



Challenges and Inclusive Practices in United States Special Education: A Meta-Analytic Research Protocol

DOI: 10.5281/zenodo.21527002

ARE-PEE M. CASTALONE

Special Education Teacher, Christine Duncan's Heritage Academy
rodelcastalone@gmail.com

ALOHA CAPACITE BARRIENTOS

Harmony Science Academy Fort Worth Middle School
simplyteachera@gmail.com

AMELIA S. ROMERO

China Lake Jr. High School, Ridgecrest, CA
romero_amelia70482@yahoo.com

Abstract:

Inclusive education has become the dominant service-delivery philosophy in United States special education, yet implementation remains uneven across districts, disability categories, and teacher preparation pipelines. This paper presents a meta-analytic research protocol examining the challenges associated with implementing inclusive practices for students with disabilities in U.S. schools, and the teacher-, student-, and system-level variables that predict successful implementation. Drawing on the global synthesis literature on teacher beliefs and professional development for inclusion (Krämer et al., 2021; Donath et al., 2023; Dignath et al., 2022; Guillemot et al., 2022; Judijanto, 2024) alongside U.S.-specific evidence on referral disparities, evidence-based practice adoption, and family collaboration (Hosp & Reschly, 2003; Burns & Ysseldyke, 2009; Reid et al., 2004; Buren et al., 2020), this protocol proposes a random-effects meta-analytic design following Cochran (1954), Higgins and Thompson (2002), and Viechtbauer (2010). It specifies search and coding procedures, effect-size conventions, heterogeneity and publication-bias diagnostics (Egger et al., 1997; Duval & Tweedie, 2000a, 2000b; Sterne & Egger, 2005), and a moderator framework organized around teacher-related, student-related, and school-related variables consistent with prior special education synthesis work. Rather than reporting completed statistical results, this paper frames anticipated patterns as testable hypotheses grounded in the existing literature, and discusses their implications for teacher preparation, administrative support structures, and future primary research. The protocol is intended as both a roadmap for conducting the proposed meta-analysis and a template for understanding how such quantitative syntheses are constructed in special education research.

Keywords: inclusive education, special education, teacher attitudes, professional development, meta-analysis, United States

Introduction:

Inclusive education—the practice of educating students with disabilities alongside their nondisabled peers in general education settings to the maximum extent appropriate—has moved from a legal aspiration to a default expectation in United States public schooling. The Individuals with Disabilities Education Act (IDEA) and its least restrictive environment (LRE) mandate have, for nearly five decades, pushed districts toward integrated placements, and



federal reporting through the Office of Special Education Programs has tracked the proportion of students with disabilities served in general education settings since the mid-1990s.

Despite this long regulatory history, the translation of inclusion policy into consistent, high-quality classroom practice remains one of the most persistent challenges in American public education. Teachers report feeling underprepared, administrators struggle to allocate co-teaching and support personnel effectively, and disparities in referral, identification, and disciplinary outcomes continue to fall unevenly across racial, linguistic, and socioeconomic lines. Understanding what predicts successful implementation—and what stands in its way—is therefore a question of both scientific and practical urgency.

The problem is not a shortage of individual studies. Over the past two decades, researchers have produced a substantial body of work on teacher attitudes toward inclusion, the adoption (or non-adoption) of evidence-based instructional practices, disparities in special education referral and identification, and the quality of collaboration between schools and families from historically underserved communities. What has been comparatively scarce, at least until recently, is quantitative synthesis that aggregates these findings into stable, generalizable effect estimates.

Narrative reviews are valuable for capturing nuance and context, but they cannot answer questions such as: how large, on average, is the association between teacher self-efficacy and willingness to implement inclusive practices? Does the size of that association vary by disability category, grade band, or professional development format? Meta-analysis exists precisely to answer such questions, by pooling correlational or comparative effect sizes across independent studies into a single, statistically weighted estimate, while also quantifying how much studies disagree with one another (heterogeneity) and whether the published literature is likely to be missing null or contrary findings (publication bias).

Recent meta-analytic work has begun to fill this gap, though much of it has been conducted with an international rather than a U.S.-specific lens. Krämer, Möller, and Zimmermann (2021) synthesized the literature on inclusive education for students with general learning difficulties, while Guillemot, Lacroix, and Nocus (2022) extended a synthesis of teacher attitudes toward inclusion across two decades, from 2000 to 2020. Dignath, Rimm-Kaufman, van Ewijk, and Kunter (2022) examined not just the direction of teachers' beliefs about inclusive education but the factors that shape those beliefs in the first place, while Donath, Lüke, Graf, Tran, and Götz (2023) asked a more applied question: does professional development actually change how teachers implement inclusive practices, and if so, under what conditions? Judijanto (2024) took a broader, more global-policy view, cataloguing the challenges and strategies reported across national contexts.

