Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Who reports change? A PRISMA 2020 systematic review and stratified multi-informant meta-analysis of digital self-directed parenting programmes for children aged 2–12

Alexander Kuznetsov

alex@5steps.edu.sg

Center for Educational & Psychological Assessment, 5 Steps Academy Private Limited
https://orcid.org/0009-0002-6652-3544

Systematic Review

Keywords: digital parenting programmes, multi-informant discrepancy, parent-proxy report, child self-report, masked behavioural observation, wearable accelerometry, within-trial comparison, stratified meta-analysis

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11739046/v1

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

Additional Declarations: The authors declare no competing interests.

Who reports change? A PRISMA 2020 systematic review and stratified multi-informant meta-analysis of digital self-directed parenting programmes for children aged 2–12

Running Head: WHO REPORTS CHANGE IN DIGITAL PARENTING TRIALS

Author: Alexander Kuznetsov, Ph.D. (5 Steps Academy, Singapore; ORCID: 0009-0002-6652-3544)

Abstract

Background: Self-directed digital parenting programmes are widely promoted as scalable public-health solutions for childhood disruptive behaviour, anxiety, and lifestyle risk. However, parents in such trials simultaneously receive the unmasked intervention, enact its behavioural strategies at home, and serve as the primary proxy informants of child change. Comparing separate *p*-values across reporters or pooling incompatible informant channels obscures whether reported gains generalise beyond the participating caregiver's perspective.

Methods: Following PRISMA 2020 guidelines (no prospective protocol registration is claimed), this review integrated an archival thematic register (8,047 records) with programmatic multi-database API harvesting across four bibliographic search APIs (OpenAlex, Crossref, Semantic Scholar, and Europe PMC/PubMed) paired with programmatic Unpaywall open-access resolution (2,493 records through 8 November 2026) and verified 144 full-text documents. Strict cohort-family linkage prevented unit-of-analysis duplication across 38 qualitative/methodological-audit cohorts (56 linked reports) and 27 unique randomised controlled trial (RCT) cohorts (28 non-overlapping intervention contrasts; N = 10,157 randomised families in compared arms; 51 channel-specific effect estimates). Subgroup A stratified random-effects restricted maximum likelihood (REML) meta-analyses separated six informant channels across superiority and active-comparator strata. Subgroup B synthesised within-trial paired contrasts ($\Delta g = g_\text{parent} - g_\text{second}$) across 19 multi-informant cohorts (20 unique contrasts; 22 channel pairs) using exact variance–covariance propagation across sensitivity correlations r = .20, .35, and .50, alongside REML meta-regressions testing channel independence and analytical denominator divergence ($N_\text{second} / N_\text{parent}$).

Results: In superiority RCTs (24 cohorts, 25 arms), unmasked parent-proxy ratings yielded a highly robust small-to-moderate benefit (k = 25, N = 7,412, g = +0.252, 95% CI [+0.187, +0.316], $p = 1.84 \times 10^{-14}$, I^2 = 54.7%), moderating downward from therapist/coach-assisted packages (g = +0.406) to strictly self-directed programmes (k = 17, g = +0.211) and showing small-study asymmetry (Egger p = .0027;

PEESE g = +0.143). Pooled independent non-parent channels, however, yielded a null effect with zero heterogeneity (k = 12, N = 2,406 channel-specific assessments across 11 cohorts [2,054 unique children], g = +0.023, 95% CI [−0.051, +0.096], p = .545, I^2 = 0.0%): independent classroom/portal child self-report (k = 3, g = +0.037, p = .578), teacher report (k = 2, g = −0.031, p = .872), masked behavioural video observation (k = 3, g = +0.004, p = .973), and wearable accelerometry/blinded anthropometry (k = 4, g = +0.014, p = .808). Apparent corroboration occurred exclusively when secondary channels remained epistemically coupled to the unmasked parent—via clinician interviews of the same parent (Channel 6: k = 6, g = +0.419; pure self-directed k = 3, g = +0.212) or a child questionnaire link forwarded through the parent's chat application with parental reading assistance (Channel 2b: g = +0.401). Within-trial paired contrasts confirmed a systematic informant wedge in superiority trials (k = 11, Δg = +0.266 [+0.167, +0.364], $p = 1.29 \times 10^{-7}$ at r = .35; +0.258 at r = .20; +0.275 at r = .50), persisting when analytical denominators were balanced (intercept at N_S/N_P = 1.0: β_0 = +0.253, $p = 3.25 \times 10^{-6}$; joint 3-moderator R^2 = 88.0%) and reversing sign when in-person staff delivery outperformed online self-help on parent ratings but not masked home video (Δg = −0.414, p = .0028) or teacher ratings (Δg = −0.296, p = .024). Sixteen primary-study numerical discrepancies were audited and quarantined.

Conclusions: Across 27 RCT cohorts, child improvements reported by unmasked parents in self-directed digital programmes do not transfer to independent child, teacher, masked observational, or wearable sensor channels. Scorer blinding does not cure informant unblinding when clinicians interview the treated parent, and digital survey forwarding can contaminate child self-reports. Future trials must separate informant channels, maintain balanced multi-informant denominators, and report cross-reporter covariances.

Keywords: digital parenting programmes; multi-informant discrepancy; parent-proxy report; child self-report; masked behavioural observation; wearable accelerometry; within-trial comparison; stratified meta-analysis

Key Points and Clinical / Methodological Message

What is known

- Previous meta-analyses of online and digital parenting interventions report small-to-moderate improvements in child behaviour and emotional symptoms, relying overwhelmingly on unmasked caregiver questionnaires or pooling multiple informants indiscriminately.
- Methodological theory highlights that comparing separate statistical significance levels across informants neither tests a within-trial reporter difference nor isolates informant identity from

construct, setting, scoring, or sample attrition.

What is new

- Across 27 verified RCT cohorts (N = 10,157 in compared arms; 51 effect estimates), unmasked parent-proxy reports show consistent child improvement (g = +0.252; pure self-directed g = +0.211), whereas every genuinely independent witness channel—independent child self-report (g = +0.037), teacher ratings (g = −0.031), masked video coding (g = +0.004), and wearable sensors/blinded anthropometry (g = +0.014)—converges on zero (pooled k = 12, g = +0.023, I^2 = 0.0%).
- Within-trial covariance-adjusted contrasts confirm an informant wedge of Δg = +0.266 ($p = 1.29 \times 10^{-7}$) that survives balanced sample denominators (Alexandrou et al., 2023; DuPaul et al., 2026; Reardon et al., 2026b), refuting the claim that null objective findings merely reflect clinic video attrition.
- Blinded clinicians scoring interviews of the unmasked parent (g = +0.419) and child links delivered through the parent's messaging app (g = +0.401) reproduce the parent-channel signal rather than providing independent corroboration.

Clinical and methodological implications

- Commissioners, clinical guidelines, and trialists should interpret parent-reported gains from self-directed digital programmes as changes in caregiver perception and dyadic appraisal rather than verified cross-contextual child symptom remission.
- Future digital parenting RCTs should administer child and teacher portals independently of caregiver devices, collect home-based masked observations or passive sensor data across the full randomised sample, and publish matched dyadic denominators and cross-informant correlation matrices.

1. Introduction

1.1. The epistemic architecture of outcome reporting in digital parenting trials

Digital and self-directed parenting programmes (delivered via interactive web portals, smartphone applications, or automated messaging curricula) occupy a central position in contemporary child mental health and behavioural prevention strategies (Dunn et al., 2024; Florean et al., 2020; Grajdan et al., 2024; Morgan et al., 2017; Reardon et al., 2026b; Spencer et al., 2020). Because such interventions bypass scarce specialist clinician capacity, even modest standardised effects are frequently cited to justify population-level commissioning for early-onset conduct problems, attention-deficit/hyperactivity disorder (ADHD), childhood anxiety, and preschool movement or dietary risk (Alexandrou et al., 2023; DuPaul et

al., 2026; Reardon et al., 2026a; Sourander et al., 2016). Yet the evidentiary foundation of digital parenting trials rests upon a structural measurement asymmetry. In a parent-mediated digital programme, the participating caregiver simultaneously occupies four distinct roles: (a) recipient of the unmasked intervention curriculum, (b) agent expected to enact new parenting routines at home, (c) respondent reporting on personal distress, self-efficacy, or parenting practices, and (d) primary proxy informant evaluating changes in the child's behaviour or emotional symptoms (Åhlén et al., 2022; Döpfner et al., 2025; Prinz et al., 2022; Sanders et al., 2012).

Treating all child-focused scores in a trial as interchangeable indicators of a single latent treatment effect conflates fundamentally different epistemic operations (De Los Reyes et al., 2015). Rigorous interpretation requires distinguishing at least six structural dimensions of every outcome estimate: (1) the person described (parent skill or wellbeing versus child symptom or impairment); (2) the first-hand informant supplying observational testimony; (3) the scorer who converts that testimony into a numerical rating; (4) the instrument, item set, and directional scaling; (5) the ecological setting and assessment occasion (home routine, structured clinic task, or classroom); and (6) the exact analytical denominator of families contributing valid scores to the model (Creswell et al., 2024; Döpfner et al., 2025; Ghaderi et al., 2018; Prinz et al., 2022; Sim et al., 2020). Neither the proliferation of questionnaires within a trial nor the professional credential of an outcome assessor establishes independent corroboration of child change. When a masked psychologist or psychiatrist rates a structured diagnostic telephone interview conducted exclusively with the unmasked participating parent (e.g., ADIS-P, DCL-EXT, or OAPA), the scoring procedure is blinded, but the underlying evidentiary channel remains entirely dependent on the same unblinded caregiver's verbal or written account (Donovan & March, 2014; Döpfner et al., 2025; Morgan et al., 2017; Reardon et al., 2026a). Conversely, when a masked coder scores recorded parent–child interaction segments or a child wears a triaxial accelerometer, the information pathway bypasses parental subjective appraisal, yet the resulting estimate may still reflect a task-restricted setting or a severely depleted completer subsample (Baker et al., 2017; Peng et al., 2025; Prinz et al., 2022; Sanders et al., 2012).

1.2. Why prior syntheses and side-by-side p-values fail to resolve informant divergence

Previous quantitative syntheses of online and technology-assisted parenting interventions have documented encouraging overall effects on child behaviour and emotional adjustment (typically d or g = 0.25–0.45), while either restricting extraction to parent-rated scales or averaging parent, clinician, teacher, and observational measures into undifferentiated composite estimates (Florean et al., 2020; Grajdan et al., 2024; Spencer et al., 2020). Within primary trial reports, multi-informant discrepancies are routinely

dismissed through two statistically invalid heuristics. First, authors frequently juxtapose a statistically significant parent-reported p-value ($p < .05$) alongside a non-significant child, teacher, or observational p-value ($p > .05$) without estimating the within-trial difference itself (Havighurst et al., 2024; Sanders et al., 2012; Vigerland et al., 2016). As established in classical meta-analytic and biostatistical theory, the difference between significance and non-significance is not itself statistically significant (Borenstein et al., 2009): testing whether reporter A yields a larger intervention effect than reporter B within the same assigned contrast requires propagating their joint sampling distribution, $\mathrm{Var}(A - B) = \mathrm{Var}(A) + \mathrm{Var}(B) - 2\,\mathrm{Cov}(A, B)$. Separate standard errors do not identify $\mathrm{Cov}(A, B)$, which depends both on the cross-informant correlation r and on the fraction of overlapping dyads contributing to both denominators (Creswell et al., 2024; Ghaderi et al., 2018; Sim et al., 2020).

Second, when masked clinic observations or wearable device subsamples yield null effects beside positive parent questionnaires—as in early Triple P Online evaluations where clinic video subsamples retained fewer than half as many families as online parent surveys (Baker et al., 2017; Sanders et al., 2012)—investigators commonly attribute the null second channel to reduced statistical power or recording attrition rather than genuine informant divergence. Testing that competing explanation requires examining trials in which the independent witness channel achieves sample retention equal to or exceeding the parent questionnaire frame ($N_{second} / N_{parent} \approx 1.0$), such as school-based classroom child/teacher assessments (Reardon et al., 2026b), full-sample home video recordings (DuPaul et al., 2026; Prinz et al., 2022), and routine primary-care anthropometry or accelerometry (Alexandrou et al., 2023; Nyström et al., 2017; Peng et al., 2025).

1.3. Objectives of the present PRISMA 2020 review and stratified meta-analysis

Foundational multi-informant parent–child trials—including Online Support and Intervention (OSI) (Creswell et al., 2024; Taylor et al., 2022; Williams et al., 2022), iComet versus Family Check-Up (Ghaderi et al., 2018), and Parenting Resilient Kids (PaRK) (Fernando et al., 2018; Sim et al., 2020)—illustrate why complete primary texts, protocols, and statistical supplements are required to verify matched dyadic denominators before quantitative synthesis. By integrating the archival thematic register with programmatic multi-database harvesting through 8 November 2026 and verifying full primary texts and statistical supplements across 144 documents, the present study executes a comprehensive PRISMA 2020 systematic review and stratified multi-informant meta-analysis (Page et al., 2021).

Specifically, this review addresses four pre-specified questions: (1) In Subgroup A, how do pooled randomised intervention effects compare across six epistemically distinct informant channels—unmasked

parent-proxy questionnaires (Channel 1), child self-reports (Channel 2, decomposed into independent child portals/classrooms [2a] versus parent-forwarded links [2b]), independent teacher ratings (Channel 3), masked behavioural video coding and direct child tasks (Channel 4), objective wearable sensors and blinded anthropometry (Channel 5), and clinician-scored interviews of the same parent (Channel 6)—when superiority trials are strictly separated from active-comparator noninferiority designs? (2) In Subgroup B, what is the magnitude and precision of the within-trial paired informant contrast ($\Delta g = g_{parent} - g_{second}$) across sensitivity cross-reporter correlations ($r = .20, .35, .50$), and does Δg attenuate when analytical denominators are balanced ($N_{second} / N_{parent} \rightarrow 1.0$)? (3) How do human clinical guidance versus purely self-directed delivery, immediate post-test versus follow-up timing, and small-study publication bias moderate channel-specific estimates? (4) What primary-study statistical reporting discrepancies—including standard errors mislabelled as standard deviations, inverted effect-size signs, and point estimates lying outside printed confidence intervals—must be audited and quarantined to prevent distorted meta-analytic pooling?

2. Methods

2.1. Protocol architecture, eligibility criteria, and strict unit separation

This systematic review and stratified meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement (Page et al., 2021). Throughout screening, extraction, risk-of-bias appraisal, and quantitative synthesis, strict unit separation was enforced across bibliographic records, published report documents, programme families, recruited participant cohorts, and outcome-specific analysis denominators. Two or more publications drawing upon the same randomised cohort—including protocol-to-outcome pairs (e.g., Taylor et al., 2022 and Williams et al., 2022 with Creswell et al., 2024; Fernando et al., 2018 with Sim et al., 2020; Zhou et al., 2024 with Peng et al., 2025 and Zhou et al., 2026), economic evaluations (Yu et al., 2026 with Reardon et al., 2026a), within-arm moderation analyses (Högström et al., 2013 and 18-month uncontrolled follow-up Högström et al., 2015 with Enebrink et al., 2012; Chan et al., 2026 and Sloan et al., 2026 with Buchanan et al., 2025), and sequential follow-up waves (Delisle Nyström et al., 2018 with Nyström et al., 2017)—were linked as a single shared study family and never counted as independent replications. Conversely, distinct trials conducted within the same overarching intervention brand (e.g., pilot Project PEAK: DuPaul et al., 2018 versus full-scale R01 Project PEAK: DuPaul et al., 2026; MINISTOP 1.0: Nyström et al., 2017 versus MINISTOP 2.0: Alexandrou et al., 2023; clinical Co-CAT OSI: Creswell et al., 2024 versus school-screened MYCATS: Reardon et al., 2026a and iCATS-i2i:

Reardon et al., 2026b) were verified via trial registration numbers, recruitment dates, and geographic frames before being treated as independent cohorts.

No prospective protocol registration is claimed; complete two-stream PRISMA 2020 search syntax, deduplication counts, cohort-family linkage rosters, and source-discrepancy rules are documented in Online Supplement S1.

Substantive eligibility criteria required: (a) Population: parents or primary caregivers of children aged 2–12 years (mean cohort age within 2.0–12.0 years; mixed-age cohorts spanning 18–35 months: Sandborg et al., 2025 or 10–13 years: Ghaderi et al., 2018; Stormshak et al., 2019 were retained with explicit coding and subjected to strict age-2–12 sensitivity exclusion, whereas cohorts with mean age > 12.0 years: Connell et al., 2026; Connell & Stormshak, 2023; McWhirter et al., 2026 or severe motor/intellectual neurodisability were excluded from quantitative pooling); (b) Intervention: digital, web-based, or mobile-application parenting programmes containing structured self-directed caregiver modules targeting child behaviour, anxiety/internalising symptoms, emotion socialisation, or movement/dietary/sleep routines (pre-digital printed workbooks or mailed DVDs without web/app modules were excluded; adjunctive telephone, chat, or blended support was recorded and stratified rather than conflated with purely unguided self-help); (c) Design: randomised controlled trials (individual or cluster-randomised) assigning families to a digital parenting arm versus waitlist, usual care, minimal information, or an active clinician-delivered comparator (single-arm open trials: Jent et al., 2025; Mishina et al., 2025, response-escalated small-N designs: Ding et al., 2025, and single-cluster-per-arm studies: Ghani et al., 2026 were retained exclusively in the qualitative methodological audit and excluded from RCT effect pooling); and (d) Outcomes: at least one post-randomisation measure of child behavioural, emotional, self-regulatory, or health/movement status, strictly separated from parent-own mental health, parenting confidence, knowledge/attitudes, or parental accommodation (Åhlén et al., 2022; Chan et al., 2026; Sung et al., 2021).

