	S2-MOD2: “...given money from the federal government... to not be able to generate an evidence-based practice within that

Theme	Group #	Representative Quotes
		timeframe seems problematic.” S2-MOD2: “We want replication... [and]... accelerating [beyond] this snail’s pace...”
Theme 5 Skepticism about “5–3–20” (especially ‘3’), but endorsement of the underlying rationale	1	S1-P3: “5 3 20 is very arbitrary...” S1-P3: “The three... is the stumbling block... why do we make it so hard for single case designs... We set the bar much higher...” S1-MOD2: “The focus is getting back on independent replication rather than these arbitrary guidelines...” S1-MOD2: “If you have one team who, how we defines our group, who’s replicating a hundred times, and then you have another team who’s unaffiliated, truly independent, who cannot replicate, which is what you were talking about, that becomes problematic. And so is it an evidence-based practice if it really is only for folks who have gone through that training and have this intuitive understanding or learned understanding about how to make these adaptations and decisions?”
	2	S2-P2: “I’m not sure that I’m that concerned about... you should have this three sites...” S2-P8: “I’m more worried... between one lab... and then going to another lab...” (<i>emphasis on cross-lab replication rather than thresholds</i>)
Theme 6 Safeguards, evidence grading (“silver vs gold”), and enabling unaffiliated replications	1	S1-P2: “...positionality... [and]... third party verifier... site visit.” S1-P2: “The data safety monitor board... stipulations... we agree to not communicate...” S1-P2: “Structured reports... [including]... decisions... adaptations...”
	2	S2-P7: “Just thinking of one, having mass assessors all the time throughout and then maybe people that are completely disconnected to the research and having that across folks that are actually measuring behavior, for example. And maybe you increase rigor in other ways, right?” S2-P8: “Masking.” S2-MOD1 (09:33): “Should there be some sort of discounting... do we get a silver standard?” S2-P2 (21:49): “Find a new faculty member who’s not part of your tradition... here’s our protocol... go do your own...” S2-P2 (22:32): “Registered replication is a much lower bar for publication...” S2-P2 (23:34): “...you may not be authors... then... you just say, go do it.”

Independence Was Treated as Functional Rather Than Structural

Across the two focus groups, experts treated independence as a functional rather than purely structural property. Participants distinguished geographic or institutional separation from the forms of separation that might matter for evaluating evidence, including differences in training history, communication, decision-making, and involvement across phases of research. One participant characterized the purpose of independence as verification by a “disconnected” team that was not closely associated with the intervention developers (S1-P1). At the same time, participants emphasized that complete separation is difficult to achieve in a small field, where trainees and collaborators may share advisors, methods, and informal professional relationships.

Participants also described independence as graded rather than binary. Rather than treating a research group as either independent or non-independent, they discussed degrees of separation in the day-to-day conduct of research. For example, one participant suggested that collaborators could share a protocol and resolve protocol questions together while conducting data collection and analysis separately, provided those interactions were documented (S2-P2). Another participant described the same collaboration as a research group during development and as independent sites during subsequent replication. This phase-dependent interpretation provided a way to distinguish coordinated development from later testing of whether the intervention could be implemented beyond the originating collaboration.

As discussion shifted from defining independence to explaining its purpose, two functions became particularly salient: reducing risks associated with researcher investment and testing whether an intervention is portable when the developers’ tacit knowledge is unavailable.

Function 1: Independence as Protection Against Bias and Ownership

The first function involved concerns about allegiance, ownership, and interpretive bias. Participants described the small size and interconnectedness of the field as making complete insulation difficult. One participant noted that researchers often develop professional relationships with others doing similar work, while another emphasized that even a seemingly independent replication may involve informal contact with the original authors. These concerns were not framed as evidence of misconduct. Instead, participants described ownership, allegiance, and contact as something that can influence interpretation even when researchers intend to be objective. One participant captured this concern by acknowledging when an intervention is their own (i.e., at their site), they might be unintentionally biased in their interpretation of outcome data (S2-P8).