Taken together, these syntheses establish that teacher beliefs, self-efficacy, and professional development quality are consistently implicated in inclusion outcomes—but because their samples are drawn from many countries with different funding structures, teacher-preparation systems, and disability categorization frameworks, their pooled estimates may not accurately describe the specific conditions under which U.S. special and general education teachers operate.

This is a meaningful limitation. The United States occupies a distinctive regulatory position: it is bound by a single federal special education law (IDEA) with uniform procedural safeguards, yet implementation is highly decentralized across roughly 13,000 school districts, each with its own hiring practices, caseload norms, and access to specialists. U.S. research has additionally documented patterns that are less prominent in the international synthesis literature—most notably persistent racial and socioeconomic disparities in special education referral rates (Hosp & Reschly, 2003), gaps between recommended and actually-implemented evidence-based instructional practices (Burns & Ysseldyke, 2009), the comparatively poor academic standing of students with emotional and behavioral disturbance relative to other disability categories (Reid et al., 2004), and continuing friction in collaboration between schools and families from nondominant cultural and linguistic communities (Buren et al., 2020). A meta-analysis that



folds these U.S.-specific bodies of evidence into the broader inclusive-practices literature, using consistent statistical conventions, would allow researchers and policymakers to see where U.S. findings converge with the international pattern and where they diverge—and to identify which teacher-, student-, and school-level variables carry the most weight in the U.S. regulatory and demographic context specifically.

The purpose of this paper is to present a meta-analytic research protocol—modeled on established synthesis methodology in special education (e.g., Park & Shin, 2020)—for examining challenges and predictors of inclusive practice implementation in U.S. special education. Rather than reporting completed pooled effect sizes, which would require primary-study data extraction beyond the scope of this paper, this protocol specifies in detail how such a study should be conducted: the search strategy, inclusion and exclusion criteria, coding procedures, effect-size metric, heterogeneity model, and publication-bias diagnostics.

It then discusses, as testable hypotheses grounded in the cited literature, the patterns a completed meta-analysis would likely reveal, and considers the implications of those hypothesized patterns for teacher preparation programs, school-level administrative supports, and future primary research. The guiding research questions are: (a) What is the association between teacher-related variables (attitudes, self-efficacy, professional development) and successful implementation of inclusive practice in U.S. schools? (b) What is the association between student-related variables (disability category, behavior, demographic background) and inclusion-related outcomes, including referral and identification? (c) What school- and system-related variables (administrative support, evidence-based practice adoption, family collaboration quality) moderate these relations?

Literature Review:

The Policy Foundation of Inclusion in the United States

Inclusive education in the U.S. context is inseparable from its statutory foundation. IDEA's least restrictive environment provision does not mandate full inclusion but does require that students with disabilities be educated with nondisabled peers "to the maximum extent appropriate," with removal to separate settings justified only when the nature or severity of a disability makes education in general classrooms, with supplementary aids and services, unsatisfactory. Decades of annual reporting to Congress by the Office of Special Education Programs have tracked placement trends, showing a long-run shift toward general education settings even as the population of students identified for special education services has grown and diversified.

This growth in identified population has occurred alongside broader shifts in diagnostic practice and awareness, comparable to trends documented for other developmental and behavioral conditions during similar periods, which increased the range and complexity of need that general education classrooms are now expected to accommodate. The policy architecture therefore sets an expectation of inclusion, but does not by itself guarantee the instructional capacity, staffing, or administrative coordination required to implement it well—a gap that has motivated a large applied research literature.

Teacher attitude toward inclusive education is arguably the most heavily studied predictor variable in this literature, and for good reason: teachers who view inclusion as unworkable or undesirable are unlikely to implement it faithfully regardless of policy mandates or resource provision. Guillemot, Lacroix, and Nocus (2022) conducted an extended meta-analysis of teacher attitudes toward inclusive education spanning the years 2000 to 2020, allowing them to examine not just the average valence of teacher attitudes but how those attitudes have shifted over two decades of accumulating policy pressure and public discourse.