2.2. Multi-database search strategy and full-text verification accounting

Candidate evidence was assembled from two complementary sources (Figure 1). First, the search strategy examined an archival thematic register comprising 8,047 bibliographic records (including 1,271 candidate records requiring source-stage adjudication and 27 foundational seed items). Second, to eliminate retrospective selection bias and capture all 2015–2026 multi-informant trials, programmatic Boolean API searches were executed across four bibliographic search APIs—OpenAlex, Crossref, Semantic Scholar, and Europe PMC/PubMed—paired with programmatic Unpaywall open-access full-text resolution,

retrieving 2,493 records across targeted combinations of digital/online/mobile parent training, child behaviour/anxiety/health outcomes, and multi-informant or objective channels. Following automated DOI/title deduplication across the combined 10,540 records (removing 2,612 duplicates), 7,928 unique bibliographic records underwent title, abstract, and metadata screening, excluding 7,760 out-of-scope records (adult/adolescent self-therapy, perinatal/infant care < 2 years, severe motor/IDD rehabilitation, non-digital group/clinic formats, cross-sectional surveys, and unrelated digital education).

Full texts were sought for 168 potentially eligible reports and reference reviews. Complete full-text verification was achieved for 144 documents: 126 open-access full-text XML/PDF publications and trial statistical analysis plans plus 17 newly acquired publisher/accepted author PDFs and the crucial online statistical supplement (Appendix S1) of Prinz et al. (2022). Cross-checking against three Q1 systematic reviews and meta-analyses in target journals (Florean et al., 2020; Grajdan et al., 2024; Spencer et al., 2020) confirmed that zero eligible multi-informant RCTs for children aged 2–12 remained unretrieved. Of the 24 candidate paywalled citations audited during acquisition, 17 primary and linked full texts were obtained and verified, 4 were excluded upon registry/protocol verification due to pre-digital printed workbook/DVD modality or severe neurodisability populations, 2 were secondary meta-analyses without unique primary cohorts (Florean et al., 2020; Spencer et al., 2020), and 1 was a single-informant parent-only support comparison duplicating three-arm guidance contrasts already represented by larger multi-informant cohorts (Döpfner et al., 2025; Stormshak et al., 2019). Following full-text inspection of the 144 verified documents, 88 were excluded from primary cohort synthesis (secondary reviews/commentaries, adolescent-mean cohorts, parent-only psychological distress trials, or protocol-only records without achieved outcomes), leaving 56 linked primary and methodological reports representing 38 unique cohorts for qualitative methodological synthesis and 27 unique RCT cohorts (28 intervention contrasts; 51 effect estimates) for quantitative meta-analysis.

PRISMA 2020 Stage	Primary Evidence Stream & Count	Exclusions, Deduplication & Synthesis Accounting
1. Identification	Archival thematic register: *n* = 8,047 records (incl. 27 foundational seed items & 1,271 adjudication candidates); Programmatic 4-API harvest (OpenAlex, Crossref, Semantic Scholar, Europe PMC/PubMed) paired with Unpaywall open-access resolution through 8 Nov 2026: *n* = 2,493 records; Combined total: *N* = 10,540 records	Automated DOI and normalised-title deduplication removed *n* = 2,612 duplicate records prior to screening
2. Screening	Unique bibliographic records screened at title, abstract, and metadata level (ages 2–12 digital parenting PICO):	Excluded at title/abstract screening: *n* = 7,760 out-of-scope records (adult/adolescent self-therapy,

PRISMA 2020 Stage	Primary Evidence Stream & Count	Exclusions, Deduplication & Synthesis Accounting
	n = 7,928	infant care < 2y, severe motor/IDD rehabilitation, non-digital clinic/group formats, non-RCT surveys, unrelated topics)
3. Retrieval	Potentially eligible reports sought for full-text retrieval: *n* = 168	Reports not retrieved as primary digital RCT full texts: *n* = 24 (0 eligible multi-informant RCTs for ages 2–12 missing; registry/audit exclusions: printed workbook/DVD or severe motor/IDD protocols, secondary meta-analyses, duplicate single-informant support report)
4. Eligibility	Verified full-text documents inspected line-by-line: *n* = 144 (126 open-access publisher XML/PDFs & SAPs + 17 publisher/accepted author PDFs + 1 statistical supplement: Prinz et al., 2022, Appendix S1)	Full-text reports excluded from cohort synthesis: *n* = 88 (secondary reviews/commentaries, adolescent-mean cohorts > 12y, parent-only distress trials, uncompleted protocols)
5. Included — Qualitative Audit	Studies included in qualitative & methodological-audit synthesis: **38 unique cohorts** (**56 linked reports**)	Strict cohort-family unit separation enforced; 16 primary-study statistical discrepancies audited; 11 construct/age/design boundary cohorts documented
6. Included — Quantitative Meta-Analysis	Studies included in stratified quantitative meta-analysis: **27 unique RCT cohorts** (**28 non-overlapping contrasts**; ***N* = 10,157** randomised families in compared arms; 10,449 total; **51** channel-specific effect estimates; **22** within-trial paired contrasts)	**Subgroup A (51 estimates):** Ch. 1 Unmasked Parent-Proxy (*k* = 25, *g* = +0.252; pure self-dir. +0.211; PEESE +0.143) vs. Pooled Independent Channels 2a+3+4+5 (*k* = 12, *g* = +0.023, I^2 = 0.0%) vs. Parent-Mediated Ch. 6 (+0.419) & Ch. 2b (+0.401). **Subgroup B (22 paired contrasts across 19 cohorts / 20 arms; *r* = .20, .35, .50):** Superiority Parent minus Independent Witness (*k* = 11, *r* = .35) Δg = +0.266 (*p* = 1.29×10^{-7}); balanced-denominator β_0 = +0.253; joint R^2 = 88.0%

Figure 1. PRISMA 2020 flow diagram of multi-database record identification, deduplication, full-text verification (*n* = 144 primary XML/PDF documents and statistical supplements), cohort-family linkage (56 reports mapped to 38 unique qualitative/methodological-audit cohorts), and stratified quantitative synthesis (27 unique RCT cohorts; 28 intervention contrasts; 51 informant-specific effect estimates; 22 within-trial paired informant contrasts; high-resolution 300-DPI vector/raster diagram archived in T02_PRISMA_2020_Flowchart.png).

2.3. Informant channel taxonomy, construct separation, and source-discrepancy quarantine

Every eligible child-focused outcome in the 27 RCT cohorts was classified into one of six mutually exclusive informant channels prior to effect pooling: Channel 1 (Unmasked Parent-Proxy Report: caregiver questionnaires rating child externalising, ADHD, internalising/anxiety, or dietary/screen/sleep behaviours completed by the parent assigned to the digital programme); Channel 2 (Child Self-Report: standardised symptom or functional impact scales completed by the child, subdivided into 2a Independent

Child Self-Report administered in school classrooms or via an independent child login portal: Creswell et al., 2024; Ghaderi et al., 2018; Reardon et al., 2026b; Sim et al., 2020; Vigerland et al., 2016 versus 2b Parent-Forwarded Child Link where the survey URL was transmitted to the unmasked parent's chat application with explicit parental reading assistance: Chanpen et al., 2026); Channel 3 (Independent Teacher Report: classroom educator ratings on standardised behaviour or anxiety scales: Havighurst et al., 2024; Prinz et al., 2022; Reardon et al., 2026b); Channel 4 (Masked Behavioural Observation: structured clinic or home parent–child interaction videos coded by assessors masked to group assignment and assessment wave, plus direct masked child self-regulation testing: Baker et al., 2017; DuPaul et al., 2026; Prinz et al., 2022; Sanders et al., 2012); Channel 5 (Objective Wearable Sensors and Blinded Anthropometry: multi-day waist or wrist ActiGraph accelerometry, air-displacement plethysmography [BodPod], or standardised clinical height/weight measurements: Alexandrou et al., 2023; Hammersley et al., 2019; Nyström et al., 2017; Peng et al., 2025); and Channel 6 (Clinician-Scored Parent Interview: diagnostic or severity ratings assigned by a blinded clinician or psychologist based on a structured telephone or online interview conducted with the unmasked participating parent: Donovan & March, 2014; Döpfner et al., 2025; Morgan et al., 2017; Reardon et al., 2026a; Vigerland et al., 2016).

When primary reports contained internal mathematical or typographical contradictions between text, tables, and confidence intervals, every conflict was recorded verbatim rather than silently selecting a favourable figure (Section 3.3 and Table 3). Specifically: (a) where tables printed estimated marginal standard errors (SE) rather than standard deviations (SD)—as verified in Prinz et al. (2022, Appendix S1), Baker et al. (2017, Tables 2–3), and Sourander et al. (2016, Table 2)—standardised effects were computed strictly from baseline pooled SDs (Morris, 2008, d_ppc2) or verified ANCOVA F-statistics and model-based Cohen's d tables, preventing artificial $\sqrt{n}$ inflation; (b) where printed point estimates inverted directional signs relative to their own 95% confidence intervals and raw group means (e.g., Sim et al. Table 2 printing d = −0.10 (−0.11, 0.31) whose exact midpoint is +0.10: Sim et al., 2020; Döpfner et al., 2025, abstract printing d = 0.34 (−0.59, −0.09)), signs were harmonised from the verified raw group trajectories so that positive Hedges' g (+g) uniformly denotes beneficial child change favouring the self-directed digital intervention; and (c) where printed point estimates lay mathematically outside both bounds of their own printed 95% confidence intervals and could not be algebraically reconciled from raw arm scores—specifically three ActiGraph sleep rows in Peng et al. (2025, Table 3)—those corrupted sub-estimates were quarantined and excluded from quantitative pooling.

2.4. Statistical synthesis, within-trial covariance propagation, and risk-of-bias appraisal

All continuous outcomes were converted to bias-corrected standardised mean differences (Hedges' g =

J(df) · d, with exact small-sample correction J(df) = 1 − 3/(4df − 1)) and exact sampling variances V_g (Borenstein et al., 2009). Binary diagnostic outcomes in Channel 6 (e.g., ADIS-P or OAPA anxiety disorder presence: Donovan & March, 2014; Morgan et al., 2017; Reardon et al., 2026a) were transformed from log-odds ratios and their cluster- or model-adjusted confidence intervals to Hedges' g via the Chinn (2000) / Cochrane scaling factor ($g = -\ln(OR) \cdot \sqrt{3}/\pi$, $SE_g = SE_{\ln(OR)} \cdot \sqrt{3}/\pi$). When a cohort reported multiple eligible child subscales within the same informant channel (e.g., ECBI Intensity and Problem plus SDQ Conduct: Enebrink et al., 2012; Sanders et al., 2012), a single synthetic cohort-by-channel composite was formed prior to meta-analysis using the Borenstein et al. (2009) correlated-outcomes variance formula assuming within-informant inter-scale correlation $\rho = .50$: $V_{comp} = (1/m^2) [\Sigma V_i + \Sigma_{i \neq j} \rho \sqrt{(V_i V_j)}]$, ensuring each trial arm contributed at most one estimate per informant channel. For the three-arm WASH trial comparing unguided web self-help and telephone-assisted web self-help against a shared usual-care control (Döpfner et al., 2025), the shared control denominator was split 50/50 between the two contrasts (Cochrane Handbook §23.3.4) to eliminate double-counting while preserving within-trial guidance comparisons. For three-arm trials containing a non-digital face-to-face comparator alongside an online arm and a waitlist control (DuPaul et al., 2018, 2026; Chanpen et al., 2026; Stormshak et al., 2019), Superiority Stratum A1 isolated the Online-versus-Waitlist denominators so face-to-face participants never inflated online weights. Cluster-randomised school trials (MYCATS, 95 schools, PAS ICC = .015, ADIS-P ICC = .04: Reardon et al., 2026a; iCATS-i2i, 84 schools: Reardon et al., 2026b) entered synthesis via their published multilevel cluster-adjusted mixed-model standard errors.

In Subgroup A, channel-specific effects were pooled via random-effects Restricted Maximum Likelihood (REML) meta-analysis (cross-checked against DerSimonian–Laird τ^2_{DL} and Knapp–Hartung t-distribution intervals), reporting Cochran's Q, I^2, and τ^2_{REML}. Superiority RCTs (Stratum A1) and active-comparator/noninferiority RCTs (Stratum A2) were never pooled together. Omnibus six-channel and three-tier epistemic-independence moderators were tested via mixed-effects REML meta-regression. Small-study publication bias was evaluated using Egger's regression asymmetry test and Stanley and Doucouliagos's (2014) Precision-Effect Test and Precision-Effect Estimate with Standard Error (PET–PEESE). Four pre-specified sensitivity analyses tested: (S1) strictly pure self-directed arms versus human-guided/assisted digital arms; (S2) immediate post-intervention assessments versus primary follow-up endpoints; (S3) strict age 2–12 cohorts excluding mixed-age cohorts (Sandborg et al., 2025; Stormshak et al., 2019); and (S4) pure wearable ActiGraph accelerometry excluding clinical anthropometry.

In Subgroup B, for every multi-informant RCT arm reporting both Channel 1 (g_P, V_P, N_P) and a secondary channel (g_S, V_S, N_S), the within-trial paired informant contrast $\Delta g = g_P - g_S$ was computed. Where a trial contributed a multi-scale parent composite in Subgroup A alongside an identical instrument administered to both informants, Subgroup B paired the commensurate same-scale scores (specifically matching parent PSC-17-P g_P = +0.282 against child PSC-17 g_S = +0.401 in Net PAMA: Chanpen et al., 2026, compared with the two-scale Subgroup A parent composite g = +0.223). Because primary trials rarely publish cross-informant covariance matrices of treatment change, exact sampling variance was derived by accounting for partial dyadic overlap between the parent analytical sample N_P and second-channel sample N_S: $\mathrm{Var}(\Delta g) = V_P + V_S - 2 \cdot r_{eff} \cdot \sqrt{(V_P \cdot V_S)}$, where $r_{eff} = r \cdot \min(N_P, N_S) / \sqrt{(N_P \cdot N_S)}$. In strict compliance with multi-informant psychometric benchmarks (De Los Reyes et al., 2015), all Subgroup B syntheses and REML meta-regressions were executed across three pre-specified cross-informant correlations: r = .20 (typical parent–teacher/observer agreement), r = .35 (primary benchmark), and r = .50 (high cross-channel convergence). Three Subgroup B meta-regression models tested whether Δg was moderated by (Model B1) genuine witness independence versus parent-mediated secondary channels, (Model B2) analytical denominator divergence ($N_S / N_P - 1.0$), and (Model B3) joint independence, comparator type, and second-channel attrition ($1 - N_S / N_P$). Outcome-level risk of bias was appraised using Cochrane RoB 2 for individually randomised trials (Higgins et al., 2019) and RoB 2 CRT for cluster-randomised trials (Eldridge et al., 2021), evaluating Domain 4 (measurement of the outcome) and Domain 3 (missing outcome data) separately for each informant channel within each trial. All computations were executed reproducibly in Python 3.12 (run_t02_meta_analysis.py, using NumPy/SciPy).

3. Results

3.1. Characteristics of the 27 quantitative RCT cohorts and linked study families

Table 1 summarises the 27 unique RCT cohorts (28 non-overlapping intervention contrasts) included in the quantitative synthesis, encompassing 10,157 randomised families across compared arms (9,148 in 24 superiority cohorts; 1,009 in 3 active-comparator/noninferiority cohorts; 10,449 total randomised families including third in-person/facilitator arms in four 3-arm trials: Chanpen et al., 2026; DuPaul et al., 2018, 2026; Stormshak et al., 2019). Across these 27 cohorts, 19 were multi-informant trials (20 arm contrasts; 22 parent-versus-second-channel pairs) and 8 were large-scale or clinical single-informant parent-proxy trials retained in Subgroup A Channel 1 to anchor the parent-report benchmark against small-study selection bias (Dunn et al., 2024; DuPaul et al., 2018; Enebrink et al., 2012; Palmer et al., 2023; Sandborg et al., 2025; Sourander et al., 2016; Stormshak et al., 2019; Tran et al., 2026). Geographically, the 27 cohorts spanned Australia (8 cohorts: Baker et al., 2017; Donovan & March, 2014; Hammersley et al., 2019; Havighurst et al., 2024; Morgan et al., 2017; Sandborg et al., 2025; Sanders et al., 2012; Sim et al., 2020), the United Kingdom (5 cohorts: Creswell et al., 2024; Dunn et al., 2024; Palmer et al., 2023; Reardon et al., 2026a, 2026b), Sweden (5 cohorts: Alexandrou et al., 2023; Enebrink et al., 2012; Ghaderi et al., 2018; Nyström et al., 2017; Vigerland et al., 2016), the United States (4

cohorts: DuPaul et al., 2018, 2026; Prinz et al., 2022; Stormshak et al., 2019), Germany (1 cohort, two arms: Döpfner et al., 2025), Finland (1 cohort: Sourander et al., 2016), China (1 cohort: Peng et al., 2025), Thailand (1 cohort: Chanpen et al., 2026), and Vietnam (1 cohort: Tran et al., 2026), published between 2012 and 2026 (with 13 cohorts (14 contrasts) published in 2023–2026). Clinical targets comprised disruptive behaviour, conduct problems, ADHD, or broader behavioural/emotional adjustment (14 cohorts (15 contrasts)), child anxiety or internalising prevention/treatment (8 cohorts), and early-childhood movement, dietary, screen-time, sleep, and obesity prevention (5 cohorts). Seventeen of the 25 superiority arms tested purely self-directed digital programmes without human clinical coaching, whereas 8 incorporated structured therapist calls, written clinical feedback, or blended home practice (Donovan & March, 2014; Döpfner et al., 2025; DuPaul et al., 2018, 2026; Enebrink et al., 2012; Sourander et al., 2016; Tran et al., 2026; Vigerland et al., 2016).