Training and mentorship were discussed as another source of dependence. Participants noted that students and postdoctoral researchers may work at different institutions while retaining substantial

shared learning histories with their advisors or originating training contexts. In this sense, institutional separation may obscure rather than eliminate the relationships that influence how an intervention is implemented or interpreted. The discussion therefore suggested that the “three investigators or research groups” clause may be attempting to control a broader set of dependencies than can be inferred from site count alone.

Function 2: Independence as a Test of Portability

The second function was more closely tied to replication and transfer. Participants repeatedly questioned whether the written intervention protocol contains the information necessary for another team to reproduce effects or whether important features of implementation remain tacit. A moderator described this as the “shorthand” that develops when collaborators make decisions with shared assumptions, and participants agreed that documenting decision rules could make otherwise hidden processes available to outsiders. One participant described the central concern as a noncontact site potentially missing decisions that had been critical to the intervention but were not accessible to the public (S1-P2).

Participants illustrated this problem with an example in which a graduate student might follow the written steps of an intervention but still receive informal coaching from the developer. That coaching could provide examples or judgments that were not contained in the protocol, thereby making the replication dependent on privileged access to the originating team (S2-P6). In this framing, independence served less as a test of researcher impartiality than as a practical test of transferability: whether a person could be given the documented protocol and implement the practice successfully without additional access to the developers. As one participant put it, “the true marker is whether you just take a protocol, hand it to somebody... and then it works” (S2-P9).

Collaboration and Independence Were Viewed as Stage-Dependent

Participants did not treat collaboration and independence as inherently competing goals. Instead, coordinated collaboration was described as particularly useful during intervention development, when researchers may need to share adaptations, solve implementation problems, and make efficient use of limited time and resources. Participants and moderators discussed funder timelines and the concern that requiring sequential, fully independent work from the outset could unnecessarily slow intervention development. One participant also suggested that collaborative decision-making may be valuable when adaptation is intrinsic to the technology, provided the decisions and adaptations are documented.

Taken together, these comments suggested a staged workflow: collaborate to develop, then branch off to replicate. The value of the later separation is not that earlier collaboration was a methodological error; rather, independent replication becomes the point at which the field can test whether the collaboratively developed technology remains effective when separated from the originating network. We use this staged formulation as an analytic shorthand for the pattern described across the two sessions.

Safeguards Addressed Different Functions of Independence

When participants considered situations in which full structural independence was impractical, they proposed safeguards that addressed different functions of independence. Transparency and documentation were primarily directed toward the portability problem. Participants recommended documenting decision rules, adaptations, the reasons for protocol changes, and places where collaborators had jointly resolved implementation questions. They also suggested structured reports and public repositories as ways to preserve information that may not fit in the main manuscript.

Other safeguards were directed more directly toward bias control. Participants suggested third-party verification, independent monitoring, restrictions on communication about certain aspects of data, masking, independent assessors, and site visits. These strategies were described as ways to “protect us from ourselves” when investigators have substantial investment in the intervention. Participants generally treated such measures as partial substitutes rather than complete replacements for unaffiliated replication, particularly when the question is whether the intervention is portable to a genuinely external team.

Evidence Strength Was Viewed as Graded Rather Than Binary

Participants were skeptical that the “3” in the 5-3-20 heuristic should be treated as a universally meaningful threshold. The concern was not that unaffiliated replication was unimportant. Rather, participants questioned whether counting research groups provides the best proxy for the functions they believed independence is intended to serve. One participant described the heuristic as “very arbitrary,” while others returned to the more substantive question of what would happen if one research group conducted many replications but an unaffiliated team could not reproduce the effect.

This distinction led participants toward a graded interpretation of evidence. Rather than treating collaborative evidence as either sufficient or insufficient, participants discussed the possibility of qualifying evidence generated within a coordinated collaboration and then strengthening the evidentiary claim when unaffiliated replications are obtained. The “silver” versus “gold” terminology introduced during the discussion captured this idea: collaborative evidence may support development and preliminary confidence, whereas unaffiliated replication provides additional evidence that the intervention can be transferred beyond the originating network.