A twenty-year window is analytically valuable because it can distinguish a stable underlying skepticism from a gradually warming (or cooling) trend as inclusion has become more normalized in professional training and public expectation. Complementing this descriptive picture, Dignath, Rimm-Kaufman, van Ewijk, and Kunter (2022) took



the further step of examining what actually predicts teachers' beliefs about inclusive education moving the field from simply cataloguing attitudes to explaining their sources. Their meta-analytic framing suggests that teacher beliefs are not free-floating opinions but are shaped systematically by variables such as training exposure, contact experience with students with disabilities, and perceived administrative support, which mirrors the logic used in other education syntheses that separate "belief" from "belief antecedent" as distinct constructs worth modeling separately.

A related but distinct body of work has focused specifically on students with general learning difficulties, a population that is large in absolute numbers but has historically received less synthesis attention than more visible categories such as autism or emotional/behavioral disorders. Krämer, Möller, and Zimmermann (2021) addressed this gap directly, synthesizing findings on the inclusive education of students with general learning difficulties. Because this population often does not require the intensive specialized supports associated with more severe or low-incidence disabilities, their inclusion is frequently treated by schools as a lower priority for dedicated resources, even though the sheer number of affected students means that outcomes here have an outsized effect on overall system performance. A synthesis focused on this population is therefore well positioned to reveal whether the predictors of successful inclusion generalize across disability severity or are specific to high-support-need populations.

Knowing that teacher attitudes and beliefs matter is only useful if there is an actionable lever to change them and professional development is the most commonly proposed one. Donath, Lüke, Graf, Tran, and Götz (2023) directly tested this proposition in a meta-analysis asking whether professional development effectively supports the implementation of inclusive education, rather than simply measuring whether it changes self-reported attitudes on a survey.

This distinction between attitude change and implementation change is analytically important, because a large existing literature has shown that positive attitude shifts following a workshop or training session do not reliably translate into changed classroom behavior once teachers return to their regular caseloads and administrative constraints. A synthesis that specifically targets implementation outcomes, rather than attitudinal proxies, therefore provides a more rigorous test of whether professional development investment is justified, and under what design conditions (e.g., duration, coaching component, follow-up support) it is more or less likely to produce durable change.

Global and Comparative Perspectives on Implementation Challenges

Beyond attitude and training-specific syntheses, broader reviews have attempted to catalogue the full range of challenges reported across national contexts. Judijanto (2024) conducted a meta-analysis of challenges and strategies in implementing inclusive education from a global-perspectives standpoint, cataloguing barriers reported across diverse educational systems. This kind of broad-scope synthesis is useful for identifying which challenges appear to be near-universal—such as insufficient teacher preparation time, large class sizes, and inadequate specialist staffing—versus which challenges are more context-dependent, such as specific funding mechanisms or categorical disability definitions that vary substantially by country.

For a U.S.-focused meta-analysis, this comparative literature functions as an important reference point: it allows researchers to ask whether variables that predict implementation success internationally (e.g., self-efficacy, administrative support, professional development quality) carry similar weight in the U.S. system, or whether U.S.-specific structural features—such as its highly decentralized district governance and its explicit procedural due-process framework under IDEA—produce a different pattern of predictors.

A meta-analysis restricted to attitudes and professional development, however, would miss a substantial portion of what makes inclusive practice challenging specifically in the United States: persistent disparities in who is identified for special education services in the first place, and whether the instructional practices actually delivered once a



student is identified reflect the evidence base. Hosp and Reschly (2003) conducted a meta-analysis of racial differences in referral rates for special education intervention or assessment, a line of inquiry that speaks directly to concerns about disproportionality that have shaped federal monitoring requirements and state-level corrective action plans for decades. Referral disparities matter for a synthesis of inclusive practice because inclusion cannot be evaluated in isolation from the identification process that precedes it: a system that over- or under-identifies particular student populations will produce placement and support patterns that reflect identification bias as much as pedagogical quality.

Once students are identified, the quality of instruction they receive becomes the next critical variable, and here again there is reason for concern. Burns and Ysseldyke (2009) examined the reported prevalence of evidence-based instructional practices in special education settings, essentially asking how large the gap is between what research recommends and what is actually delivered in classrooms. A persistent implementation gap of this kind would suggest that even in settings where inclusion is nominally occurring—students physically present in general education classrooms—the instructional quality they receive may not reflect the specialized, evidence-based methods that justify inclusive placement as an educationally sound choice rather than merely a compliance exercise.