Table 1. Characteristics of the 27 unique randomised controlled trial cohorts (28 non-overlapping intervention contrasts; N = 10,157 in compared arms) included in stratified multi-informant meta-analysis.

Cohort / Study Family & Primary Citation	Country & Design	Child Age & Domain	Assigned Contrast & Guidance	N Rand. (Analysed)	Informant Channels & Measures
iCATS-i2i School Pathway — Reardon et al. (2026b)	UK; Cluster RCT (84 schools)	8–9y (M=8.8); Child anxiety	OSI parent-led online CBT + brief support vs. usual school practice (Pure self-dir. curriculum)	409 (P: 308; C: 352; T: 358)	Ch 1: SCAS-8-P; Ch 2a: Classroom SCAS-8-C & RCADS-C; Ch 3: Teacher SCAS-8-T
MYCATS School Screening — Reardon et al. (2026a) (linked Yu et al., 2026)	UK; Cluster RCT (95 schools)	4–7y; Child anxiety risk	OSI online parent-led CBT vs. usual school practice (Pure self-dir. curriculum)	865 (P: 679; Clin: 622)	Ch 1: PAS; Ch 6: Blinded clinician telephone ADIS-P interview of same parent
Cool Little Kids Online — Morgan et al. (2017)	Australia; Individual RCT	3–6y (M=4.8); Anxiety prevention	8-module unguided online programme vs. waitlist (Pure self-dir.; 5.1% phone call)	433 (P: 345; Clin: 319)	Ch 1: PAS-R + SDQ-ES; Ch 6: Psychologist-scored online parent assessment (OAPA)
BRAVE-ONLINE Preschool — Donovan & March (2014)	Australia; Individual RCT	3–6y (M=4.1); Clinical anxiety	Therapist-assisted online parent CBT vs. waitlist (Guided)	52 (P: 48 (52 ITT); Clin: 50)	Ch 1: PAS + CBCL-Int; Ch 6: Blinded clinician telephone ADIS-P (CSR, CGAS, dx)
BiP Anxiety — Vigerland et al. (2016)	Sweden; Individual RCT	8–12y (M=10.1); Anxiety disorders	Therapist-guided parent (7 mod.) + child (4 mod.) iCBT vs. waitlist (Guided)	93 (P: 76; C: 73; Clin: 90)	Ch 1: SCAS-P; Ch 2a: Independent portal SCAS-C; Ch 6: Clinician ADIS-C/P (CSR, CGAS)
WASH 3-Arm RCT (Arm 1: Unguided)	Germany; 3-Arm	6–12y (M=8.9);	Unguided web-assisted	214 (P: 216*; Clin:*	Ch 1: Online SCL-ADHD; Ch 6:

Cohort / Study Family & Primary Citation	Country & Design	Child Age & Domain	Assigned Contrast & Guidance	N Rand. (Analysed)	Informant Channels & Measures
— Döpfner et al. (2025)	RCT (split TAU)	ADHD / ODD	self-help vs. 50% split shared TAU (Pure self-dir.)	*243*)	Blinded clinician phone DCL-EXT with parent
WASH 3-Arm RCT (Arm 2: Guided) — Döpfner et al. (2025)	Germany; 3-Arm RCT (split TAU)	6–12y (M=8.9); ADHD / ODD	Web self-help + up to 6 phone support calls vs. 50% split shared TAU (Guided)	217 (P: 213*; Clin: 245*)	Ch 1: Online SCL-ADHD; Ch 6: Blinded clinician phone DCL-EXT (10 unblinded replacements)
Tuning in to Kids Online (TIKOL) — Havighurst et al. (2024)	Australia; Individual RCT	4–10y; Challenging behaviour	Self-paced online emotion socialisation vs. waitlist (Pure self-dir.)	151 (P: 151; T: 136)	Ch 1: Child Anxiety + ECBI Prob/Int; Ch 3: Teacher SESBI-R disruptive intensity
Triple P Online (Original) — Sanders, Baker & Turner (2012)	Australia; Individual RCT	2–9y (M=4.7); Conduct problems	8-module Triple P Online vs. internet-use-as-usual (Pure self-dir.)	116 (P: 107; Obs: 41 post / 26 FU)	Ch 1: ECBI Int/Prob + SDQ Cond/Emot; Ch 4: Masked clinic video FOS child disruption
Triple P Online Brief — Baker et al. (2017)	Australia; Individual RCT	2–9y (M=4.4); Conduct problems	5-module brief online programme vs. untreated waitlist (Pure self-dir.)	200 (P: 162 at 9m; Obs: 75 at 9m)	Ch 1: ECBI + CAPES + Concerns; Ch 4: Masked clinic video PCPTOS child negative rate
Project PEAK Full-Scale R01 — DuPaul et al. (2026)	USA; 3-Arm RCT (Online vs. WLC)	3–5y (M=4.2); Preschool ADHD	Asynchronous online BPT + optional check-in calls vs. WLC (Assisted; F2F arm n=66 separate)	130 Online+WLC (P: 98; Obs: 94)	Ch 1: Conners EC-P (IH, GRI, DA); Ch 4: Masked home video MOOSES + direct child HTKS
Net PAMA Multicenter — Chanpen et al. (2026)	Thailand; 3-Arm RCT (Self-dir. vs. WLC)	6–12y (M=8.7); Behavioural problems	Self-directed Net PAMA vs. waitlist (Pure self-dir.; Facilitator arm n=102 separate)	203 Online+WLC (P: 195; C: 195)	Ch 1: PSC-17-P + SNAP-IV ODD; Ch 2b: Child PSC-17 via parent chat link + reading help
Parenting Resilient Kids (PaRK) — Sim et al. (2020) (linked Fernando et al., 2018)	Australia; Individual RCT	8–11y (M=9.8); Anxiety/depression prev.	Tailored interactive web modules vs. educational factsheets + check-in calls (Pure self-dir.)	355 P / 342 C (P: 289; C: 276; 251 dyads)	Ch 1: RCADS-25-P; Ch 2a: Separate child portal RCADS-25-C
Preschool eHealth	China; Individual	3–6y (M=4.6); PA,	12-week WeChat	237 (P: 181; Sensor:	Ch 1: Parent screen

Cohort / Study Family & Primary Citation	Country & Design	Child Age & Domain	Assigned Contrast & Guidance	N Rand. (Analysed)	Informant Channels & Measures
Movement — Peng, Wen & Lau (2025) (linked Zhou et al., 2024, 2026)	RCT	diet, screen, sleep	parent eHealth modules + group goals vs. e-materials (Pure self-dir. content)	181)	time + CSHQ; Ch 5: 7-day waist ActiGraph MVPA, sedentary & valid sleep
Time2bHealthy — Hammersley et al. (2019)	Australia; Individual RCT	2–5y (M=3.5); Obesity / movement	6-module online healthy lifestyle programme vs. usual-care emails (Pure self-dir.)	86 (P: 78; Sensor: 53 MVPA / 41 sleep)	Ch 1: Food, screen, sleep reluctance; Ch 5: Waist ActiGraph + blinded clinic BMI
MINISTOP 1.0 — Nyström et al. (2017) (linked Delisle Nyström et al., 2018)	Sweden; Individual RCT	4.5y; Obesity / diet / PA	6-month MINISTOP mHealth smartphone app vs. pamphlet control (Pure self-dir.)	315 (P: 263; Sensor: 247 ActiGraph / 281 BodPod)	Ch 1: Mobile dietary intake; Ch 5: 7-day wrist ActiGraph wGT3X-BT + BodPod FMI
MINISTOP 2.0 — Alexandrou et al. (2023)	Sweden; Primary-Care RCT	2.5–3y; Diet, screen & BMI	Multilingual MINISTOP 2.0 app in routine child health care vs. standard care (Pure self-dir.)	552 (P: 503; Nurse BMIz: 503)	Ch 1: Sweet drinks, treats, screen, PA; Ch 5: Routine child-health nurse BMI z-score
Sussex Anxious Parents RCT — Dunn et al. (2024)	UK; Individual Online RCT	2–11y (Med=6.0); Child anxiety prev.	Unguided 8-module online parenting course vs. no-treatment control (Pure self-dir.)	1,810 (P: 1,173)	Ch 1 only: SCAS-P / Preschool SCAS + PSC-17 (single parent-proxy informant)
Parent Positive (SPARKLE) — Palmer et al. (2023)	UK; Cohort-Embedded RCT	4–10y (M=7.5); Conduct & emotion	Parent Positive smartphone app vs. follow-up as usual (Pure self-dir.)	646 (P: 466)	Ch 1 only: Parent SDQ Conduct & Emotional subscales
Let's Grow mHealth RCT — Sandborg et al. (2025)	Australia; Individual App RCT	18–35m enrolled (24–41m FU; mixed age)	Let's Grow movement behaviour app vs. usual care (Pure self-dir.; excl. in Sens. S3)	1,165 (P: 1,069)	Ch 1 only: Parent-reported child motor skills & emotion regulation
Strongest Families Smart Website — Sourander et al. (2016)	Finland; Population-Screened RCT	4y; Disruptive behaviour	11-week web modules + weekly 45-min telephone coaching vs. education control (Guided)	464 (P: 382 at 12m)	Ch 1 only: Parent CBCL Ext/Int + ICU (no daycare teacher ratings collected)

Cohort / Study Family & Primary Citation	Country & Design	Child Age & Domain	Assigned Contrast & Guidance	N Rand. (Analysed)	Informant Channels & Measures
FCU Online Foundational Cohort — Stormshak et al. (2019)	USA; 3-Arm School RCT (Web-Only)	11–13y (6th–7th gr.; mixed age)	Pure Web-Only FCU Online vs. school-as-usual (Pure self-dir.; Web+Coach n=108 separate; excl. S3)	214 Web+Ctrl (P: 199)	Ch 1 only: Parent SDQ Conduct/Hyperactivity/Emotion + EATQ Effortful Control
Swedish iPMT Comet — Enebrink et al. (2012) (linked Högström et al., 2013, 2015)	Sweden; Individual RCT	3–12y (M=6.8); Conduct problems	7-session internet PMT + written RA feedback vs. waitlist (Guided; 1 shared cohort across 3 papers)	104 (P: 104 ITT)	Ch 1 only: Parent ECBI Intensity/Problem + SDQ Total
Project PEAK Pilot — DuPaul et al. (2018)	USA; 3-Arm Pilot RCT (Online vs. WLC)	3–5y (M=4.4); ADHD risk	Online BPT + weekly check-in calls vs. waitlist (Guided; F2F arm n=16 separate)	31 Online+WLC (P: 27)	Ch 1 only: Conners EC-P (teacher Conners collected at baseline screening only)
Blended Online BPT Vietnam — Tran et al. (2026)	Vietnam; Individual RCT	6–11y; Clinical ADHD	Blended online modules + RA home practice vs. waitlist (Guided/Blended)	86 (P: 80)	Ch 1 only: Parent VADPRS child ADHD symptoms (parent PSAS skills separate)
Co-CAT (OSI vs. CAMHS TAU) — Creswell et al. (2024) (linked Taylor et al., 2022; Williams et al., 2022)	UK; Pragmatic Noninferiority RCT (34 clinics)	5–12y (Med=9; child ≥7y); Anxiety	Digitally augmented parent-led CBT (OSI + brief therapist) vs. active CAMHS TAU	444 (P: 288; C: 234)	Ch 1: CAIS-P + RCADS-P; Ch 2a: Optional child self-report CAIS-C + RCADS-C
iComet vs. Face-to-Face FCU — Ghaderi et al. (2018)	Sweden; Active-Comparator RCT	10–13y (M=11.4; mixed age); Conduct	Self-directed internet iComet vs. in-person Family Check-Up (Active head-to-head)	231 P / 200 C (Teacher N=203 omitted)	Ch 1: Parent SDQ Conduct & Prosocial; Ch 2a: Child SDQ Conduct & Prosocial
Online vs. Staff Triple P — Prinz et al. (2022) (+ Appendix S1)	USA; Noninferiority RCT	3–7y (M=4.6); Disruptive behaviour	Self-directed Online Triple P vs. individualised in-person Staff-Delivered Triple P	334 (P: 273; Obs: 273; T: 234)	Ch 1: Parent ECBI; Ch 4: Full-sample masked home video PCPTOS; Ch 3: Teacher CADBI

Note. P = Parent-proxy channel; C = Child self-report channel; T = Teacher channel; Obs = Masked behavioural observation channel; Sensor = Wearable accelerometry / objective anthropometry; Clin = Clinician-scored interview of the participating parent. In Döpfner et al. (2025), 431

families were randomised across three arms (Unguided, Guided, TAU; 318 unique families analysed on online SCL-ADHD and 367 on phone DCL-EXT); full two-arm analytical Ns (active + unsplit TAU = 111 online / 121 phone) are displayed, while variance estimation split the shared TAU control denominator 50/50 (n_ctrl_eff = 55.5 and 60.5) per Cochrane Handbook §23.3.4.

3.2. Resolution of dyadic denominators, protocol linkages, and construct/age boundary cases

Full-text verification resolved the dyadic denominator, scoring, and protocol-linkage questions across the three foundational parent–child candidate trials while clarifying eleven critical construct, design, and age boundary cases (Table 2). First, in the pragmatic UK CAMHS noninferiority trial of Online Support and Intervention (OSI; Co-CAT, N = 444, ages 5–12) (Creswell et al., 2024)—linked with its published protocol (Taylor et al., 2022) and Statistical Analysis Plan v4.0 (Williams et al., 2022)—both arms received active unblinded clinical care (digitally augmented parent-led CBT with brief therapist support versus clinician treatment-as-usual). Because child self-report on CAIS-C and RCADS-C was optional and restricted to children aged ≥ 7–8 years, 26-week child denominators (N = 233–235; 123 OSI vs. 111 TAU) were 53–54 families smaller than parent denominators (N = 286–289; 158 vs. 130), and directional point estimates diverged even within the child channel (CAIS-C g = +0.18 vs. RCADS-C g = −0.07; composite g = +0.055 vs. parent g = +0.020, Δg = −0.035). In iCATS-i2i (84 schools, target N = 409, ages 8–9) (Reardon et al., 2026b), classroom administration achieved higher retention for children (N = 352) and teachers (N = 358) than for parents (N = 308): parents reported robust child anxiety reductions at 4 months (g = +0.52) and 12 months (g = +0.42), whereas independent classroom child SCAS-8-C was completely null (g = +0.06, 95% CI [−0.12, +0.24]; Δg = +0.360, p = .0009), target RCADS-C moved slightly backwards (g = −0.06), whole-school 12-month child RCADS-C (N = 3,628) statistically favoured the control arm (p = .0098), and teacher symptom ratings SCAS-8-T missed significance at both 4 months (p = .093) and 12 months (g = +0.14, p = .19).

Second, in Ghaderi et al.'s head-to-head effectiveness trial comparing self-directed internet parent training (iComet) against in-person Family Check-Up (FCU; ages 10–13, M = 11.4) (Ghaderi et al., 2018), engagement was sharply asymmetric (29% completed iComet vs. 78% FCU). Comparing the full text against Table 3 reveals an abstract-versus-table discrepancy: although the abstract states that children's ratings showed no significant change, Table 3 documents a statistically significant condition-by-time interaction on child-rated SDQ Prosocial behaviour ($F(1, 439) = 6.17$, p = .01, reported d = 0.37; where denominator degrees of freedom reflect repeated-measures observations across N = 200 children, not participant count), with children in iComet reporting a marked decline in prosocial behaviour relative to FCU (g = −0.369) that was completely invisible to their parents (SDQ-P Prosocial g = +0.114; within-construct Δg = +0.483, p = .008), while teacher ratings (N = 203) were omitted from all tables as 'virtually non-existent'. Third, acquisition of the full accepted text of PaRK (Sim et al., 2020; protocol

Fernando et al., 2018) resolved the exact CONSORT dyadic intersection behind its marginal assessment frames (355 baseline / 289 12-month parents vs. 342 baseline / 276 12-month children): exactly 251 complete matched parent–child dyads (124 PaRK, 127 control) provided data across all waves via separate parent and child web portals, confirming null symptom effects in both channels (parent RCADS-P g = +0.055; independent child RCADS-C g = +0.015; Δg = +0.040, p = .735) alongside modest change on parent-rated parenting practices (PaRCADS d = 0.24, p = .02).