Across the two sessions, the discussion therefore returned to two practical questions: Is the evidence vulnerable to bias or ownership effects, and does the intervention remain effective when the only thing transferred is what is explicitly documented? Participants generally viewed transparency and external checks as ways to address portions of these problems, while retaining unaffiliated replication as the strongest direct test of portability.

Independence as a Functional Evidentiary Construct

Across the two focus groups, experts did not reject independence as a meaningful feature of evidence standards. Instead, they questioned whether a simple requirement for “three different investigators or research groups” adequately represents the reasons independence matters. Their discussion reframed independence as a mechanism serving at least two functions: limiting bias and ownership effects, and testing whether an intervention is portable without privileged access to the developers. This functional interpretation helps explain why participants could be skeptical of the literal threshold while continuing to endorse the underlying rationale for cross-team replication. Figure 1 summarizes this functional interpretation and its implications for safeguards and replication.

This interpretation is consistent with the broader logic of single-case evidence standards. Early proposals for establishing evidence-based practices emphasized replication across investigators and studies as a means of strengthening confidence that effects were not idiosyncratic to a single research program (Horner et al., 2005). More recent guidance has retained attention to replication while emphasizing that methodological recommendations should be interpreted in relation to the claims and contexts under consideration rather than treated as inflexible minimum thresholds (BLINDED FOR REVIEW; Ledford et al., 2023). The present findings extend that logic by showing how experts themselves reasoned about the purposes that an independence requirement might serve.

The numerical “3” can therefore be understood as an observable proxy for less readily observable evidentiary properties. Counting investigators or research groups is straightforward; determining whether allegiance has been sufficiently insulated or whether an intervention can be implemented without privileged access to developers is considerably more difficult. The present findings do not establish that the proxy is invalid. Rather, they suggest that interpreting it without attending to the functions it is intended to serve may obscure meaningful differences among ostensibly independent and non-independent replications (Ledford et al., 2023).

Because focus groups generate information through interaction among participants, the patterns identified here should be understood as products of collective discussion rather than as independent votes on a definition of independence (Guerrero & Xicola, 2018). Participants built on, qualified, and sometimes redirected one another’s observations. That feature of the method is particularly relevant to this study because the distinction between bias control and portability became clearer as participants discussed the consequences of different forms of dependence rather than simply selecting a definition from a list.

Different Functions Imply Different Safeguards

The distinction between bias control and portability also suggests that no single safeguard is likely to reproduce all of the evidentiary value associated with independence. External checks such as masking, independent assessors, and third-party monitoring may reduce opportunities for invested investigators to influence measurement or interpretation, but they do not directly establish that another team can

implement the intervention without assistance. Conversely, detailed reporting and decision logs can make tacit procedures more visible, but documentation does not by itself establish that a genuinely external team can reproduce the effect. The expert discussion therefore supports treating these practices as complementary rather than interchangeable.

This distinction also clarifies why transparency infrastructure matters even when unaffiliated replication remains the preferred test of portability. Ledford et al. (2023) emphasized the importance of reporting practices that allow readers and reviewers to understand what was done, under what conditions, and with what limitations. The participants in the present study identified a more specific version of that problem: intervention development can produce decision rules and adaptations that become obvious to collaborators but remain invisible to outsiders. Documenting those decisions may therefore improve the portability of the intervention while also making the basis for replication more inspectable (e.g., Lambert et al., 2026).

A Staged Model May Reconcile Development and Replication

The collaborate-to-develop, branch-to-replicate model may be particularly useful for interventions that are still being refined. Participants described collaboration as valuable for accelerating development, sharing expertise, and resolving adaptation problems. Requiring strong structural independence before a technology has been sufficiently specified may therefore constrain the very collaboration needed to make the technology usable. At the same time, the findings suggest that development-stage dependence should not be allowed to stand in for later evidence of portability. Once the intervention and its decision rules are sufficiently specified, branching to unaffiliated teams creates a distinct evidentiary opportunity: testing whether the technology survives the removal of informal support from the originating network.