Academic outcomes themselves provide a further lens on implementation quality. Reid, Gonzalez, Nordness, Trout, and Epstein (2004) conducted a meta-analysis of the academic status of students with emotional and behavioral disturbance (EBD), a population frequently identified in the broader burnout and behavior-management literature as particularly demanding for teachers and particularly likely to be removed from general education settings despite LRE requirements. Their synthesis provides a benchmark against which the success (or failure) of inclusive placement can be measured for one of the most instructionally challenging disability categories, and offers a natural point of comparison for a broader meta-analysis examining whether disability category moderates the relation between inclusion-supportive teacher variables and actual student outcomes.

Finally, inclusive practice does not occur in a vacuum bounded by the classroom; it depends heavily on the quality of collaboration between schools and the families of students they serve, particularly families from communities that have been historically underserved or marginalized within special education systems. Buren, Maggin, and Brown (2020) conducted a meta-synthesis of the experiences of families from nondominant communities in special education collaboration, surfacing recurring themes of communication breakdown, cultural mismatch, and unequal power dynamics in individualized education program (IEP) meetings. Because family engagement is both a procedural requirement under IDEA and a practical predictor of whether accommodations and supports are implemented consistently across home and school contexts, this line of research provides an essential school-related (and arguably system-related) variable that a comprehensive meta-analysis of U.S. inclusive practice should not omit.

Taken as a whole, the reviewed literature converges on several recurring variable clusters: teacher attitudes and beliefs (Guillemot et al., 2022; Dignath et al., 2022), the design and effectiveness of professional development (Donath et al., 2023), broad implementation barriers across contexts (Judijanto, 2024), disability-population-specific considerations (Krämer et al., 2021; Reid et al., 2004), and system-level equity concerns specific to U.S. governance (Hosp & Reschly, 2003; Burns & Ysseldyke, 2009; Buren et al., 2020). No existing synthesis, however, integrates all of these variable clusters within a single U.S.-focused meta-analytic framework using a consistent effect-size metric and moderator structure. The present protocol is designed to address that integration gap.

Methodology:

This study proposes a random-effects meta-analysis following the general approach used in prior special education synthesis work (e.g., Park & Shin, 2020; Krämer et al., 2021), in which correlational and comparative effect sizes from independent primary studies are converted to a common metric, weighted by inverse variance, and pooled to produce an overall estimate for each predictor variable, disaggregated by outcome category (teacher implementation behavior, student placement/referral outcome, and family/school collaboration quality). A random-



effects rather than fixed-effects model is specified a priori because primary studies in this literature vary substantially in sample composition (grade band, disability category, geographic region, and study design), making it implausible that all studies estimate a single common population effect (Higgins & Thompson, 2002; Viechtbauer, 2005).

Consistent with standard three-step search procedures used in comparable special education meta-analyses, the search would proceed as follows. First, electronic databases—including ERIC, PsycINFO, Education Source, and EBSCOhost—would be searched using three concept fields: (a) terms related to inclusive education (e.g., "inclusion," "inclusive practice," "least restrictive environment," "mainstreaming"); (b) terms related to the U.S. special education population and setting (e.g., "special education teacher," "students with disabilities," "IEP," "IDEA"); and (c) terms related to outcome constructs (e.g., "teacher attitude," "self-efficacy," "professional development," "referral," "evidence-based practice," "family collaboration"). Boolean combinations across these three fields would generate the initial candidate pool. Second, titles and abstracts would be screened by at least two independent coders against a priori inclusion criteria (below), with disagreements resolved through discussion and a calculated inter-rater agreement statistic reported using the percent-agreement method common in this literature. Third, full-text review of the remaining candidate studies would be conducted, with reasons for exclusion logged and reported in a PRISMA-style flow diagram, and a supplementary inverse citation search of included studies' reference lists would be conducted to reduce omission of relevant studies not captured by the database search terms (a step taken explicitly to guard against the file-drawer problem discussed below).

Studies would be included if they (a) sampled U.S. public school settings (K–12) or U.S. teacher-preparation contexts; (b) reported quantitative data—correlational, regression, or group-comparison—linking a teacher-, student-, or school-related variable to an inclusion-relevant outcome (implementation fidelity, placement/referral decision, disciplinary outcome, or family-reported collaboration quality); (c) were published in English in a peer-reviewed journal or as a dissertation; and (d) reported sufficient statistical information (e.g., correlation coefficient, regression coefficient with degrees of freedom, or means and standard deviations for comparison groups) to compute or approximate an effect size. Intervention studies without a comparison or correlational structure suitable for effect-size extraction, non-U.S. samples, and studies focused exclusively on postsecondary or adult populations would be excluded.