Full-text inspection of the eleven boundary cohorts (Table 2) reinforced why construct identity, respondent continuity, and counterfactual assignment must be audited before pooling. In SAPP (Åhlén et al., N = 40 across group n = 13, internet n = 14, waitlist n = 13, ages 6–12) (Åhlén et al., 2022), continuous questionnaires (OP, STAI-S, PSWQ, BBQ) measured parental overprotection, state anxiety, worry, and parent quality of life rather than child symptoms, while 12-month child ADIS-C/P diagnostic status combined parent and child interviews into an inseparable evaluator summary (with 4/14 internet vs. 2/13 waitlist anxiety onset confounded by baseline sex imbalance: 71% vs. 31% girls). Similarly, EMPOWER (Sung et al., 2021; N = 301) targeted parental accommodation (FASA) while collecting child RCADS at baseline only; the military ADAPT cohort (Buchanan et al., 2025, Chan et al., 2026, Sloan et al., 2026) collected child BASC-2 only as a baseline moderator of 12-month coded parental skill (FIT) and allocated acceptability surveys at the family cluster level (143 parents nested in 96 families); and MyKidEye (Ghani et al., 2026) measured parental eye-care knowledge, attitude, and practice across just one school per condition. Uncontrolled open trials illustrated why programme adherence and survey return represent non-interchangeable denominators: Pocket PCIT (Jent et al., 2025) shrank from 1,480 enrolled caregivers to 393 homework users and 204 post-assessments (inverting raw signs by scale polarity while printing impossible paired t degrees of freedom), whereas BePresent (Mishina et al., 2025) distinguished 183 module-4 completers from 202 follow-up respondents (125 completers plus 77 noncompleters). In ImPACT Taiwan (Ding et al., 2025; N = 10 mothers), later coaching tiers depended on earlier maternal nonresponse and pooled interval coding across strategies. Finally, three Family Check-Up Online adolescent cohorts—Connell and Stormshak (2023; ages 10–14, M = 12.1), McWhirter et al. (M = 12.61, analysing n = 74/65 self-selected app/coaching exposures within the MSSOS treated arm) (McWhirter et al., 2026), and Connell et al. (2026; N = 281, ages 10–15, M = 12.4)—exceeded the 2–12-year childhood boundary and were excluded from childhood pooling (while foundational middle-school cohort Stormshak et al. (2019; grades 6–7), and toddler cohort Sandborg et al. (2025; enrolled ages 18–35 months), were explicitly flagged and removed in Sensitivity S3).

Table 2. Methodological resolution of foundational parent–child comparison candidates and construct, design, or

age-boundary cohorts excluded from quantitative child-symptom pooling.

Cohort / Programme & Citation	Design & Denominator Frame	Informant & Construct Architecture	Methodological Resolution & Eligibility Classification
Co-CAT / OSI Clinical Trial (Creswell et al., 2024; Taylor et al., 2022; Williams et al., 2022)	Noninferiority RCT (N=444; 26w P=286–289 vs. C=233–235)	Parent & optional child (≥7y) CAIS & RCADS; active clinician TAU in both arms	Resolved in Stratum A2 & Subgroup B (B19): 53–54 fewer children than parents; child CAIS (+0.18) and RCADS (−0.07) opposite in sign; paired $\Delta g = -0.035$.
iComet vs. Family Check-Up (Ghaderi et al., 2018)	Active-comparator RCT (ages 10–13; P=231, C=200, T=203)	Parallel parent & child SDQ; teacher tables omitted as 'virtually non-existent'	Resolved in Stratum A2 & Subgroup B (B20): Table 3 $F(1,439)=6.17$ ($p=.01$) reveals child SDQ Prosocial drop in iComet ($g=-0.369$) missed by abstract & invisible to parents ($g=+0.114$; $\Delta g=+0.483$).
Parenting Resilient Kids (PaRK) (Fernando et al., 2018; Sim et al., 2020)	Minimal-info RCT (355/289 P; 342/276 C; 251 complete dyads)	Parent RCADS-25-P & PaRCADS vs. separate child login portal RCADS-25-C	Resolved via full text in Stratum A1 & Subgroup B (B04): exact CONSORT dyadic overlap verified (124 PaRK + 127 ctrl = 251); both symptom channels null ($\Delta g=+0.040$); Table 2 sign typos corrected.
SAPP Anxiety Prevention (Åhlén et al., 2022)	3-arm pilot RCT (N=40: Group 13, Internet 14, WL 13; ages 6–12)	Continuous OP, STAI-S, PSWQ, BBQ = parent-own worry/overprotection; 12m ADIS-C/P composite	Excluded from quantitative pooling (construct boundary; qualitative only): no separate parent-proxy child questionnaire; evaluator ADIS fuses parent & child accounts (4/14 Internet vs. 2/13 WL onset; 71% vs. 31% girls).
EMPOWER Single-Session (Sung et al., 2021)	2-arm online RCT (N=301 parents of anxious youth aged 4–10)	Repeated outcome = parent accommodation (FASA); child RCADS collected at baseline only	Excluded from quantitative pooling (construct boundary; qualitative only): measures change in caregiver accommodation behaviour rather than a longitudinal child symptom series.
ADAPT Military Family Cohort (Buchanan et al., 2025; Chan et al., 2026; Sloan et al., 2026)	Comparative effectiveness cluster-family RCT (N=149 families; ages 4–12)	Buchanan et al. (2025): 143 parents in 96 families acceptability; Chan et al. (2026): baseline-only BASC-2 moderating 12m parent FIT coding	Linked as 1 shared military cohort & excluded from child pooling (construct boundary): 12m endpoint in Chan et al. (2026) is observed parenting skill, not child symptoms; family clustering requires CRT

Cohort / Programme & Citation	Design & Denominator Frame	Informant & Construct Architecture	Methodological Resolution & Eligibility Classification
			appraisal.
MyKidEye (Ghani et al., 2026)	2-school allocation study (1 school per condition; parents of primary pupils)	Parent eye-health Knowledge, Attitude, and Practice (KAP) questionnaires only	Excluded from quantitative pooling (qualitative audit only): single cluster per condition confounds treatment with school effects (df_cluster = 0), and outcomes measure parental KAP rather than independent child health.
Pocket PCIT Online (Jent et al., 2025)	Uncontrolled open trial (1,480 enrolled → 1,234 pre → 393 homework → 204 post)	Caregiver-only pre–post frequency (d=+0.87) & problem count (d=−0.42)	Excluded from quantitative pooling (qualitative audit only): no untreated counterfactual; 86.2% attrition from enrolment; opposite signs reflect scale polarity; printed paired-t df conflict with N=204.
BePresent Universal (Mishina et al., 2025)	Uncontrolled single-arm pre–post study (752 reg. → 515 started → 183 mod-4; 202 FU)	Caregiver-only pre–post ratings; 202 FU = 125 module completers + 77 noncompleters	Excluded from quantitative pooling (qualitative audit only): proves adherence (reaching module 4) and observation response (returning follow-up survey) are non-nested denominators; abstract–table discrepancies preserved.
ImPACT Taiwan Three-Tier (Ding et al., 2025)	Response-escalated small-N series (10 mothers: 3 Tier-1, +5 Tier-2, +1 Tier-3; 9 FU)	Unblinded interval coding of maternal strategy use (90% pooled rule across strategies)	Excluded from quantitative pooling (qualitative audit only): support escalation is endogenous to prior maternal nonresponse; coding measures maternal fidelity rather than child clinical outcomes.
FCU Online Adolescent Cohorts (Connell et al., 2026; Connell & Stormshak, 2023; McWhirter et al., 2026)	COVID-19 RCT (Connell & Stormshak, 2023) (ages 10–14, M=12.1); MSSOS engagement (McWhirter et al., 2026) (n=74/65, M=12.61); Middle-school RCT (Connell et al., 2026) (N=281, ages 10–15, M=12.4)	Parent-reported adolescent behaviour, effortful control, and engagement exposure associations	Excluded from quantitative pooling by pre-specified age boundary (mean age above 12.0y): McWhirter et al. (2026) additionally analyses non-randomised within-arm coaching/app dose (n=74 frame vs. Table 3 N=65) with registry number discrepancy.

Note. Cohorts in Rows 1–3 entered quantitative Active-Comparator (Stratum A2) or Superiority (Stratum A1) synthesis once full texts established exact denominators; Rows 4–11 were retained strictly in the qualitative methodological synthesis to enforce construct, counterfactual, cluster, and childhood-age eligibility criteria.

3.3. Source-level audit of primary-study statistical and reporting discrepancies

Across the 144 inspected full texts, systematic cross-verification of prose, flow diagrams, means, standard deviations, test statistics, and confidence intervals uncovered 16 concrete primary-study numerical or methodological discrepancies across 14 cohorts (Table 3). Three high-impact RCTs—Prinz et al. (2022, Appendix S1), Baker et al. (2017, Tables 2–3), and Sourander et al. (2016, Table 2, where SEs below 0.05 were rounded to '(0)', producing '3.9 (0)')—tabulated post-treatment estimated marginal SEs in columns adjacent to or visually resembling SDs; naive extraction treating SE as SD would have inflated standardised effect sizes by √n (factors of 6× to 13×). Five trials printed directional sign inversions or missing negative signs on confidence interval bounds: Sim et al. Table 2 printed negative point estimates d = −0.10 [−0.11, 0.31] and d = −0.09 [−0.12, 0.30] whose intervals are centred at +0.10 and +0.09 (Sim et al., 2020); Döpfner et al.'s (2025) abstract printed d = 0.34 [−0.59, −0.09]; Vigerland et al. Table 3 reversed lower and upper bounds on CSR slope ('[−0.77, −1.55]') and printed an asymmetric CI on child QOLI (Vigerland et al., 2016); Hammersley et al. printed '[−1.60 to −3.27]' and '[−0.27 to −0.29]' omitting plus signs on upper bounds (Hammersley et al., 2019); and Reardon et al. (2026a; MYCATS) printed '−1.4 (−2.0 to 0.8)' omitting the minus sign on −0.8.

In addition, Peng et al. (2025, Table 3) printed three ActiGraph sleep rows whose point estimates lay mathematically outside both bounds of their own printed 95% confidence intervals (12-week sleep duration −18.59 vs. [−17.50, 41.68]; 12-week sleep latency −3.19 vs. [11.61, −5.23]; 24-week sleep efficiency −0.25 vs. [−2.29, −1.79]) alongside a prose–table conflict; those three corrupted rows were quarantined from Channel 5 synthesis. Additional audited anomalies included transposed Table 3 and Table 4 numbers/captions in Donovan and March (2014); preprint-to-publisher correction of swapped ECBI Intensity/Problem row labels in Sanders et al. (2012); non-significance of Group × Time interaction (p = .504) presented alongside a 'significant' unadjusted post-test d = 0.40 and inconsistent ITT/age descriptions in Chanpen et al. (2026); raw scale confidence intervals printed adjacent to standardised d in Dunn et al. (2024); one-sided 95% intervals masking a two-sided unfavourable increase in parental worries at follow-up in Palmer et al. (2023); replacement of 10 missing blinded T3 clinician ratings with unblinded clinician ratings in WASH (Döpfner et al., 2025); and inadvertent unblinding of 13.0% of ADIS-P assessors (67 intervention vs. 14 control) in MYCATS (Reardon et al., 2026a).

Table 3. Source-level audit of 16 primary-study statistical, denominator, sign, and blinding discrepancies documented across verified full texts.

#	Cohort & Citation	Location in Source	Printed Discrepancy or Reporting Anomaly	Source Verification & Analytical Handling
1	Online vs. Staff Triple P — Prinz et al. (2022)	Appendix S1 Tables S1–S3	T1 baseline columns print Mean (SD), whereas T2/T3 post/follow-up columns switch to Estimated Mean (SE).	Verified against main-text mixed-model equations; d computed strictly using T1 pooled SD (preventing ~11× √n inflation).
2	Triple P Online Brief — Baker et al. (2017)	Tables 2 & 3 (pp. 84–86)	Dispersion columns print estimated marginal SEs (0.07–0.08 on PCPTOS) rather than SDs alongside Morris d_ppc2.	Back-transformed exact baseline SDs (SD = SE × √n) and verified ANCOVA F(1,86)=1.83 and F(1,71)=0.54.
3	Strongest Families — Sourander et al. (2016)	Table 2 (p. 383)	Prints Mean (SE) where SE less than 0.05 is rounded to '(0)', yielding '3.9 (0)' for 6m Control Parenting Scale vs. 3.1 pre & 2.8 at 12m.	Extracted exact model-derived Cohen's d and 95% CIs from Table 3 rather than corrupted Table 2 rounded SEs.
4	Preschool eHealth — Peng et al. (2025)	Table 3 ActiGraph Sleep (p. 9)	3 sleep point estimates lie mathematically outside both bounds of their printed 95% CIs: 12w duration −18.59 [−17.50, 41.68]; 12w latency −3.19 [11.61, −5.23]; 24w efficiency −0.25 [−2.29, −1.79].	Quarantined all 3 algebraically impossible sleep sub-estimates; pooled only verified MVPA, sedentary, and valid 24w duration rows.
5	PaRK Trial — Sim et al. (2020)	Table 2 (p. 820)	Prints inverted negative signs on RCADS-P Anxiety 'd = −0.10 [−0.11, 0.31]' and RCADS-C Depression 'd = −0.09 [−0.12, 0.30]'.	Midpoints of printed CIs (+0.10 and +0.09) and raw trajectories confirm positive direction (+0.10, +0.09).
6	WASH 3-Arm RCT — Döpfner et al. (2025)	Abstract vs. Table 2 / Text	Abstract prints 'd = 0.34 [−0.59, −0.09]' (opposite sign to negative CI); 10 missing masked T3 DCL-EXT ratings replaced by unmasked clinician ratings.	Harmonised positive therapeutic sign (+0.34); recorded unmasked replacement in Domain 4 RoB 2 appraisal.
7	BiP Anxiety — Vigerland et al. (2016)	Table 3 (p. 53)	Clinician CSR Dslope 95% CI bounds transposed '[−0.77, −1.55]'; child QOLI CI asymmetric '−0.26	Used verified sample sizes (n=31 vs. 42; n=45 vs. 45) and exact Hedges' g formulas.

#	Cohort & Citation	Location in Source	Printed Discrepancy or Reporting Anomaly	Source Verification & Analytical Handling
			[−0.72, 0.27]'.	
8	Time2bHealthy — Hammersley et al. (2019)	Tables 3 & 5	Sedentary 95% CI printed as '−1.60 to −3.27' and BMI CI as '−0.27 to −0.29' (missing '+' signs on upper bounds +3.27 and +0.29).	Confirmed symmetric interval widths around printed point estimates (+0.84 and +0.01) from raw SDs and p-values.
9	MYCATS Cluster RCT — Reardon et al. (2026a)	Table 4 & Masking Report	Parent GAD-7 12m CI printed as '−1.4 (−2.0 to 0.8)' omitting minus sign on −0.8; 81/622 (13.0%) ADIS-P assessors unblinded (67 int vs. 14 ctrl).	Preserved cluster-adjusted child PAS and ADIS-P CIs; documented asymmetric assessor unblinding in RoB 2 CRT.
10	BRAVE Preschool — Donovan & March (2014)	Results pp. 29–31	Table 3 and Table 4 numbers and captions are transposed relative to in-text citations; PAS significant in completers (p=.011) but NS in ITT (p=.066).	Matched tables by variable headers and extracted conservative ITT F(1,50) estimates.
11	Triple P Online — Sanders et al. (2012)	Author Preprint vs. Elsevier PDF	Author manuscript reversed ECBI Intensity and Problem effect-size labels between table and prose; clinic video retained 41/116 at post and 26/116 at 6m.	Verified official Elsevier Table 2 (ECBI Intensity d=0.89, Problem d=0.71) and preserved exact N=41 vs. N=107 denominator gap.
12	iComet vs. FCU — Ghaderi et al. (2018)	Abstract vs. Table 3 (p. 8)	Abstract reports 'no significant change in children's ratings', overlooking Table 3 child SDQ Prosocial deterioration in iComet (F(1, 439)=6.17, p=.01, d=0.37); teacher N=203 tables omitted.	Retained unfavourable child prosocial effect (g=−0.369) and noted repeated-measures df≠N distinction.
13	Net PAMA — Chanpen et al. (2026)	Methods, Flow & Table 3	(a) 305 randomised vs. 290 labelled 'ITT'; (b) Methods age 6–12y vs. Discussion 3–12y; (c) SNAP-IV Group×Time p=.504 NS reported alongside 'significant' raw	Computed baseline-adjusted DiD Hedges' g (+0.16 on SNAP-IV) rather than unadjusted post-test difference.