This interpretation aligns with a broader distinction between developing an intervention under optimized conditions and determining whether it works under the conditions in which it will actually be used (Fahmie et al. 2023; Lambert et al., 2024). Ledford et al. (2023) and BLINDED FOR REVIEW argued that quality considerations should be separated by domain and interpreted flexibly in relation to research purpose. The present findings suggest that independence may likewise be weighted differently across research stages: coordination may be especially valuable during development, whereas structural separation becomes more consequential when the research question is whether the resulting technology is portable beyond its developers. This stage-dependent interpretation is also consistent with distinctions between intervention development and later efforts to determine whether an intervention is effective beyond optimized development conditions; in such programs, the dominant evidentiary uncertainty can change as a technology moves toward broader use (Ghaemmaghami et al., 2021).

Implications for Evidence Synthesis

The findings also have implications for how evidence syntheses interpret independence. If the primary purpose of independence is to test portability, then a binary code indicating whether a study was conducted by an “independent group” may conceal information that is more directly relevant to interpretation. Reviews could instead record the kinds of interdependence that were present, including shared training lineages, ongoing communication with developers, joint decision-making, and the availability of detailed protocols or adaptation records. Such information would allow reviewers to distinguish evidence that is coordinated because of development needs from evidence that remains dependent on the originating team during later replication.

Independence should also be considered alongside the visibility of null and non-effect findings. Evidence syntheses can be shaped by the selective availability of published studies, particularly when gray literature and studies with non-effects are difficult to locate or are excluded from review. Authors have argued for broader consideration of practice definitions, gray literature, and non-effects in evidence synthesis (Gilroy et al., 2026; Ledford et al., 2021; Tincani et al., 2018). The present findings suggest that independence expectations cannot solve portability uncertainty if unsuccessful replications are systematically less visible. Registered replications, public repositories, and other practices that increase the visibility of replication outcomes may therefore complement independence requirements rather than sit outside them. This concern is especially relevant to independence because the evidentiary value of unaffiliated replication depends on unsuccessful as well as successful replication outcomes being visible; publication practices that selectively favor positive findings can therefore obscure the very portability failures that independent replication is intended to reveal (Tincani & Travers, 2019).

Implications for Evidence Grading and Future Research

The participants’ movement toward “silver” and “gold” evidence suggests a graded interpretation of independence rather than a simple pass-fail rule. Importantly, the present study does not establish a replacement threshold for 5-3-20, nor does it demonstrate that a particular degree of independence is sufficient. Instead, the findings suggest that the evidentiary value of independence may depend on which uncertainty is at stake. A collaborative evidence base may provide strong information about whether an intervention can be developed and refined, while an unaffiliated replication may provide stronger information about whether the resulting technology can be transferred beyond the originating research network.

This distinction yields several testable directions for future research. Studies could examine whether documented interdependence, restricted communication, independent assessment, or unaffiliated replication predict differences in replication success. Evidence reviews could also code these characteristics rather than reducing studies to an independent/not-independent designation. Such approaches would retain the substantive rationale for replication while making the underlying basis for evidence judgments more transparent. Consistent with calls to treat methodological guidance as recommendations that warrant empirical testing, focus groups such as these can provide an initial source of feasibility and meaning-making information for refining evidence standards (Ledford et al., 2023; Tennyson et al., 2026).