Following the coding conventions used in comparable syntheses, two independent coders would extract study-level descriptive information (sample size, disability category represented, grade band, geographic region, publication year, study design) and effect-size-relevant statistics for each reported relation. Variables would be organized into three moderator categories mirroring the structure of the literature review above: **teacher-related** (attitude/beliefs, self-efficacy, professional development exposure, years of experience, certification pathway); **student-related** (disability category, age/grade, behavior severity, race/ethnicity as it relates to referral outcomes); and **school-related** (administrative support, evidence-based practice availability, family/school collaboration quality, caseload and staffing ratios). Inter-rater reliability would be calculated using the percent-agreement formula ($\text{agreements} / [\text{agreements} + \text{disagreements}] \times 100$), with a target threshold of at least 90% agreement prior to independent full-set coding, consistent with reliability standards reported in comparable special education meta-analyses.

Effect Size Calculation

Pearson's r would serve as the common effect-size metric. Where studies report regression results with a single predictor ($df = 1$), r would be derived directly; where multiple regression is reported, partial correlation or standardized beta coefficients would be used as an approximation, flagged as a lower-confidence conversion in sensitivity analyses. Where studies report group comparisons (e.g., inclusive versus non-inclusive placement) rather than continuous correlations, standardized mean differences (Cohen's d) would be computed and converted to r using standard transformation formulas to preserve a single common metric across the full study pool. All correlation coefficients would then be converted to Fisher's Z using the transformation $Z = 0.5 \times \log[(1 + r) / (1 - r)]$, following Borenstein et al.'s widely used convention, to stabilize variance prior to pooling.



Statistical Analysis Plan

Analyses would be conducted in R using the metafor package (Viechtbauer, 2010, 2017), with meta (Schwarzer, 2010) used as a cross-validation tool, following the R computing environment described by the R Development Core Team (2011). For each moderator variable within each outcome category, a random-effects model would be fit using restricted maximum likelihood estimation, with between-study variance (τ^2) estimated following Viechtbauer's (2005) recommended approach for the random-effects model.

Heterogeneity would be assessed using Cochran's Q statistic (Cochran, 1954) and quantified using the I^2 statistic (Higgins & Thompson, 2002), with I^2 values above approximately 50% interpreted as indicating substantial between-study variability warranting further moderator or subgroup analysis. Outlier and influence diagnostics would be conducted following Viechtbauer and Cheung (2010) to identify individual studies exerting disproportionate influence on pooled estimates.

Publication bias would be assessed using multiple complementary methods rather than any single test, consistent with recommendations in the methodological literature (Rothstein et al., 2006). Funnel plot asymmetry would be evaluated visually and statistically using Egger's regression test (Egger et al., 1997; Sterne & Egger, 2005), and Duval and Tweedie's trim-and-fill procedure (2000a, 2000b) would be applied to estimate the number of "missing" studies implied by funnel plot asymmetry and to compute a bias-adjusted pooled effect size for comparison against the unadjusted estimate. Where publication bias indicators are ambiguous, results would be reported with explicit caveats regarding the confidence that can be placed in the pooled estimate.

Moderator and Subgroup Analyses

Given the conceptual importance of disability category, grade band, and geographic/administrative region in the U.S. special education context, planned subgroup analyses would examine whether pooled effect sizes differ significantly by these categorical moderators, using mixed-effects meta-regression models. This would allow the study to test, for example, whether the association between teacher self-efficacy and inclusive implementation is stronger for high-incidence categories (e.g., specific learning disabilities) than for categories associated with more intensive behavioral support needs (e.g., emotional/behavioral disturbance), a question directly motivated by the EBD-specific findings in Reid et al. (2004).

Results and Discussion:

Teacher-Related Variables: Anticipated Patterns

Based on the convergence of findings across Guillemot et al. (2022), Dignath et al. (2022), and Donath et al. (2023), the proposed meta-analysis would be expected to find a moderate, statistically reliable association between teacher attitudes toward inclusion and measures of implementation fidelity, in the direction where more positive attitudes correspond to higher-fidelity implementation of inclusive practices. This expectation follows directly from the extended two-decade synthesis window used by Guillemot and colleagues, which would be well positioned to show whether this association has strengthened, weakened, or remained stable as inclusion has become a more normalized expectation in U.S. teacher preparation over the same period.

If the U.S.-specific subset of studies shows a comparable or larger effect than the international pooled estimate, this would support the inference that U.S. teacher-preparation reforms emphasizing inclusive pedagogy have had a measurable downstream effect on attitude-behavior consistency; if the U.S. subset shows a smaller effect, this would instead suggest that structural features of U.S. school staffing (e.g., large caseloads, limited co-teaching time) suppress the translation of positive attitudes into consistent classroom behavior regardless of teacher belief.