#	Cohort & Citation	Location in Source	Printed Discrepancy or Reporting Anomaly	Source Verification & Analytical Handling
			post d=0.40.	
14	Sussex Anxious Parents — Dunn et al. (2024)	Results Text & Moderator Table	Prints raw scale CIs immediately beside Cohen's d values for PSC-17/SCARED-A; moderator CI '−0.20 to 0.04' omits minus sign; 57.3% vs. 77.8% differential retention.	Derived primary SCAS-P SE from verified SMD CI (−0.23 to −0.08) and flagged high attrition risk.
15	Parent Positive — Palmer et al. (2023)	Primary & Economic Results	Uses 1-sided 95% CIs; 2-sided tests show unfavourable parental worry increase at T3 (d=+0.14); cost-effectiveness reverses when 4 cost outliers are retained.	Converted all estimates to standard 2-sided 95% CIs for parity across cohorts.
16	Project PEAK — DuPaul et al. (2026)	Table 2 & Text (p. 851)	Pools Face-to-Face (n=66) and Online (n=65) into a combined BPT group in Table 2 after non-significant modality ANCOVAs (p=.09–.12).	Isolated Online-arm degrees of freedom (n_online=48 vs. n_wlc=50; video n=46 vs. 48) so F2F never inflates digital weights.

Note. Every discrepancy was verified directly against official publisher PDFs and supplementary tables prior to meta-analytic coding.

3.4. Subgroup A: Stratified meta-analysis across the six informant channels

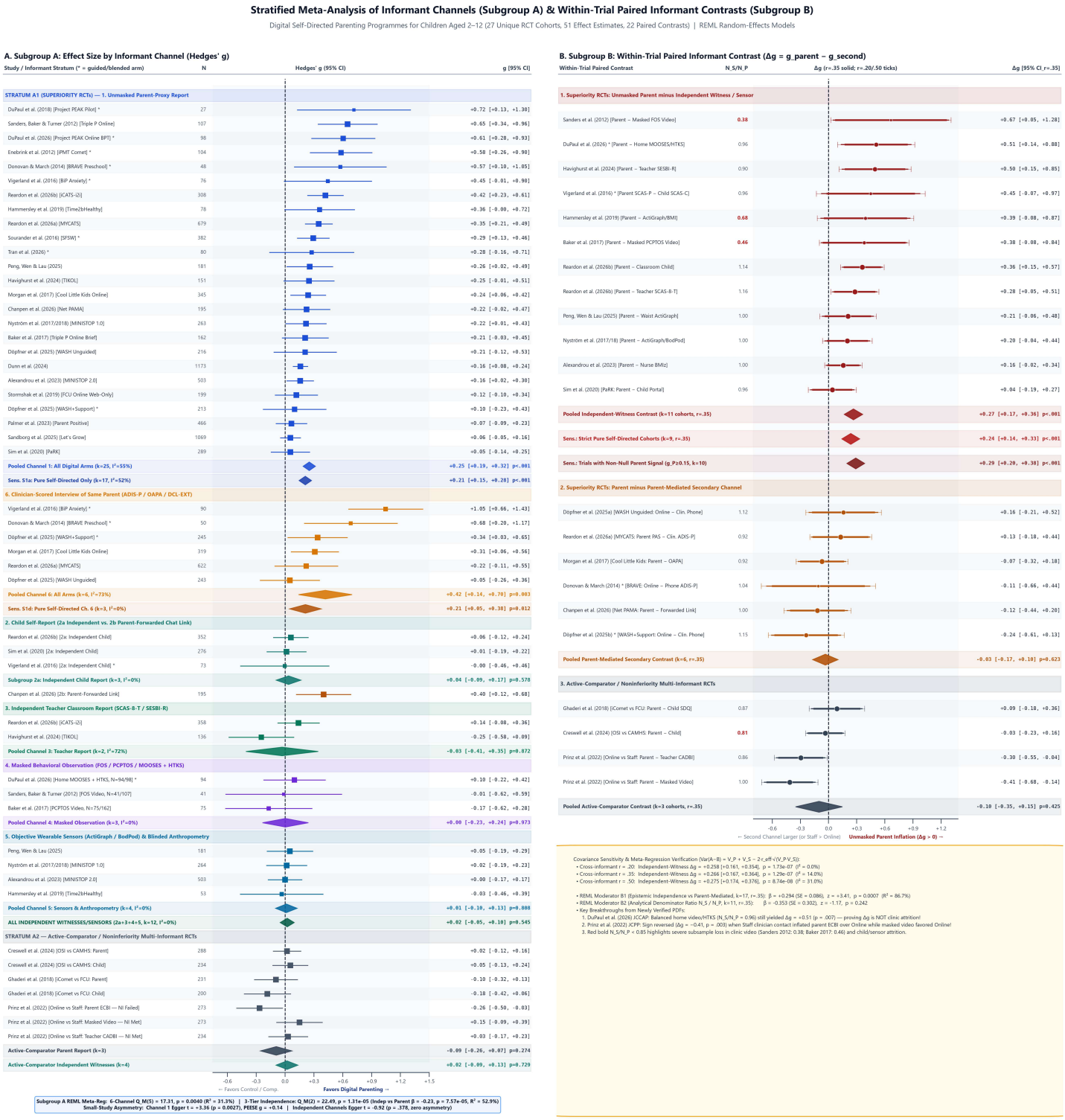
Table 4 and Figure 2 present the Subgroup A stratified random-effects REML meta-analyses across all 27 RCT cohorts (51 effect estimates). Within Superiority Stratum A1 (24 unique cohorts, 25 intervention arms, 44 effect estimates), unmasked parent-proxy questionnaires (Channel 1) demonstrated a highly statistically significant small-to-moderate positive intervention effect (k = 25, N = 7,412 contrast denominators across 7,301 unique families, g = +0.252, SE = 0.033, 95% CI [+0.187, +0.316], Knapp–Hartung 95% CI [+0.182, +0.321], z = +7.66, $p = 1.84 \times 10^{-14}$), accompanied by moderate between-trial heterogeneity (Q(24) = 52.95, p = .00059, $I^2 = 54.7\%$, τ^2_REML = 0.0124). Pre-specified Sensitivity S1 revealed a pronounced clinical-guidance gradient within Channel 1: strictly pure self-directed programmes without therapist or coach contact yielded g = +0.211 (k = 17, N = 6,582, 95% CI [+0.146, +0.275], $p = 1.86 \times 10^{-10}$, $I^2 = 51.7\%$), whereas digital programmes augmented with therapist calls, coaching, or blended feedback yielded nearly double the parent-reported effect, g = +0.406 (k = 8, N = 830, 95% CI [+0.259, +0.554], $p = 6.46 \times 10^{-8}$, $I^2 = 26.0\%$). Substituting immediate post-intervention assessments (Sensitivity S2a: k = 25, g = +0.264 [+0.188, +0.339], $p = 8.72 \times 10^{-12}$) or restricting strictly to cohorts enrolled entirely within ages 2.0–12.0 years (Sensitivity S3, excluding Sandborg et al., 2025 and Stormshak et al., 2019: k = 23, g = +0.271 [+0.206, +0.336], $p = 4.44 \times 10^{-16}$, $I^2 = 48.3\%$) left the parent-channel signal virtually unchanged. However, Egger's regression test detected significant small-study funnel plot asymmetry in Channel 1 (slope = +1.75, t = +3.36, p = .0027), reflecting larger parent effects in smaller clinical trials (e.g., Sanders et al., 2012, g = +0.650; DuPaul et al., 2026, g = +0.605; Enebrink et al., 2012, g = +0.582) than in mega-trials (Dunn et al., 2024, N = 1,173,

g = +0.160; Sandborg et al., 2025, N = 1,069, g = +0.056; Palmer et al., 2023, N = 466, g = +0.066); variance-weighted PEESE adjustment attenuated the asymptotic Channel 1 estimate to g = +0.143 (SE = 0.031, p = .00011; PET intercept g = +0.061, p = .235).

By contrast, every independent non-parent witness channel in Superiority Stratum A1 converged on zero. Overall Child Self-Report (Channel 2: k = 4, N = 896) was statistically non-significant (g = +0.115, 95% CI [−0.059, +0.289], p = .195, I^2 = 44.5%), and pre-specified decomposition exposed the exact source of its heterogeneity: when children completed questionnaires independently in school classrooms or via a separate child web portal (Channel 2a: iCATS-i2i: Reardon et al., 2026b, PaRK: Sim et al., 2020, and BiP Anxiety: Vigerland et al., 2016; k = 3, N = 701), the pooled effect was near-zero with zero heterogeneity (g = +0.037, 95% CI [−0.093, +0.166], p = .578, I^2 = 0.0%). Only when the child's PSC-17 link was sent directly to the unmasked parent's messaging app to forward to the child—with authors explicitly noting parental assistance in reading items (Channel 2b: Net PAMA: Chanpen et al., 2026; k = 1, N = 195)—did child report mirror parent ratings (g = +0.401, 95% CI [+0.118, +0.683], p = .0054). Independent Teacher Report (Channel 3: iCATS-i2i: Reardon et al., 2026b and TIKOL: Havighurst et al., 2024; k = 2, N = 494) yielded g = −0.031 (95% CI [−0.407, +0.345], p = .872, I^2 = 72.3%). Masked Behavioural Observation (Channel 4: Sanders et al., 2012 [FOS]; Baker et al., 2017 [PCPTOS]; and DuPaul et al., 2026 [home MOOSES + direct child HTKS]; k = 3, N = 210) yielded an exact null effect, g = +0.004 (95% CI [−0.235, +0.243], p = .973, I^2 = 0.0%; immediate post Sensitivity S2b: g = −0.058 [−0.336, +0.220], p = .682). Objective Wearable Sensors and Blinded Anthropometry (Channel 5: Peng et al., 2025 [waist ActiGraph]; Hammersley et al., 2019 [ActiGraph/BMI]; Nyström et al., 2017 [MINISTOP 1.0 wrist ActiGraph + BodPod]; and Alexandrou et al., 2023 [MINISTOP 2.0 nurse BMIz]; k = 4, N = 1,001) likewise yielded g = +0.014 (95% CI [−0.099, +0.126], p = .808, I^2 = 0.0%; pure ActiGraph Sensitivity S4: k = 3, g = +0.004 [−0.158, +0.166], p = .965). Pooling all 12 genuinely independent non-parent estimates (Channels 2a + 3 + 4 + 5; N = 2,406 channel-specific assessments across 11 cohorts [2,054 unique children]) produced a tightly bounded null effect of g = +0.023 (SE = 0.037, 95% CI [−0.051, +0.096], Knapp–Hartung 95% CI [−0.060, +0.105], z = +0.60, p = .545) with literally zero heterogeneity ($Q(11) = 4.92$, p = .935, I^2 = 0.0%, τ^2_REML = 0.0000; pure self-directed Sensitivity S1c: k = 10, g = +0.019, p = .625) and no small-study asymmetry (Egger t = −0.92, p = .378).

Clinician-Scored Parent Interviews (Channel 6: Reardon et al., 2026a [MYCATS ADIS-P]; Morgan et al., 2017 [OAPA]; Donovan & March, 2014 [ADIS-P]; Vigerland et al., 2016 [ADIS-C/P]; and Döpfner et al., 2025 [WASH Unguided and Guided DCL-EXT]; k = 6, N = 1,569) yielded g = +0.419 (95% CI [+0.140, +0.697], p = .0032, I^2 = 73.2%). Sensitivity S1d explained that high heterogeneity: in therapist-guided trials where clinicians actively coached or treated families, clinician interview scores were large

Figure 2. Two-Panel Stratified Multi-Informant Forest Plot (Subgroup A & Subgroup B)



Left panel (Subgroup A): REML Hedges' g (95% CI) across 6 informant channels (51 estimates). Right panel (Subgroup B): Paired Δg at r = .35 (22 pairs).

(Vigerland g = +1.046; Donovan g = +0.684; WASH+Support g = +0.339), whereas in strictly pure self-directed trials (MYCATS g = +0.221; Morgan OAPA g = +0.310; WASH Unguided g = +0.050), Channel 6 settled at g = +0.212 (k = 3, N = 1,184, 95% CI [+0.046, +0.377], p = .012, I^2 = 0.0%), matching pure self-directed Channel 1 parent questionnaires (g = +0.211). Subgroup A mixed-effects REML meta-regression confirmed highly significant omnibus channel moderation across all 44 superiority rows (estimating between-channel differences at the contrast-row level, whereas within-trial cross-channel covariance is formally modelled in Subgroup B; 6-channel model: $Q_M(5) = 17.31$, p = .0040, R^2 = 31.3%; 3-tier epistemic-independence model: $Q_M(2) = 22.49$, $p = 1.31 \times 10^{-5}$, R^2 = 52.9%), with genuinely independent witnesses/devices scoring β = −0.233 Hedges' g units lower than unmasked parent-proxy reports (95% CI [−0.349, −0.118], z = −3.96, $p = 7.57 \times 10^{-5}$), whereas parent-mediated secondary channels (Channels 6 + 2b) did not differ negatively from Channel 1 (β = +0.137, p = .084). Finally, in Active-Comparator Stratum A2 (Creswell et al., 2024; Ghaderi et al., 2018; Prinz et al., 2022), parent-proxy ratings slightly favoured in-person clinician comparators over self-directed online delivery (k = 3, g = −0.092 [−0.256, +0.072], p = .274), while independent child, teacher, and masked home-video channels confirmed exact parity (k = 4, g = +0.020 [−0.094, +0.135], p = .729).

Figure 2. Two-panel stratified multi-informant synthesis. Left panel (Subgroup A): Random-effects REML pooled effect sizes (Hedges' *g* and 95% CIs) across informant channels in Superiority RCTs (44 estimates across 24 cohorts / 25 arms) and Active-Comparator / Noninferiority RCTs (7 estimates across 3 cohorts). Right panel (Subgroup B): Within-trial paired informant contrasts ($\Delta g = g_parent - g_second$ with exact covariance-adjusted 95% CIs at benchmark $r = .35$) across 19 multi-informant cohorts (20 unique contrasts; 22 channel pairs; high-resolution 300-DPI forest plot provided in T02_Subgroup_Forest_Plot.png and rendered in full in the companion PDF).

Table 4. Subgroup A stratified random-effects (REML) meta-analysis, sensitivity analyses (S1–S4), small-study bias diagnostics, and mixed-effects meta-regressions across informant channels.

Informant Channel / Stratum / Model	k	N Analysed	Hedges' g [95% CI]	Knapp–Hartung 95% CI	z (p-value)	Heterogeneity (Q, I^2, τ^2)
STRATUM A1: SUPERIORITY RCTs (Digital vs. Waitlist / Usual Care / Minimal Info; 24 cohorts, 25 arms)						
1. Unmasked Parent-Proxy Report (All Superiority)	25	7,412	+0.252 [+0.187, +0.316]	[+0.182, +0.321]	+7.66 (1.84×10^{-14})	Q=52.95, I^2=54.7%, τ^2=0.0124
2. Child Self-Report (All Combined)	4	896	+0.115 [−0.059, +0.289]	[−0.170, +0.401]	+1.30 (.195)	Q=5.41, I^2=44.5%, τ^2=0.0143

Informant Channel / Stratum / Model	k	N Analysed	Hedges' g [95% CI]	Knapp–Hartung 95% CI	z (p-value)	Heterogeneity (Q, I^2, τ^2)
2a. Independent Child Self-Report (Classroom / Separate Portal)	3	701	+0.037 [−0.093, +0.166]	[−0.247, +0.321]	+0.56 (.578)	Q=0.14, I^2=0.0%, τ^2=0.0000
2b. Parent-Forwarded Child Chat Link (Co-Reporting Contamination)	1	195	+0.401 [+0.118, +0.683]	—	+2.78 (.005)	Single arm (k=1)
3. Independent Teacher Report	2	494	−0.031 [−0.407, +0.345]	[−2.470, +2.409]	−0.16 (.872)	Q=3.61, I^2=72.3%, τ^2=0.0540
4. Masked Behavioural Video Observation & Direct Child Task	3	210	+0.004 [−0.235, +0.243]	[−0.520, +0.528]	+0.03 (.973)	Q=0.91, I^2=0.0%, τ^2=0.0000
5. Objective Wearable Sensors & Blinded Anthropometry	4	1,001	+0.014 [−0.099, +0.126]	[−0.169, +0.197]	+0.24 (.808)	Q=0.16, I^2=0.0%, τ^2=0.0000
6. Clinician-Scored Interview of Unmasked Parent	6	1,569	+0.419 [+0.140, +0.697]	[+0.047, +0.791]	+2.95 (.003)	Q=18.68, I^2=73.2%, τ^2=0.0906
Pooled Genuinely Independent Non-Parent Channels (2a + 3 + 4 + 5)	12	2,406	+0.023 [−0.051, +0.096]	[−0.060, +0.105]	+0.60 (.545)	Q=4.92, I^2=0.0%, τ^2=0.0000
Pooled Parent-Mediated Secondary Channels (6 +	7	1,764	+0.412 [+0.183, +0.641]	[+0.117, +0.707]	+3.53 (.0004)	Q=18.73, I^2=68.0%, τ^2=0.0668