Limitations

Several limitations constrain interpretation of these findings. The focus groups were brief, occurred at an invitation-only conference, and relied on self-selection into the sessions. Participants were methodological experts, but we did not collect demographic information beyond professional role and field, and we did not conduct member checking. Attendance was estimated rather than formally recorded, and participants were defined operationally as active speakers; thus, nonvocal agreement or disagreement was not captured. The two sessions also occurred within a single convening, which precluded prospective sampling to saturation. In addition, the moderators introduced the problem context and occasionally summarized apparent agreement, so some formulations may reflect moderator framing as well as participant-generated language. These features make the study most appropriately understood as an exploratory account of expert reasoning about an underspecified methodological issue rather than as an estimate of consensus among SCD researchers.

Future studies could extend the present analysis by sampling researchers with different levels of methodological involvement, including researchers who are not strongly invested in the independence question or who work in fields with different collaboration structures. Additional groups could also examine whether the two functions identified here are stable across research stages and whether participants distinguish between different forms of dependence in consistent ways. Structured data collection could be paired with qualitative discussion in future work if researchers wished to examine both the content of reasoning and the distribution of agreement across participants.

Conclusion

Methodological standards can serve important functions in SCD research by supporting credible inference, replication, reporting, and evidence synthesis. However, recent guidance has cautioned against treating methodological indicators as context-independent thresholds when their relevance depends on the scientific claim and purpose of a study (Ledford et al., 2023; BLINDED FOR REVIEW; BLINDED FOR REVIEW). This distinction may be particularly important for practice-level evidence standards—a numerical requirement such as “three investigators or research groups” is readily counted, but its evidentiary meaning depends on what independence among those investigators is intended to establish.

Experts in these focus groups treated independence as valuable not because separation is inherently preferable, but because separation can serve identifiable evidentiary functions. Participants described independence as a means of reducing bias and ownership effects and, importantly, as a test of whether an intervention remains portable without undocumented support from its developers. They viewed coordinated collaboration as useful during development, proposed transparency and external checks as partial safeguards, and favored unaffiliated replication when portability remained uncertain. These findings suggest that evidence standards may be more informative when the function of independence

is made explicit and when evidence is interpreted in stages rather than reduced to a single categorical threshold.

Declarations

All authors declare no conflicts of interest. Pseudonymized versions of the data that support the findings of this study are available as supporting materials. The research reported here was partially supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A240009 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

We gratefully acknowledge Makayla Hieb for her assistance with focus-group coordination.

Author Contribution

JML & MJR conceptualized and moderated the focus groups. JRL supported focus group development and operationalization and facilitated implementation. JML lead data analysis, wrote the original draft of the manuscript, and lead the editing/revision process. MJR and JRL supported data analysis and the editing/revision process.

Acknowledgement

We gratefully acknowledge Makayla Hieb for her assistance with focus-group coordination.

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

References

1. Burney, V., Arnold-Saritepe, A., & McCann, C. M. (2023). Rethinking the place of qualitative methods in behavior analysis. *Perspectives on Behavior Science*, 46(1), 185-200.
2. Clay, C. J., Bloom, S. E., & Lambert, J. M. (2018). Behavioral interventions for inappropriate sexual behavior in individuals with developmental disabilities and acquired brain injury: A review. *American journal on intellectual and developmental disabilities*, 123(3), 254-282.
3. Contreras, B. P., Hoffmann, A. N., & Slocum, T. A. (2022). Ethical behavior analysis: Evidence-based practice as a framework for ethical decision making. *Behavior analysis in practice*, 15(2), 619-634.

4. Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
5. Fahmie, T. A., Rodriguez, N. M., Luczynski, K. C., Rahaman, J. A., Charles, B. M., & Zangrillo, A. N. (2023). Toward an explicit technology of ecological validity. *Journal of Applied Behavior Analysis*, 56(2), 302-322. <https://doi.org/10.1002/jaba.972>
6. Gilroy, S. P., Tincani, M., & Podlesnik, C. A. (2026). Recommendations and guidance for enhancing systematic reviews of single-case experimental design research. *Journal of Applied Behavior Analysis*, 59(3), e70064.
7. Guerrero, L., & Xicola, J. (2018). New approaches to focus groups. In G. Ares & P. Varela (Eds.), *Methods in consumer research* (Vol. 1, pp. 49-77). Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-102089-0.00003-0>
8. Fahmie, T. A., Rodriguez, N. M., Luczynski, K. C., Rahaman, J. A., Charles, B. M., & Zangrillo, A. N. (2023). Toward an explicit technology of ecological validity. *Journal of Applied Behavior Analysis*, 56(2), 302-322. <https://doi.org/10.1002/jaba.972>
9. Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2021). Functional communication training: From efficacy to effectiveness. *Journal of Applied Behavior Analysis*, 54(1), 122-143. <https://doi.org/10.1002/jaba.762>
10. Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165-179. <https://doi.org/10.1177/001440290507100203>
11. Lambert, J., Fettig, A., & Ledford, J. (2026). Toward a technology of formative triangulation: A mixed methods case study of socially valid intervention programming. *Journal of Behavioral Education*, 1-24. <https://doi.org/10.1007/s10864-026-09651-w>
12. Lambert, J. M., Ledford, J. R., Fahmie, T., & Gast, D. L. (2024). External validity and generalizable knowledge. In *Single Case Research Methodology* (pp. 15-27). Routledge.
13. Ledford, J. R., Lambert, J. M., Barton, E. E., & Ayres, K. M. (2021). The evidence base for interventions for individuals with ASD: A call to improve practice conceptualization and synthesis. *Focus on Autism and Other Developmental Disabilities*, 36(3), 135-147. <https://doi.org/10.1177/10883576211023349>
14. Ledford, J. R., Lambert, J. M., Pustejovsky, J. E., Zimmerman, K. N., Hollins, N., & Barton, E. E. (2023). Single-case-design research in special education: Next-generation guidelines and considerations. *Exceptional Children*, 89(4), 379-396. <https://doi.org/10.1177/00144029221137656>
15. Meadan, H., Alqunaysi, R., Jennings, S., Douglas, S. N., Reilly, A. M., & Xiong, J. (2026). "Nothing Can Replace Authentic Human Interaction": EI/ECSE Professionals' Perception of AI. *Topics in Early Childhood Special Education*, 02711214261435935.
16. Patton, M. Q. (2014). *Qualitative research & evaluation methods* (4th ed.). SAGE Publications.
17. Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Ozkan, S., & Savage, M. N. (2020). Evidence-based practices for children,

- youth, and young adults with autism. Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice.
18. Tennyson, A., May, R., & Mc Tiernan, A. (2026). Establishing core outcome sets to advance evidence synthesis in behavior analysis. *Perspectives on Behavior Science*, 1-20.
 19. Tincani, M., Gilroy, S. P., & Dowdy, A. (2024). Extensions of open science for applied behavior analysis: Preregistration for single-case experimental designs. *Journal of Applied Behavior Analysis*, 57(4), 808-820.
 20. Tincani, M., & Travers, J. (2018). Publishing single-case research design studies that do not demonstrate experimental control. *Remedial and Special Education*, 39(2), 118-128.
 21. Tincani, M., & Travers, J. (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science*, 42(1), 59–75.
 22. Travers, J. C., Tincani, M., Dowdy, A., Forbes, H., & Johnson, J. V. (2026). A registered report on selective reporting in single-case experimental research. *Remedial and Special Education*, 47(1), 36-52.
 23. What Works Clearinghouse. (2022). What Works Clearinghouse procedures and standards handbook, version 5.0. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE).
<https://ies.ed.gov/ncee/wwc/Handbooks>.