A second, closely related hypothesis concerns the specific mechanism identified by Dignath et al. (2022): if teacher beliefs about inclusion are themselves predicted by identifiable antecedents (training exposure, direct contact experience, perceived administrative support), then a U.S.-focused meta-analysis should be able to decompose the "teacher attitude" variable into these upstream predictors and test which antecedent carries the largest effect size within the U.S. system specifically.

It is anticipated that direct, sustained contact experience with students with disabilities during teacher preparation (student teaching placements, practicum hours) would emerge as a stronger predictor of durable positive attitude than one-time coursework exposure alone, consistent with contact-based attitude change models used broadly in social psychology and imported into the special education attitude literature. If confirmed, this would carry a clear implication for teacher-preparation program design: brief coursework modules on inclusion, without accompanying sustained field placement in inclusive classrooms, may be an insufficient lever for producing durable attitude change.

The professional development question posed directly by Donath et al. (2023)—whether professional development effectively supports implementation, not merely attitude—is anticipated to produce a more qualified pattern than the attitude literature alone would suggest. Specifically, it is hypothesized that the pooled effect of professional development on implementation fidelity would be smaller and more heterogeneous (higher I^2) than the pooled effect of professional development on self-reported attitude change, reflecting the well-documented gap between attitudinal and behavioral outcomes in professional development research more broadly.

Meta-regression is expected to reveal that professional development programs incorporating an ongoing coaching or follow-up component produce substantially larger implementation effects than one-time workshop formats, a moderator finding that would have direct implications for how U.S. districts allocate professional development budgets—suggesting a reallocation away from single-day trainings toward sustained instructional coaching models, even if this requires serving fewer teachers per budget cycle.

Teaching experience and certification pathway are anticipated to show a more modest and possibly non-significant relation to implementation fidelity once self-efficacy and attitude are accounted for in a multivariate framework, mirroring a pattern noted in the broader teacher-variable literature where experience alone often fails to predict practice quality once more proximal psychological variables (self-efficacy, belief) are modeled directly. This would suggest that policy interventions targeting years-of-service thresholds or credentialing requirements alone, without corresponding attention to belief and skill development, are unlikely to move implementation fidelity substantially.

Student-Related Variables: Anticipated Patterns

The student-related moderator analysis is anticipated to reveal the most policy-consequential findings in the entire synthesis, because it directly engages the disparities documented by Hosp and Reschly (2003). It is hypothesized that referral rate disparities associated with student race and ethnicity would remain a statistically significant, non-trivial predictor of special education identification even after controlling for behavior-severity indicators, replicating the core finding that motivated federal disproportionality monitoring requirements.

If the proposed meta-analysis is able to assemble a sufficiently large and recent set of primary studies, a secondary hypothesis worth testing is whether this disparity effect has narrowed, remained stable, or widened over the study period, given that federal and state corrective-action requirements targeting disproportionality have intensified over the past two decades. A stable or widening disparity effect despite intensified regulatory attention would represent an important and sobering finding, suggesting that procedural compliance requirements alone are insufficient to close referral gaps without more direct intervention in local decision-making practices.

Consistent with Reid et al. (2004), disability category is anticipated to function as a strong moderator of academic outcome measures, with students identified under the emotional/behavioral disturbance category showing



significantly poorer pooled academic standing relative to other high-incidence categories such as specific learning disabilities or speech/language impairment. This pattern, if confirmed, would have a direct bearing on the inclusion literature specifically: it would suggest that EBD students are simultaneously the population for whom LRE placement is most difficult to sustain administratively (due to behavior-management demands on general education teachers) and the population showing the poorest academic outcomes regardless of placement setting raising the analytically important question of whether poor outcomes for this group reflect placement decisions, instructional quality, or the severity of the underlying condition itself, a question the proposed meta-analysis could partially disentangle through moderator analysis comparing outcomes across placement settings within the EBD subgroup specifically.

Building on Krämer et al. (2021), it is hypothesized that students with general learning difficulties—despite representing one of the largest identified populations—would show a comparatively smaller and less heterogeneous set of reported predictor-outcome relations than more intensively studied categories such as autism spectrum disorder, reflecting a genuine research-attention gap rather than a true absence of implementation challenge. This would itself be a notable finding: a "low visibility" effect in which the size of the primary-study literature is inversely related to the size of the affected population, a pattern that would argue for a deliberate redirection of future primary research funding toward this underserved population within the disability-category landscape.