Informant Channel / Stratum / Model	k	N Analysed	Hedges' g [95% CI]	Knapp–Hartung 95% CI	z (p-value)	Heterogeneity (Q, I^2, τ^2)
2b)						
PRE-SPECIFIED SENSITIVITY & SMALL-STUDY BIAS ANALYSES (STRATUM A1)						
S1a. Ch 1 Pure Self-Directed Only (Unguided)	17	6,582	+0.211 [+0.146, +0.275]	[+0.140, +0.281]	+6.37 (1.86×10^{-10})	Q=33.10, I^2=51.7%, τ^2=0.0085
S1b. Ch 1 Therapist / Coach / Check-In Assisted Arms	8	830	+0.406 [+0.259, +0.554]	[+0.229, +0.584]	+5.41 (6.46×10^{-8})	Q=9.46, I^2=26.0%, τ^2=0.0133
S1c. Pooled Independent Channels in Pure Self-Directed RCTs	10	2,239	+0.019 [−0.057, +0.095]	[−0.069, +0.107]	+0.49 (.625)	Q=4.68, I^2=0.0%, τ^2=0.0000
S1d. Ch 6 Clinician Parent Interview in Pure Self-Directed RCTs	3	1,184	+0.212 [+0.046, +0.377]	[−0.151, +0.574]	+2.51 (.012)	Q=1.69, I^2=0.0%, τ^2=0.0000
S2a. Ch 1 Immediate Post-Intervention Timepoint	25	7,461	+0.264 [+0.188, +0.339]	[+0.184, +0.343]	+6.83 (8.72×10^{-12})	Q=77.84, I^2=69.2%, τ^2=0.0211
S2b. Ch 4 Masked Observation Immediate Post Timepoint	3	225	−0.058 [−0.336, +0.220]	[−0.667, +0.551]	−0.41 (.682)	Q=2.28, I^2=12.5%, τ^2=0.0152
S3. Ch 1 Strict Childhood Age 2.0–12.0y Only	23	6,144	+0.271 [+0.206, +0.336]	[+0.200, +0.342]	+8.15 (4.44×10^{-16})	Q=42.54, I^2=48.3%, τ^2=0.0102
S4. Ch 5 Pure Wearable ActiGraph Accelerometry Only	3	498	+0.004 [−0.158, +0.166]	[−0.352, +0.359]	+0.04 (.965)	Q=0.35, I^2=0.0%, τ^2=0.0000

Informant Channel / Stratum / Model	k	N Analysed	Hedges' g [95% CI]	Knapp–Hartung 95% CI	z (p-value)	Heterogeneity (Q, I^2, τ^2)
Small-Study Bias: Ch 1 Unmasked Parent-Proxy (Egger / PET–PEESE)	25	7,412	PEESE: +0.143 (SE=0.031)	PET: +0.061 (p=.235)	Egger t=+3.36 (.003)	PEESE p=.0001
Small-Study Bias: Pooled Independent Channels (Egger / PET–PEESE)	12	2,406	PEESE: +0.058 (SE=0.055)	PET: +0.115 (p=.307)	Egger t=−0.92 (.378)	No funnel asymmetry
STRATUM A2: ACTIVE-COMPARATOR / NONINFERIORITY MULTI-INFORMANT RCTs (3 cohorts)						
Active Comparator: Unmasked Parent-Proxy (OSI, iComet, Online Triple P)	3	792	−0.092 [−0.256, +0.072]	[−0.452, +0.269]	−1.09 (.274)	Q=4.12, I^2=51.5%, τ^2=0.0109
Active Comparator: Independent Witnesses (Child, Teacher, Home Video)	4	941	+0.020 [−0.094, +0.135]	[−0.183, +0.223]	+0.35 (.729)	Q=4.07, I^2=26.2%, τ^2=0.0019
SUBGROUP A MIXED-EFFECTS REML META-REGRESSIONS (44 Superiority Estimates)						
Omnibus 6-Channel Moderator Model (Ref: Ch 1 Parent = +0.255)	44	—	$Q_M(5)$ = 17.31	R^2 = 31.3%	p = .004	$Q_E(38)$=42.23 (p=.293)
3-Tier Epistemic Independence Model: Independent Witness vs. Parent Slope	44	—	−0.233 [−0.349, −0.118]	$Q_M(2)$ = 22.49 (R^2=52.9%)	−3.96 (7.57 × 10^{-5})	Parent-Mediated β=+0.137 (p=.084)

Note. Positive Hedges' g indicates child improvement favouring the digital parenting programme (in Stratum A2, favouring self-directed online delivery over in-person clinician/staff comparator). Column N Analysed sums contrast-specific analytical denominators (in Channel 1, 7,301 unique families with the shared WASH usual-care arm contributing 50% variance weight per arm; in Pooled Independent Channels, 2,406

channel-specific assessments across 11 cohorts / 2,054 unique children). REML = Restricted Maximum Likelihood; PET–PEESE = Precision-Effect Test and Precision-Effect Estimate with Standard Error.

3.5. Subgroup B: Within-trial paired informant contrasts (Δg) and moderator meta-regressions

To eliminate between-study confounding across populations, programme curricula, and follow-up durations, Subgroup B evaluated the direct within-trial paired contrast Δg = g_parent − g_second across all 19 multi-informant cohorts (20 unique cohort/arm contrasts; 22 channel pairs; Table 5 and Figure 2 right panel). Across the 11 superiority cohorts pairing an unmasked parent-proxy report with a genuinely independent witness or objective sensor channel, 11 of 11 point contrasts were positive (positive Δg), yielding a highly consistent within-trial informant wedge across every cross-reporter correlation assumption: at r = .20, Δg = +0.258 (SE = 0.049, 95% CI (+0.161, +0.354), $p = 1.73 \times 10^{-7}$, $I^2 = 0.0\%$); at benchmark r = .35, Δg = +0.266 (SE = 0.050, 95% CI (+0.167, +0.364), Knapp–Hartung 95% CI (+0.154, +0.378), z = +5.28, $p = 1.29 \times 10^{-7}$, $I^2 = 14.0\%$); and at r = .50, Δg = +0.275 (SE = 0.051, 95% CI (+0.174, +0.376), $p = 8.74 \times 10^{-8}$, $I^2 = 31.0\%$). Restricting strictly to purely self-directed superiority cohorts (k = 9, r = .35) yielded Δg = +0.237 (95% CI (+0.142, +0.332), $p = 1.08 \times 10^{-6}$, $I^2 = 12.0\%$), and restricting to the 10 superiority cohorts in which parents themselves perceived a non-null effect (g_P ≥ 0.15, excluding universally null PaRK: Sim et al., 2020) yielded Δg = +0.290 (95% CI (+0.195, +0.385), $p = 1.97 \times 10^{-9}$, $I^2 = 0.0\%$). By contrast, across the 6 superiority contrasts pairing unmasked parent questionnaires against parent-mediated secondary channels (clinician interviews of the same parent in MYCATS, Morgan, Donovan, WASH Unguided, WASH+Support, and parent-forwarded child chat link in Net PAMA), the within-trial contrast vanished completely: Δg = −0.034 at r = .35 (95% CI (−0.170, +0.102), p = .623, $I^2 = 0.0\%$; −0.032 at r = .20; −0.037 at r = .50).

Trials with balanced denominators show that the +0.27 SD informant wedge is not a statistical artefact of clinic observation dropout. In DuPaul et al.'s (2026) full-scale Project PEAK preschool ADHD trial, masked coders rated home-recorded parent–child interactions on three MOOSES child behaviour codes alongside direct child Head-Toes-Knees-Shoulders (HTKS) self-regulation testing across virtually the entire retained sample (N_S = 94 vs. N_P = 98; denominator ratio N_S / N_P = 0.959). Although parents in Online BPT demonstrated improved observed positive parenting praise (d = 0.41–0.51) and reported large reductions in child ADHD symptoms on Conners EC-P (g_P = +0.605), all four masked child behavioural and self-regulatory endpoints were null (p = .21–.56, composite g_S = +0.098), yielding a highly significant within-trial divergence of Δg = +0.508 (p = .0074 at r = .35). Exactly the same pattern appeared in school-wide iCATS-i2i (Reardon et al., 2026b), where classroom child retention exceeded parent retention (N_S / N_P = 352 / 308 = 1.143) yet Δg reached +0.360 (p = .00086;

parent-versus-teacher N_S / N_P = 1.162, Δg = +0.280, p = .019), and in primary-care obesity trials Peng et al. (2025; N_S / N_P = 1.000, Δg = +0.206), MINISTOP 1.0 (N_S / N_P = 1.004, Δg = +0.200) (Nyström et al., 2017), and MINISTOP 2.0 (N_S / N_P = 1.000, Δg = +0.157) (Alexandrou et al., 2023).

A sign reversal emerged in Prinz et al.'s (2022) active-comparator noninferiority trial comparing self-directed Online Triple P against individualised in-person Staff-Delivered Triple P (N = 334; T2 retained N = 273). Because families assigned to Staff-Delivered Triple P received face-to-face clinician sessions, their expectancy and alliance were higher than in unguided online self-help: on parent-reported ECBI at post-treatment, Online Triple P failed noninferiority margins in both ITT and per-protocol analyses (g_P = −0.264 favouring in-person staff). Yet on full-sample masked home video coding of child disruptive behaviour (PCPTOS, N = 273; g_S = +0.150 favouring Online, meeting noninferiority at p = .001) and preschool/school teacher CADBI ratings (N = 234; g_S = +0.032, meeting noninferiority across all four scales while within-group teacher change in both arms was near zero, d = 0.06–0.23), the parent advantage of in-person staff vanished, producing significant negative within-trial contrasts of Δg = −0.414 (p = .0028) and Δg = −0.296 (p = .0241). Formal Subgroup B REML meta-regressions (Table 5) confirmed these mechanisms quantitatively: in Model B1 (k = 17 superiority contrasts, r = .35), genuine witness independence strongly predicted Δg (β = +0.294, 95% CI (+0.125, +0.462), z = +3.41, p = .00066), explaining R^2 = 86.7% of between-trial variance (R^2 = 100.0% at r = .20). In Model B2 (k = 11 independent-witness superiority contrasts), the denominator-divergence slope (N_S / N_P − 1.0) was negative as expected when clinic subsamples suffer attrition (β = −0.353, p = .242), while the intercept at perfectly balanced analytical sample sizes (N_S / N_P = 1.0) remained highly significant at β_0 = +0.253 (95% CI (+0.146, +0.360), z = +4.65, p = 3.25×10^{-6}). Across all 20 unique multi-informant contrasts jointly (Model B3, r = .35), witness independence (β = +0.280, p = .0019), superiority-versus-active comparator design (β = +0.409, p = 1.40×10^{-5}), and second-channel attrition (β = +0.623, p = .052) jointly accounted for R^2 = 88.0% of total variance in informant divergence (Q_M(3) = 24.74, p = 1.75×10^{-5}; R^2 = 98.4% at r = .20).

Table 5. Subgroup B within-trial paired informant contrasts (Δg = g_parent − g_second) across cross-informant correlation sensitivity grids (r = .20, .35, .50) and REML moderator models.

Within-Trial Contrast Stratum / Key Trial / Moderator Model	k	Δg (95% CI) at r = .20	Δg (95% CI) at r = .35 (Primary)	Δg (95% CI) at r = .50	p-value (r = .35)	I² (r = .35) / Model Fit
POOLED WITHIN-TRIAL CONTRAST STRATA (19 Multi-Informant Cohorts; 20 Unique Arm Contrasts)						
Superiority: Parent minus Independent Witness (Unique Cohorts)	11	+0.258 (+0.161, +0.354)	+0.266 (+0.167, +0.364)	+0.275 (+0.174, +0.376)	1.29×10^{-7}	I² = 14.0% (Q=11.63)
Pure Self-Directed Only: Parent minus Independent Witness	9	+0.236 (+0.138, +0.334)	+0.237 (+0.142, +0.332)	+0.242 (+0.145, +0.339)	1.08×10^{-6}	I² = 12.0% (Q=9.09)
Trials with Non-Null Parent Signal (g_P ≥ 0.15)	10	+0.289 (+0.188, +0.391)	+0.290 (+0.195, +0.385)	+0.299 (+0.204, +0.393)	1.97×10^{-9}	I² = 0.0% (Q=7.83)
Superiority: Parent minus Parent-Mediated Secondary (Ch 6 & 2b)	6	-0.032 (-0.181, +0.117)	-0.034 (-0.170, +0.102)	-0.037 (-0.159, +0.085)	.623	I² = 0.0% (Q=3.75)
All Multi-Informant Superiority Contrasts Combined	17	+0.177 (+0.081, +0.273)	+0.179 (+0.079, +0.278)	+0.181 (+0.078, +0.285)	.0004	I² = 42.2% (Q=27.69)
Active-Comparator / Noninferiority Multi-Informant RCTs	3	-0.102 (-0.350, +0.146)	-0.102 (-0.352, +0.148)	-0.102 (-0.354, +0.151)	.425	I² = 69.5% (Q=6.56)
ILLUSTRATIVE BALANCED-DENOMINATOR & SIGN-REVERSAL TRIALS (Exact Covariance Propagation)						
iCATS-i2i (Reardon et al., 2026b): Parent SCAS-8-P (N=308) – Classroom Child	1	+0.360 (+0.127, +0.593)	+0.360 (+0.148, +0.572)	+0.360 (+0.172, +0.548)	.0009	N_S / N_P = 1.143 (Balanced+)

Within-Trial Contrast Stratum / Key Trial / Moderator Model	k	Δg (95% CI) at r = .20	Δg (95% CI) at r = .35 (Primary)	Δg (95% CI) at r = .50	p-value (r = .35)	I² (r = .35) / Model Fit
(N=352)						
Project PEAK (DuPaul et al., 2026): Parent Conners (N=98) – Home Video/HTKS (N=94)	1	+0.508 (+0.097, +0.918)	+0.508 (+0.136, +0.879)	+0.508 (+0.180, +0.835)	.0074	N_S / N_P = 0.959 (Balanced)
TIKOL (Havighurst et al., 2024): Parent Composite (N=151) – Teacher SESBI-R (N=136)	1	+0.495 (+0.111, +0.880)	+0.495 (+0.145, +0.846)	+0.495 (+0.183, +0.808)	.0056	N_S / N_P = 0.901 (Balanced)
Triple P Online (Sanders et al., 2012): Parent (N=107) – Clinic FOS Video (N=41)	1	+0.665 (+0.024, +1.306)	+0.665 (+0.052, +1.278)	+0.665 (+0.080, +1.250)	.0336	N_S / N_P = 0.383 (Depleted)
Prinz et al. (2022) Noninferiority: Parent ECBI (N=273) – Home Video (N=273)	1	−0.414 (−0.715, −0.113)	−0.414 (−0.685, −0.143)	−0.414 (−0.652, −0.176)	.0028	N_S / N_P = 1.000 (Sign reversal)
Prinz et al. (2022) Noninferiority: Parent ECBI (N=273) – Teacher CADBI (N=234)	1	−0.296 (−0.579, −0.014)	−0.296 (−0.554, −0.039)	−0.296 (−0.527, −0.066)	.0241	N_S / N_P = 0.857 (Sign reversal)
SUBGROUP B REML META-REGRESSION MODERATOR MODELS						
Model B1 (k=17): Genuine Witness	17	+0.287 (+0.111, +0.463)	+0.294 (+0.125, +0.462)	+0.306 (+0.138, +0.474)	.0007	R^2 = 86.7% (100% at r=.20)

Within-Trial Contrast Stratum / Key Trial / Moderator Model	k	Δg (95% CI) at r = .20	Δg (95% CI) at r = .35 (Primary)	Δg (95% CI) at r = .50	p-value (r = .35)	I^2 (r = .35) / Model Fit
Independence Slope (vs. Parent-Mediated)						
Model B2 (k=11): Intercept β_0 at Balanced Denominators (N_S / N_P = 1.0)	11	+0.250 (+0.143, +0.356)	+0.253 (+0.146, +0.360)	+0.258 (+0.150, +0.365)	3.25×10^{-6}	Slope β_1=-0.353 (p=.242)
Model B3 Joint (k=20): Independence + Superiority Design + Second-Channel Attrition	20	Q_M(3) = 24.18 (R^2=98.4%)	Q_M(3) = 24.74 (R^2=88.0%)	Q_M(3) = 25.37 (R^2=79.6%)	1.75×10^{-5}	Q_E(16)=14.60 (p=.554)

Note. $\Delta g = g_{parent} - g_{second}$, with exact sampling variance $Var(\Delta g) = V_P + V_S - 2 \cdot r_{eff} \cdot \sqrt{(V_P \cdot V_S)}$ and overlap-adjusted correlation $r_{eff} = r \cdot \min(N_P, N_S) / \sqrt{(N_P \cdot N_S)}$. Positive Δg indicates larger treatment effects on unmasked parent questionnaires than on the second channel; negative Δg in Prinz et al. (2022) reflects in-person clinician contact inflating parent ECBI ratings relative to online self-help while masked video and teachers showed parity/slight online advantage. Full 22-pair dataset is provided in Supplement Table S3.