Figures

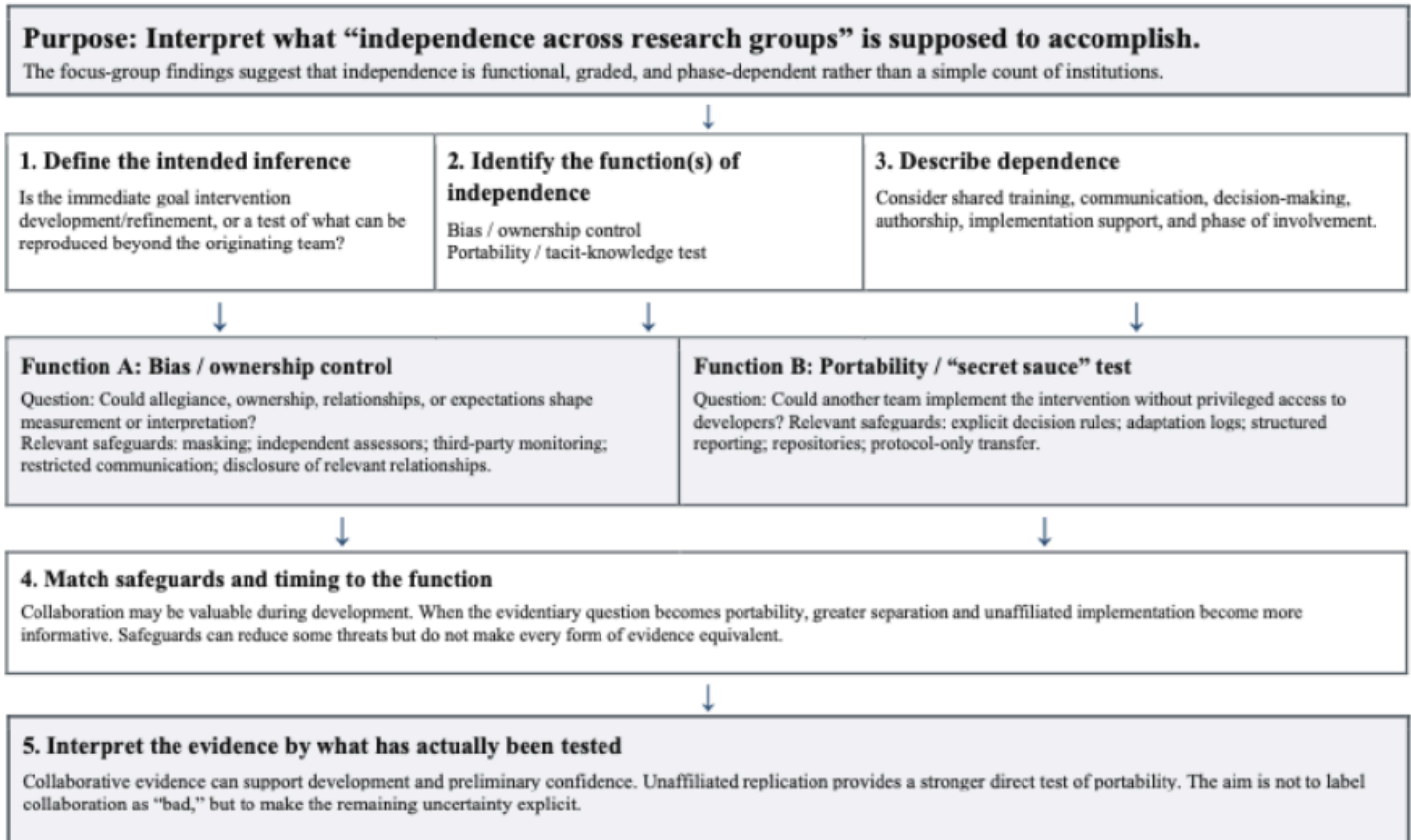


Figure 1

Conceptual framework summarizing expert focus-group judgments about the role of “independence across research groups” in SCD evidence standards.

Note. Independence is treated as functional, graded, and phase-dependent. The framework distinguishes two primary functions identified in the focus groups—bias/ownership control and portability—and links each to corresponding safeguards and to the transition from collaborative development to independent replication.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [WISCC5320SupportingMaterials.docx](#)
- [FullTranscriptSession1.pdf](#)
- [FullTranscriptSession2.pdf](#)