School- and System-Related Variables: Anticipated Patterns

At the school and system level, administrative support is anticipated to emerge, consistent with patterns documented broadly in the special education literature, as one of the largest and most consistent predictors across all three outcome domains (teacher implementation, student outcome, and family collaboration quality) mirroring the centrality of "support from school personnel" as a predictor in comparable syntheses of special education professional experience. This would suggest that interventions targeting administrator training and instructional leadership specifically around inclusive practice may produce broader downstream benefits than interventions targeting teachers alone, since administrative support appears to condition the effectiveness of teacher-level variables such as self-efficacy and attitude.

Building on Burns and Ysseldyke (2009), it is hypothesized that the gap between recommended and actually-implemented evidence-based instructional practices would itself moderate the relation between positive teacher attitude and student outcome: that is, positive teacher attitude toward inclusion is anticipated to translate into meaningfully improved student outcomes only in school contexts where evidence-based practices are also reliably available and supported administratively, and to show a substantially attenuated (or even null) relation to student outcome in contexts where such practices are largely unavailable regardless of individual teacher attitude or effort.

If confirmed, this interaction would be among the most policy-relevant findings of the proposed study, because it would imply that teacher-attitude-focused interventions (the most commonly funded type of intervention in this literature) are unlikely to improve student outcomes in the absence of parallel investment in evidence-based practice infrastructure and availability—a finding that would argue against treating attitude change and practice availability as independent, separately fundable initiatives.

Finally, drawing on Buren, Maggin, and Brown (2020), family collaboration quality—particularly for families from nondominant linguistic and cultural communities—is anticipated to show a significant association with sustained implementation fidelity over time, even after controlling for teacher- and school-level variables, reflecting the practical reality that accommodations and supports agreed upon in IEP meetings require consistent reinforcement across home and school contexts to remain effective.

It is further hypothesized that this relation would be moderated by whether family communication is conducted in the family's preferred language and cultural idiom, with studies employing culturally and linguistically responsive



collaboration practices showing significantly stronger family-reported collaboration quality than studies of standard IEP meeting procedures, a moderator finding that would provide empirical support for continued investment in interpreter services, culturally responsive family liaison roles, and flexible meeting scheduling as concrete, implementable school-level interventions rather than aspirational equity language alone.

Heterogeneity and Publication Bias: Anticipated Diagnostic Patterns

Given the wide range of disability categories, grade bands, and geographic contexts represented in the anticipated primary-study pool, it is expected that most pooled effect-size estimates in this synthesis would show at least moderate heterogeneity (I^2 in the 40–75% range), justifying the random-effects modeling choice specified in the methodology above and motivating the extensive moderator analyses proposed throughout this section. Regarding publication bias, it is anticipated—consistent with patterns noted in comparable special education meta-analyses that funnel plot asymmetry would be modest for well-established, frequently studied predictors such as teacher attitude and self-efficacy (where null results are increasingly publishable given the maturity of the research area) but potentially more pronounced for newer or more contested constructs such as family collaboration quality metrics, where the file-drawer effect (the tendency for non-significant findings to go unpublished) may be more severe given the smaller and less established primary-study base.

Trim-and-fill adjustment would be expected to produce only modest downward corrections to the largest, most-studied effect sizes (teacher attitude, self-efficacy, administrative support) but potentially larger corrections for less-studied predictors, a pattern that would itself argue for cautious interpretation of pooled estimates in newer subfields of this literature until additional primary studies accumulate.

Considered together, the hypothesized pattern of findings across teacher-, student-, and school-related domains would suggest a layered model of inclusive practice implementation in U.S. special education: teacher attitude and self-efficacy function as necessary but not sufficient conditions for high-fidelity implementation; administrative support and evidence-based practice availability function as conditioning variables that determine whether positive teacher-level variables translate into improved student outcomes; and persistent identification-stage disparities, along with variable family collaboration quality, function as upstream and parallel processes that shape which students experience inclusive placement in the first place and how consistently accommodations are reinforced once placement occurs.

This layered framing is broadly consistent with the ecological systems perspective used in comparable special education synthesis work, in which individual teacher variables are nested within school-level structures that are themselves nested within broader policy and demographic contexts—suggesting that single-lever interventions targeting only one layer of this system (e.g., teacher training alone, or compliance monitoring alone) are less likely to produce durable improvement in inclusive practice outcomes than coordinated interventions addressing multiple layers simultaneously.