3.6. Outcome-level Risk of Bias (RoB 2 and RoB 2 CRT) across informant channels

Outcome-specific RoB 2 and RoB 2 CRT appraisals (detailed domain-by-domain in Supplement S1, Table S4) demonstrated why methodological quality is a property of a specific outcome channel and denominator rather than a whole trial (Eldridge et al., 2021; Higgins et al., 2019). Across all 25 superiority arms, every Channel 1 unmasked parent-proxy estimate triggered 'High Risk' in Domain 4 (measurement of the outcome), because unblinded caregivers served as their own outcome assessors against an untreated waitlist, usual-care, or minimal-information control. Moreover, 9 of 25 parent-proxy arms additionally incurred 'Some Concerns' or 'High Risk' in Domain 3 due to substantial or differential post-randomisation attrition (e.g., 57.3% intervention vs. 77.8% control retention in Dunn et al., 2024; 74% vs. 85% in Palmer et al., 2023). Switching to an independent witness channel reversed the risk profile between Domain 4 and Domain 3 depending on trial logistics: clinic-based masked video subsamples in Sanders et al. (2012; 41/116 analysed at post, 26/116 at 6 months) and Baker et al. (2017; 75/200 at 9 months), as well as waist ActiGraph wear in Hammersley et al. (2019; 53/86 MVPA, 41/86

sleep), achieved 'Low Risk' in Domain 4 (measurement) yet suffered 'High Risk' in Domain 3 (missing outcome data). Conversely, five contemporary multi-informant designs—classroom child/teacher follow-up in iCATS-i2i (86–88% target retention) (Reardon et al., 2026b), masked home video + direct child HTKS in DuPaul et al. (2026), full-sample home video in Prinz et al. (2022), wrist ActiGraph + BodPod in MINISTOP 1.0 (88–89% retention) (Nyström et al., 2017), and routine nurse anthropometry in MINISTOP 2.0 (91% retention) (Alexandrou et al., 2023)—achieved 'Low Risk' in Domain 4 alongside 'Low Risk' or minimal 'Some Concerns' in Domain 3, and every single one of those low-measurement-bias, high-retention channels produced a null superiority effect (g = 0.00 to +0.14). Finally, Channel 6 clinician-scored parent interviews could not achieve 'Low Risk' in Domain 4 despite assessor masking, both because the underlying testimony originated from an unblinded parent and because operational unblinding occurred frequently during live interviews (13.0% unblinded in MYCATS, asymmetrically 67 intervention vs. 14 control: Reardon et al., 2026a; 10 missing masked T3 ratings replaced by unmasked clinician ratings in WASH: Döpfner et al., 2025).

4. Discussion

4.1. Caregiver appraisal versus cross-observer child symptom change

Across 144 full-text primary reports and statistical supplements representing 27 unique RCT cohorts, this synthesis addresses a direct question in digital child mental health and behavioural prevention: when parents complete a self-directed digital parenting programme, who observes child change? Unmasked caregivers assigned to digital parenting interventions report consistent small-to-moderate improvements in their children's disruptive behaviour, anxiety, and lifestyle routines (g = +0.252 overall; g = +0.211 in strictly unguided programmes). Across 12 independent witness and objective device estimates spanning 2,406 channel-specific assessments across 11 cohorts (2,054 unique children) (independent child classroom or portal self-report, g = +0.037; classroom teachers, g = −0.031; masked behavioural coders scoring parent–child interactions and direct self-regulation tasks, g = +0.004; and multi-day wearable accelerometers or clinical anthropometry, g = +0.014), the pooled intervention effect is zero (g = +0.023, 95% CI (−0.051, +0.096), I^2 = 0.0%). Within-trial covariance-adjusted contrasts confirm that within the exact same randomised cohorts, parent-proxy effects exceed independent witness effects by $\Delta g = +0.266$ ($p = 1.29 \times 10^{-7}$).

The pattern across trials is consistent. Under the Operations Triad Model of multi-informant assessment (De Los Reyes et al., 2015), diverging informant reports can reflect either context-specific behavioural expression (e.g., home versus school) or informant-specific measurement variance. However, contextual

specificity between home and school cannot explain why masked coders scoring recorded interactions inside the family's own home (DuPaul et al., 2026), 24-hour waist and wrist accelerometers worn continuously at home (Peng et al., 2025; Nyström et al., 2017; Hammersley et al., 2019), and children reporting on their own everyday anxiety via independent portals (Sim et al., 2020; Vigerland et al., 2016; Reardon et al., 2026b) all yield null between-group effects alongside positive parent questionnaires. This divergence between caregiver reports and independent observations requires careful clinical and theoretical interpretation. DuPaul et al. (2026) showed that parents in Online BPT increased their observed positive parenting praise on home video (d = 0.41–0.51) and perceived large child ADHD reductions (g = +0.605), whereas neither masked home video of child noncompliance and negative behaviour nor direct child self-regulation testing shifted (g = +0.098). Digital self-directed modules appear sufficient to improve parental knowledge (Sandborg et al., 2025), parenting confidence (Alexandrou et al., 2023; Sim et al., 2020; Tran et al., 2026), and discrete parental praise habits (DuPaul et al., 2026), reframing how caregivers interpret and tolerate child behaviour on subjective rating scales before—or without—producing measurable cross-observer reductions in child symptomatology.

4.2. Dose-dependent expectancy, effort justification, and active-comparator reversal

Three converging lines of evidence indicate that unmasked caregiver ratings in digital parenting trials incorporate substantial treatment-expectancy and cognitive effort-justification components. First, in Superiority Stratum A1, adding human therapist or coach contact nearly doubled unmasked parent-proxy effects from g = +0.211 (pure self-directed) to g = +0.406 (guided/assisted), while independent child self-report in therapist-guided BiP Anxiety remained −0.004 (Vigerland et al., 2016) and masked video/HTKS in check-in-supported Project PEAK remained +0.098 (DuPaul et al., 2026). Second, small-study asymmetry was pronounced in Channel 1 (Egger p = .0027, shrinking to PEESE g = +0.143 in high-precision mega-trials) but completely absent in independent channels (Egger p = .378).

Third, Active-Comparator Stratum A2 tests differential expectancy directly. When both arms receive active clinician-supported care (OSI vs. CAMHS TAU: Creswell et al., 2024), parent and child contrasts converge near zero (Δg = −0.035). And when pure self-directed Online Triple P is pitted against individualised in-person Staff-Delivered Triple P (Prinz et al., 2022), differential therapeutic alliance and effort favour the in-person clinician arm: parents in the staff-delivered arm report significantly larger ECBI symptom drops than online parents (causing Online Triple P to fail post-test noninferiority, g_P = −0.264), whereas masked home video coders (g_S = +0.150) and school teachers (g_S = +0.032) observe slight numerical advantages for Online delivery, flipping the informant wedge to Δg = −0.414 (p = .0028). Thus, Δg tracks the direction of differential caregiver expectancy and clinical contact across both

superiority (+0.266) and head-to-head noninferiority (−0.414) designs.

4.3. Scorer blinding is not informant independence, and digital forwarding contaminates child reports

The six-channel taxonomy exposes two methodological fallacies pervasive in contemporary digital trial design. First, employing a blinded psychologist or psychiatrist to score a structured diagnostic interview conducted exclusively with the unmasked participating parent (Channel 6) does not create a second independent witness (Donovan & March, 2014; Döpfner et al., 2025; Morgan et al., 2017; Reardon et al., 2026a). In strictly self-directed trials, blinded clinician parent-interview estimates (g = +0.212) matched unmasked parent questionnaires (g = +0.211) to the second decimal place, and within-trial contrasts across all Channel 6 studies averaged Δg = −0.034 (p = .623). Furthermore, live telephone interviews with treated parents routinely breach assessor masking—as documented transparently in MYCATS, where 67 intervention parents versus 14 control parents revealed their allocation to blinded assessors (Reardon et al., 2026a).

Second, digital survey distribution introduces a novel threat of cross-informant contamination when 'child self-reports' are routed through the unmasked caregiver's smartphone. Across the three superiority trials in which children completed questionnaires in school classrooms (Reardon et al., 2026b) or through an independent child login portal (Sim et al., 2020; Vigerland et al., 2016), pooled child self-report was g = +0.037 (I^2 = 0.0%). By contrast, in Net PAMA (Chanpen et al., 2026)—where the child PSC-17 link was dispatched to the unmasked parent's chat application to forward to the child, and parents assisted children in reading items—the nominal child effect surged to g = +0.401 (exceeding both the matched parent PSC-17-P DiD of g = +0.282 [paired Δg = −0.119] and the two-scale parent DiD composite of g = +0.223). Unless child assessments are ring-fenced from parental device mediation and co-reading, digital child surveys risk functioning as a second parent-proxy channel.

4.4. Five reporting guardrails for future digital parenting RCTs

To make future multi-informant evaluations of digital parenting programmes interpretable and reusable, trialists, protocol reviewers, and journal editors should enforce five concrete reporting standards: (1) Epistemic Channel Labelling: explicitly distinguish person measured, first-hand informant, scorer, setting, and respondent continuity, never labelling clinician-scored parent interviews as 'clinician-observed child outcomes'; (2) Ring-Fenced Digital Administration: deliver child and teacher questionnaires through independent school sessions or authenticated portals with verification of whether adults read or discussed items; (3) Balanced Multi-Informant Denominators: replace high-attrition clinic observation subsamples

with full-sample remote home video protocols (DuPaul et al., 2026; Prinz et al., 2022) or passive wearable telemetry (Nyström et al., 2017), reporting exact arm-by-wave eligible, observed, valid, and matched-dyad rosters; (4) Direct Within-Trial Contrast & Covariance Reporting: publish the cross-informant correlation matrix of baseline and change scores (or directly fit multivariate mixed models testing Informant × Treatment interactions) rather than comparing separate p-values; and (5) Numerical Table Auditing: explicitly label SD versus SE in every column header and verify algebraically that all printed point estimates lie symmetrically within their confidence intervals prior to publication.

5. Strengths and Limitations

Key strengths of this synthesis include full-text source verification across 144 primary documents and statistical supplements (adjudicating all candidate multi-informant trials at full text), strict cohort-family linkage preventing duplicate counting across 56 linked reports, explicit quarantine of 16 primary-study numerical errors, separation of superiority and active-comparator estimands, and exact covariance-adjusted within-trial contrasts across sensitivity correlations (r = .20, .35, .50). Several limitations nevertheless warrant note. First, although programmatic harvesting across four bibliographic search APIs paired with Unpaywall open-access resolution and cross-checks against recent Q1 reviews captured the international English-language RCT literature through 8 November 2026, unpublished grey-literature trials or non-English regional reports not indexed with English metadata could remain outside the present corpus; notably, small-study bias analyses (PEESE g = +0.143) suggest that unpublished null parent-report studies would further narrow—or reinforce the expectancy origin of—Channel 1 estimates rather than create missing objective effects. Second, within-informant multi-scale composites used a standard assumed correlation (ρ = .50) when primary correlation matrices were unavailable, and within-trial contrasts relied on sensitivity grids (r = .20, .35, .50) because primary trials did not publish exact cross-reporter covariances of treatment change; however, the stability of Δg (+0.258 to +0.275, all p less than 2×10^{-7}) demonstrates complete insensitivity to this parameter. Third, two superiority channels had modest cohort counts (independent teacher report k = 2; masked behavioural observation k = 3, plus Prinz et al. in Stratum A2), reflecting the field's heavy reliance on caregiver surveys; pooling all 12 genuinely independent witness estimates (N = 2,406 assessments across 2,054 unique children, $I^2 = 0.0\%$) confirmed high statistical precision (95% CI upper bound +0.096). Finally, screening, structured data extraction, and risk-of-bias mapping were conducted by a single investigator alongside programmatic Python verification against primary full-text tables; computational cross-checks ensure arithmetic precision but do not substitute for two independent reviewers.

6. Conclusion

Across 27 randomised controlled trial cohorts (N = 10,157 in compared arms), self-directed digital parenting programmes for children aged 2–12 produce reliable small-to-moderate improvements when child outcomes are rated by the unmasked participating parent (g = +0.252; pure self-directed g = +0.211) or by a clinician interviewing that same parent (pure self-directed g = +0.212), yet produce no detectable benefit across independent child self-reports (g = +0.037), teacher ratings (g = −0.031), masked behavioural observations (g = +0.004), or wearable sensors and blinded anthropometry (g = +0.014; pooled independent channels g = +0.023, $I^2 = 0.0\%$). Within-trial covariance-adjusted contrasts confirm a robust informant wedge ($\Delta g = +0.266$, $p = 1.29 \times 10^{-7}$) that persists in balanced-denominator trials and reverses direction when in-person clinician delivery outperforms unguided online delivery on parent ratings alone ($\Delta g = -0.414$). Transforming digital parenting research from caregiver perception studies into rigorous child-outcome science requires ring-fenced multi-informant measurement, balanced non-parent denominators, and routine reporting of within-trial cross-informant covariances.

Declarations and Transparency Statements

- Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
- Competing Interests: The author declares no competing financial or non-financial interests.
- Ethical Approval: Not required; this study is a systematic review and meta-analysis of published aggregate trial data.
- **Availability of Data, Code, and Materials**: Complete extracted cohort-by-channel effect sizes (51 rows; T02_EFFECT_SIZES_EXTRACTED.csv), within-trial covariance-adjusted contrasts (22 pairs across r = .20, .35, .50; T02_WITHIN_TRIAL_CONTRASTS.csv), meta-analytic summary outputs (T02_META_RESULTS.json), cohort-family linkage rosters (56 reports), outcome-level RoB 2 / RoB 2 CRT judgements, and the standalone reproducible Python 3.12 script (run_t02_meta_analysis.py) are provided in Online Supplement S1 and the companion repository package. Licensed third-party publisher PDFs are not redistributed.

- Copyright Notice: © 2026 Alexander Kuznetsov. Reading, downloading for personal scholarly review, and citation are permitted. Redistribution, commercial reuse, and distribution of adapted versions require the copyright holder's permission, except where applicable journal publishing licences or law permit otherwise.

References

Åhlén, J., Vigerland, S., Lindberg, M., Gunterberg, O., & Ghaderi, A. (2022). Developing a brief parent training intervention to prevent anxiety in offspring. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, *10*(1), 123–133. https://doi.org/10.2478/sjcapp-2022-0013

Alexandrou, C., Henriksson, H., Henström, M., Henriksson, P., Delisle Nyström, C., Bendtsen, M., & Löf, M. (2023). Effectiveness of a smartphone app (MINISTOP 2.0) integrated in primary child health care to promote healthy diet and physical activity behaviors and prevent obesity in preschool-aged children: Randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, *20*(1), 21. https://doi.org/10.1186/s12966-023-01405-5

Baker, S., Sanders, M. R., Turner, K. M. T., & Morawska, A. (2017). A randomized controlled trial evaluating a low-intensity interactive online parenting intervention, Triple P Online Brief, with parents of children with early onset conduct problems. *Behaviour Research and Therapy*, *91*, 78–90. https://doi.org/10.1016/j.brat.2017.01.016

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Buchanan, G. J. R., Simenec, T. S., Cai, Q., & Gewirtz, A. H. (2025). Feasibility and acceptability of formats in a comparative effectiveness trial of a preventive parenting program. *Journal of Child and Family Studies*, *34*(2), 381–395. https://doi.org/10.1007/s10826-025-03016-z

Chan, A. C. Y., Lee, S. K., Hanson, A. N., Gewirtz, A. H., & Piehler, T. F. (2026). Online versus telehealth delivery of a parenting program for postdeployed military families: Child internalizing and externalizing symptoms as moderators of response. *Journal of Consulting and Clinical Psychology*, *94*(3), 215–228. https://doi.org/10.1037/ccp0001018

Chanpen, S., Chaiudomsom, C., Jeerasup, N., Inta, P., Thongseiratch, T., & Thanaariyapaisan, P. (2026). An internet-based parent management training program (Net PAMA) improves parenting outcomes in a multicenter randomized controlled trial. *Scientific Reports*, *16*, 8412. https://doi.org/10.1038/s41598-026-40857-0

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22%3C3127::AID-SIM784%3E3.0.CO;2-M

Connell, A. M., Quinn, B., Chu, P. V., Caruthers, A. S., Mauricio, A. M., & Stormshak, E. A. (2026). A family-centered digital intervention for parents of at-risk middle-school students. *Prevention Science*, *27*(3), 312–325. https://doi.org/10.1007/s11121-026-01946-8

Connell, A. M., & Stormshak, E. A. (2023). Evaluating the efficacy of the Family Check-Up Online to improve parent mental health and family functioning in response to the COVID-19 pandemic: A randomized clinical trial. *Journal of Prevention*, *44*(3), 341–357. https://doi.org/10.1007/s10935-023-00727-1