Conclusion:

This paper has presented a detailed meta-analytic research protocol for examining challenges and predictors of inclusive practice implementation in United States special education, integrating an international synthesis literature on teacher attitudes and professional development (Krämer et al., 2021; Guillemot et al., 2022; Dignath et al., 2022; Donath et al., 2023; Judijanto, 2024) with U.S.-specific evidence on referral disparities, evidence-based practice adoption, disability-category-specific academic outcomes, and family collaboration quality (Hosp & Reschly, 2003; Burns & Ysseldyke, 2009; Reid et al., 2004; Buren et al., 2020). The protocol specifies a random-effects meta-analytic design following established conventions in the field (Cochran, 1954; Higgins & Thompson, 2002; Viechtbauer, 2005, 2010; Egger et al., 1997; Duval & Tweedie, 2000a, 2000b), organized around teacher-, student-, and school-related moderator categories consistent with prior special education synthesis work.



Rather than reporting completed statistical findings, this paper has framed the anticipated pattern of results as a set of testable hypotheses, grounded closely in the cited literature: that teacher attitude and self-efficacy will show moderate, reliable associations with implementation fidelity; that these associations will be conditioned by administrative support and the availability of evidence-based practices; that referral disparities linked to student demographic characteristics will persist as a significant predictor of identification outcomes independent of behavior severity; that disability category, and emotional/behavioral disturbance in particular, will moderate academic outcome patterns substantially; and that family collaboration quality, particularly for nondominant communities, will emerge as an independent predictor of sustained implementation success. These hypotheses, and the moderator and subgroup analyses proposed to test them, are offered as a concrete roadmap for the primary-study extraction and statistical analysis that a completed version of this meta-analysis would require.

The protocol carries several limitations worth acknowledging in advance of any completed analysis. First, correlational synthesis of this kind cannot establish causal direction between, for example, administrative support and teacher implementation fidelity; experimental and quasi-experimental primary studies would be needed to test causal claims directly. Second, the moderator structure proposed here, while grounded in the existing literature, may need revision once the actual distribution of available primary studies is known, particularly if certain disability categories or geographic regions are underrepresented in the retrievable literature. Third, as with any meta-analysis, the quality of pooled estimates is bounded by the quality and completeness of the underlying primary-study reporting, and publication bias diagnostics—while informative cannot fully compensate for a genuinely incomplete evidence base in newer subfields such as family collaboration quality measurement.

Notwithstanding these limitations, a completed version of this meta-analysis would offer meaningful value to U.S. special education policy and practice: by quantifying which variables carry the largest, most reliable associations with successful inclusive implementation, it would help teacher-preparation programs, district administrators, and state policymakers direct limited professional development and staffing resources toward the levers most likely to produce durable improvement in inclusive practice for students with disabilities across the United States.

References:

- Buren, M. K., Maggin, D. M., & Brown, C. (2020). Meta-synthesis on the experiences of families from nondominant communities and special education collaboration. *Exceptionality*, 28(4), 259–278.
- Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *The Journal of Special Education*, 43(1), 3–11.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101–129.
- Dignath, C., Rimm-Kaufman, S., van Ewijk, R., & Kunter, M. (2022). Teachers' beliefs about inclusive education and insights on what contributes to those beliefs: A meta-analytical study. *Educational Psychology Review*, 34(4), 2609–2660.
- Donath, J. L., Lüke, T., Graf, E., Tran, U. S., & Götz, T. (2023). Does professional development effectively support the implementation of inclusive education? A meta-analysis. *Educational Psychology Review*, 35(1), Article 30.
- Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98.
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.



Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634.

Guillemot, F., Lacroix, F., & Nocus, I. (2022). Teachers' attitude towards inclusive education from 2000 to 2020: An extended meta-analysis. *International Journal of Educational Research Open*, 3, Article 100175.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.

Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education*, 37(2), 67–80.

Judijanto, L. (2024). Challenges and strategies in implementing inclusive education: A meta-analysis of global perspectives. *Indonesian Journal of Education (INJOE)*, 4(1), 991–1001.

Krämer, S., Möller, J., & Zimmermann, F. (2021). Inclusive education of students with general learning difficulties: A meta-analysis. *Review of Educational Research*, 91(3), 432–478.

Park, E. Y., & Shin, M. (2020). A meta-analysis of special education teachers' burnout. *Sage Open*, 10(2), Article 2158244020918297.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, 38(3), 130–143.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Schwarzer, G. (2010). *Meta: Meta-analysis with R* (R package version 1.6-1). R Foundation for Statistical Computing.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). John Wiley & Sons.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.

Viechtbauer, W. (2017). Package 'metafor'. The Comprehensive R Archive Network.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125.