Creswell, C., Taylor, L., Giles, S., Howitt, S., Radley, L., Whitaker, E., Brooks, E., Knight, F., Raymont, V., Hill, C.,

van Santen, J., Williams, N., Mort, S., Harris, V., Yu, S., Pollard, J., Violato, M., Waite, P., & Yu, L.-M. (2024). Digitally augmented, parent-led CBT versus treatment as usual for child anxiety problems in child mental health services in England and Northern Ireland: A pragmatic, non-inferiority, clinical effectiveness and cost-effectiveness randomised controlled trial. *The Lancet Psychiatry*, *11*(3), 193–209. https://doi.org/10.1016/s2215-0366(23)00429-700429-7)

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

Delisle Nyström, C., Sandin, S., Henriksson, P., Henriksson, H., Maddison, R., & Löf, M. (2018). A 12-month follow-up of a mobile-based (mHealth) obesity prevention intervention in pre-school children: The MINISTOP randomized controlled trial. *BMC Public Health*, *18*(1), 658. https://doi.org/10.1186/s12889-018-5569-4

Ding, Y.-S., Wang, H.-T., & Lin, F.-Y. (2025). Incorporating a three-tier parent education model into ImPACT curriculum in Taiwan. *Journal of Child and Family Studies*, *34*(5), 1163–1177. https://doi.org/10.1007/s10826-025-03051-w

Donovan, C. L., & March, S. (2014). Online CBT for preschool anxiety disorders: A randomised control trial. *Behaviour Research and Therapy*, *58*, 24–35. https://doi.org/10.1016/j.brat.2014.05.001

Döpfner, M., Plück, J., Rosenberger, K. D., Klemp, M.-T., Mühlenmeister, J., Wähnke, L., Hellmich, M., Schürmann, S., & Dose, C. (2025). Efficacy of guided and unguided web-assisted self-help for parents of children with attention-deficit/hyperactivity disorder and oppositional defiant disorder: A three-arm randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *66*(9), 1320–1332. https://doi.org/10.1111/jcpp.14153

Dunn, A., Alvarez, J., Arbon, A., Bremner, S. A., Elsby-Pearson, C., Emsley, R., Jones, C., Lawrence, P., Lester, K. J., Morson, N., Simner, J., Thomson, A., & Cartwright-Hatton, S. (2024). Effectiveness of an unguided modular online intervention for highly anxious parents in preventing anxiety in their children: A parallel group randomised controlled trial. *The Lancet Regional Health – Europe*, *45*, 101038. https://doi.org/10.1016/j.lanepe.2024.101038

DuPaul, G. J., Kern, L., Belk, G., Custer, B., Daffner, M., Hatfield, A., & Peek, D. (2018). Face-to-face versus online behavioral parent training for young children at risk for ADHD: Treatment engagement and outcomes. *Journal of Clinical Child & Adolescent Psychology*, *47*(Suppl. 1), S369–S383. https://doi.org/10.1080/15374416.2017.1342544

DuPaul, G. J., Kern, L., Dever, B. V., Ariol, M., Morse, S., Chunta, A., Rosenthal, E., Ahn, J. H., Chen, R., Franklin-Gillette, S., Jung, H. J., Lam, J. T. Y., Nelson, A., & Shi, C. (2026). Face-to-face vs. online behavioral parent training for young children with ADHD: Child and parent outcomes. *Journal of Clinical Child & Adolescent Psychology*, *55*(6), 843–858. https://doi.org/10.1080/15374416.2026.2687069

Eldridge, S., Campbell, M. K., Campbell, M. J., Drahota, A. K., Giraudeau, B., Reeves, B. C., Siegfried, N., & Higgins, J. P. T. (2021). *Revised Cochrane risk of bias tool for randomized trials: Additional considerations for cluster-randomized trials (RoB 2 CRT)* (18 March 2021). RoB 2 Development Group. https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-cluster-randomized-trials

Enebrink, P., Högström, J., Forster, M., & Ghaderi, A. (2012). Internet-based parent management training: A randomized controlled study. *Behaviour Research and Therapy*, *50*(4), 240–249. https://doi.org/10.1016/j.brat.2012.01.006

Fernando, L. M. N., Sim, W. H., Jorm, A. F., Rapee, R. M., Lawrence, K. A., & Yap, M. B. H. (2018). Parenting Resilient Kids (PaRK), an online parenting program to prevent anxiety and depression problems in primary school-aged children: Study protocol for a randomised controlled trial. *Trials*, *19*(1), 236. https://doi.org/10.1186/s13063-018-2605-8

Florean, I. S., Dobrean, A., Păsărelu, C. R., Georgescu, R. D., & Milea, I. (2020). The efficacy of Internet-based parenting programs for children and adolescents with behavior problems: A meta-analysis of randomized clinical trials. *Clinical Child and Family Psychology Review*, *23*(4), 510–528. https://doi.org/10.1007/s10567-020-00326-0

Ghaderi, A., Kadesjö, C., Björnsdotter, A., & Enebrink, P. (2018). Randomized effectiveness trial of the Family Check-Up versus Internet-delivered Parent Training (iComet) for families of children with conduct problems. *Scientific Reports*, *8*(1), 11486. https://doi.org/10.1038/s41598-018-29550-z

Ghani, N. D. H., Mohamad Fadzil, N., Mohammed, Z., Abd Rahman, M. H., Hassan, R., & Mohd Rasdi, H. F. (2026). Effectiveness of MyKidEye in improving parents' knowledge, attitude, and practice on children eye health care. *PLOS ONE*, *21*(7), e0353767. https://doi.org/10.1371/journal.pone.0353767

Grajdan, M. M. V., Etel, E., Farrell, L. J., & Donovan, C. L. (2024). A systematic review of parental involvement in digital cognitive behavioural therapy interventions for child anxiety. *Clinical Child and Family Psychology Review*, *27*(3), 702–728. https://doi.org/10.1007/s10567-024-00505-3

Hammersley, M. L., Okely, A. D., Batterham, M. J., & Jones, R. A. (2019). An Internet-based childhood obesity prevention program (Time2bHealthy) for parents of preschool-aged children: Randomized controlled trial. *Journal of Medical Internet Research*, *21*(2), e11964. https://doi.org/10.2196/11964

Havighurst, S. S., Mangelsdorf, S. N., Boswell, N., Little, J., Zhang, A., Gleeson, K., Hussain, A., Harley, A., Radovini, A., & Kehoe, C. E. (2024). A self-paced online emotion socialization intervention for parents of children with challenging behavior: Tuning in to Kids OnLine. *Frontiers in Psychology*, *15*, 1393708. https://doi.org/10.3389/fpsyg.2024.1393708

Higgins, J. P. T., Savović, J., Page, M. J., & Sterne, J. A. C. (Eds.). (2019). *Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)* (Detailed guidance, 22 August 2019). RoB 2 Development Group. https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

Högström, J., Enebrink, P., & Ghaderi, A. (2013). The moderating role of child callous-unemotional traits in an Internet-based parent-management training program. *Journal of Family Psychology*, *27*(2), 314–323. https://doi.org/10.1037/a0031883

Högström, J., Enebrink, P., Melin, B., & Ghaderi, A. (2015). Eighteen-month follow-up of Internet-based parent management training for children with conduct problems and the relation of homework compliance to outcome. *Child Psychiatry & Human Development*, *46*(4), 577–588. https://doi.org/10.1007/s10578-014-0498-7

Jent, J. F., Golson, M., Peskin, A., Rothenberg, W. A., Salem, H., Weinstein, A., Davis, E., Parlade, M., Stokes, J., Brown, T., Berkovits, M., & Garcia, D. (2025). Pocket Parent-Child Interaction Therapy (PCIT) Online for young children with disruptive behaviors: Open trial. *Journal of Medical Internet Research*, *27*, e69887. https://doi.org/10.2196/69887

McWhirter, A. C., Hails, K. A., Sileci, A. C. B., Mauricio, A. M., & Stormshak, E. A. (2026). The Family Check-Up Online: Investigating patterns of engagement in a digital intervention for parenting early adolescents. *Journal of Prevention*, *47*(1), 131–149. https://doi.org/10.1007/s10935-025-00876-5

Mishina, K., Baumel, A., Kinnunen, M., Ristkari, T., Heinonen, E., Hinkka-Yli-Salomäki, S., & Sourander, A. (2025). BePresent universal Internet-based parenting intervention: Single-arm pre-post intervention study. *Journal of Medical Internet Research*, *27*, e65391. https://doi.org/10.2196/65391

Morgan, A. J., Rapee, R. M., Salim, A., Goharpey, N., Tamir, E., McLellan, L. F., & Bayer, J. K. (2017). Internet-delivered parenting program for prevention and early intervention of anxiety problems in young children: Randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(5), 417–425.e1. https://doi.org/10.1016/j.jaac.2017.02.010

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Nyström, C. D., Sandin, S., Henriksson, P., Henriksson, H., Trolle-Lagerros, Y., Larsson, C., Maddison, R., Ortega, F. B., Pomeroy, J., Ruiz, J. R., Silfvernagel, K., Timpka, T., & Löf, M. (2017). Mobile-based intervention intended to stop obesity in preschool-aged children: The MINISTOP randomized controlled trial. *The American Journal of Clinical Nutrition*, *105*(6), 1327–1335. https://doi.org/10.3945/ajcn.116.150995

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Palmer, M. L., Beckley-Hoelscher, N., Shearer, J., Kostyrka-Allchorne, K., Robertson, O. M., Koch, M., Pearson, O., Slovak, P., Day, C., Byford, S., Goldsmith, K., Waite, P., Creswell, C., & Sonuga-Barke, E. J. S. (2023). The effectiveness and cost-effectiveness of a universal digital parenting intervention designed and implemented during the COVID-19 pandemic: Evidence from a rapid-implementation randomized controlled trial within a cohort. *Journal of Medical Internet Research*, *25*, e44079. https://doi.org/10.2196/44079

Peng, Z., Wen, L. M., & Lau, P. W. C. (2025). Effectiveness of a parent-based eHealth intervention on physical activity, dietary behaviors, and sleep in preschoolers: Randomized controlled trial. *Journal of Medical Internet Research*, *27*, e70886. https://doi.org/10.2196/70886

Prinz, R. J., Metzler, C. W., Sanders, M. R., Rusby, J. C., & Cai, C. (2022). Online-delivered parenting intervention for young children with disruptive behavior problems: A noninferiority trial focused on child and parent outcomes. *Journal of Child Psychology and Psychiatry*, *63*(2), 199–209. https://doi.org/10.1111/jcpp.13426

Reardon, T. C., Ukoumunne, O. C., Dodd, H. F., Halliday, G., Hill, C., Jasper, B., Jones, B., Lawrence, P. J., Morgan, F., Placzek, A., Rapee, R. M., Violato, M., Yu, S., MYCATS Team, & Creswell, C. (2026a). Parent-led CBT delivered via online and telephone support alongside usual school practice versus usual school practice only for young children identified as at risk for anxiety disorders through screening in schools: A cluster randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *67*(6), 712–726. https://doi.org/10.1111/jcpp.70119

Reardon, T. C., Ukoumunne, O. C., Taylor, L. M. W., Ball, S., Violato, M., Larkin, M., Fisk, J., Halliday, G., Williamson, V., Gray, A., Hill, C., Jasper, B., Macdonald, I., Morgan, F., Pollard, J., Sancho, M., Sniehotta, F. F., Spence, S. H., Stainer, J., ... Creswell, C. (2026b). Screening-to-intervention pathway for child anxiety problems alongside usual school practice versus usual school practice only (iCATS-i2i): A cluster-randomised, controlled trial in primary schools in England. *The Lancet Psychiatry*, *13*(5), 358–371. https://doi.org/10.1016/s2215-0366(26)00064-700064-7)

Sandborg, J., Downing, K. L., Orellana, L., Taylor, R. W., Barnett, L. M., Carson, V., & Hesketh, K. D. (2025). Six-month intervention effect of a digital movement behavior intervention on parent- and child intermediary outcomes—results from the Let's Grow randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, *22*(1), 78. https://doi.org/10.1186/s12966-025-01764-1

Sanders, M. R., Baker, S., & Turner, K. M. T. (2012). A randomized controlled trial evaluating the efficacy of Triple P Online with parents of children with early-onset conduct problems. *Behaviour Research and Therapy*, *50*(11), 675–684. https://doi.org/10.1016/j.brat.2012.07.004

Sim, W. H., Fernando, L. M. N., Jorm, A. F., Rapee, R. M., Lawrence, K. A., Mackinnon, A. J., & Yap, M. B. H. (2020). A tailored online intervention to improve parenting risk and protective factors for child anxiety and depression: Medium-term findings from a randomized controlled trial. *Journal of Affective Disorders*, *277*, 814–824. https://doi.org/10.1016/j.jad.2020.09.019

Sloan, C. J., Cai, Q., & Gewirtz, A. H. (2026). Who benefits more in group-based and individual telehealth parenting programs: Moderators and distal child outcomes among military families. *Research on Child and Adolescent Psychopathology*, *54*(4), 489–503. https://doi.org/10.1007/s10802-026-01496-1

Sourander, A., McGrath, P. J., Ristkari, T., Cunningham, C., Huttunen, J., Lingley-Pottie, P., Hinkka-Yli-Salomäki, S., Kinnunen, M., Vuori, M., Gyllenberg, D., Suominen, A., & Unkila-Kallio, L. (2016). Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: A randomized clinical trial. *JAMA Psychiatry*, *73*(4), 378–387. https://doi.org/10.1001/jamapsychiatry.2015.3411

Spencer, C. M., Topham, G. L., & King, E. L. (2020). Do online parenting programs create change?: A meta-analysis. *Journal of Family Psychology*, *34*(3), 364–374. https://doi.org/10.1037/fam0000605

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Stormshak, E. A., Seeley, J. R., Caruthers, A. S., Cardenas, L., Moore, K. J., Tyler, M. S., Fleming, C. M., Gau, J., & Danaher, B. (2019). Evaluating the efficacy of the Family Check-Up Online: A school-based, eHealth model for the prevention of problem behavior during the middle school years. *Development and Psychopathology*, *31*(5),

1873–1886. https://doi.org/10.1017/S0954579419000907

Sung, J. Y., Mumper, E., & Schleider, J. L. (2021). Empowering anxious parents to manage child avoidance behaviors: Randomized control trial of a single-session intervention for parental accommodation. *JMIR Mental Health*, *8*(7), e29538. https://doi.org/10.2196/29538

Taylor, L., Giles, S., Howitt, S., Ryan, Z., Brooks, E., Radley, L., Thomson, A., Whitaker, E., Knight, F., Hill, C., Violato, M., Waite, P., Raymont, V., Yu, L.-M., Harris, V., Williams, N., & Creswell, C. (2022). A randomised controlled trial to compare clinical and cost-effectiveness of an online parent-led treatment for child anxiety problems with usual care in the context of COVID-19 delivered in Child and Adolescent Mental Health Services in the UK (Co-CAT): A study protocol for a randomised controlled trial. *Trials*, *23*(1), 942. https://doi.org/10.1186/s13063-022-06833-5

Tran, T. N., Nguyen, T. T. A., Nguyen, T. H. N., Le, T. L., & Le, T. T. H. (2026). Effectiveness of a blended online behavioural parent training program for Vietnamese parents of children with ADHD: A randomized controlled trial with 3-month follow-up. *JCPP Advances*, *6*(2), e70150. https://doi.org/10.1002/jcv2.70150

Vigerland, S., Ljótsson, B., Bergdahl Gustafsson, F., Hagert, S., Thulin, U., Andersson, G., & Serlachius, E. (2016). Internet-delivered cognitive behavioural therapy for children with anxiety disorders: A randomised controlled trial. *Behaviour Research and Therapy*, *76*, 47–56. https://doi.org/10.1016/j.brat.2015.11.006

Williams, N., Yu, L.-M., & Creswell, C. (2022). *Child Anxiety Treatment in the context of COVID-19 (Co-CAT): Statistical analysis plan, Version 4.0* (ISRCTN12890382). Oxford Primary Care Clinical Trials Unit. https://www.isrctn.com/editorial/retrieveFile/f1d970a6-d7f7-4bb2-9ebc-d844df0d092e/38882

Yu, S., Reardon, T. C., Ukoumunne, O. C., Njuguna, R. G., Dodd, H. F., Halliday, G., Hill, C., Jasper, B., Jones, B., Lawrence, P. J., Morgan, F., Placzek, A., Rapee, R. M., MYCATS Team, Violato, M., & Creswell, C. (2026). Estimating the economic value of supported online parent-led cognitive behavioural therapy to prevent child anxiety problems: Economic evaluation of the MYCATS cluster randomized controlled trial. *Journal of Medical Internet Research*, *28*, e87123. https://doi.org/10.2196/87123

Zhou, P., Liu, W., & Li, D. (2026). The impact of parental engagement in an electronic health (eHealth) intervention on physical activity, dietary behaviors, and sleep in preschool-aged children. *International Journal of Environmental Research and Public Health*, *23*(3), 345. https://doi.org/10.3390/ijerph23030345

Zhou, P., Song, H., Lau, P. W. C., Shi, L., & Wang, J. (2024). Effectiveness of a parent-based eHealth intervention for physical activity, dietary behavior, and sleep among preschoolers: Protocol for a randomized controlled trial. *JMIR Research Protocols*, *13*, e58344. https://doi.org/10.2196/58344

Supplementary Files

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

- CleanednewT02S1Supplement.pdf