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.

Beyond the Immediate Task: Durability and Everyday Relevance of Process and Person Praise in Children: A Systematic Review and Stratified Meta-Analysis

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: process praise, person praise, generic language, child motivation, persistence, mindset, durability, naturalistic observation, systematic review, stratified meta-analysis

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11795652/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.

Beyond the Immediate Task: Durability and Everyday Relevance of Process and Person Praise in Children: A Systematic Review and Stratified Meta-Analysis

Alexander Kuznetsov, Ph.D.
Center for Educational & Psychological Assessment, 5 Steps Academy Private Limited, Singapore | alex@5steps.edu.sg | ORCID: 0009-0002-6652-3544

Highlights

- PRISMA 2020 synthesis of 41 independent child praise cohorts (N = 5,088).
- Immediate lab setbacks favor process over person praise (g = +0.68; PEESE +0.26).
- Same-session process praise equals neutral feedback across 16 cohorts (g = +0.11).
- Delayed post-withdrawal praise evidence rests on two small cohorts (g = +0.46, *ns*).
- Prospective family cohorts show null mindset (r = +0.01) and exam/IQ links (+0.04).

Abstract

Does praising effort instead of ability build lasting motivation in children's daily lives? Or does the famed benefit depend on scripted laboratory setbacks pitted against fragile ability labels? Adhering to PRISMA 2020 standards, we reconciled 36 legacy bibliographic exports (1,870 unique APA records, with 789 topic candidates adjudicated at full text) against a six-database harvest across OpenAlex, Crossref, Semantic Scholar, ERIC, Europe PMC/PubMed, and arXiv (2,495 deduplicated records) plus first-hand inspection of 74 primary reports and online supplements spanning 1998 to 2026. Cross-study family linkage resolved 42 quantitative study units into 37 study families and k = 41 strictly independent cohorts (N = 5,088 analytic participants; 5,030 unique youths; 133 primary effect sizes alongside cluster-confounded sensitivity rows). Restricted maximum likelihood (REML) random-effects models separated temporal and ecological distance (Subgroup A), comparator wording (Subgroup B), and endpoint realism (Subgroup C). Incompatible estimands were never pooled. During same-session laboratory tasks after staged failures, process or nongeneric praise beat person or generic praise (k = 23, g = +0.68, 95% CI [+0.43, +0.93], I^2 = 85%), yet funnel asymmetry was pronounced (Egger p = .050; PEESE-adjusted g = +0.26), preregistered replication was null (g = +0.05), and older youths occasionally showed attributional or behavioral backfire. Against neutral or objective feedback, immediate process praise yielded no gain across 16 cohorts (N = 2,229, g = +0.11, 95% CI [−0.04, +0.27], p = .160; mixed-effects comparator test $Q_M(2)$ = 11.39, p = .0034, R^2 = 23.7%). Person-praise harm drove those three-arm gaps. Post-withdrawal behavioral follow-ups 10 to 21 days later rested on two small samples marred by asymmetric memory exclusions (g = +0.46, 95% CI [−0.01, +0.94], p = .058). In naturalistic caregiver speech (k = 10, N = 1,295), overall correlations were weak (r = +0.10, Knapp–Hartung 95% CI [−0.001, +0.204]) and dropped to zero across all four multi-wave prospective home and diary cohorts spanning 6 months to 7 years (N = 525, r = +0.05, 95% CI [−0.04, +0.14], p = .289), whether predicting growth mindset questionnaires (k = 4, N = 722, r = +0.01, p = .775) or standardized achievement and IQ (k = 3, N = 310, r = +0.04, p = .548). Adults should drop fixed ability labels during setbacks rather than expecting effort praise alone to lift long-term trajectories.

Keywords: process praise; person praise; generic language; child motivation; persistence; mindset; durability; naturalistic observation; systematic review; stratified meta-analysis

1. Introduction

1.1. Praise Wording, Comparator Logic, and the Gap Between Immediate Scripts and Everyday Life

Adults praise children in distinct registers. A parent or teacher may highlight the tactic a child tried, applaud a finished drawing, or label an enduring personal trait. Grammar is only the surface difference. Process statements ('you worked hard' or 'you found a good way to do it') point toward controllable tactics, product remarks endorse a tangible outcome, and person or generic labels ('you are so smart' or 'you are a good drawer') invite inferences about fixed diagnostic capacity (Cimpian et al., 2007; Kamins & Dweck, 1999; Mueller & Dweck, 1998). Real-world praise in kitchens and classrooms also varies along sincerity, hyperbole, contingency on actual progress, autonomy support, and age-graded social pragmatics (Amemiya & Wang, 2018; Brummelman et al., 2016; Corpus & Good, 2021; Henderlong & Lepper, 2002). Popular parenting guides and teacher workshops have compressed twenty-five years of heterogeneous findings into one tidy rule: swap person praise for process praise, and children will adopt incremental mindsets, persist through frustration, and earn higher grades.

Does the empirical record actually support that causal chain? Testing it demands separating three inferential leaps that narrative reviews routinely blur. Start with outcome realism. Seconds after a scripted error in a puppet vignette, a preschooler says their doll would redraw the spoiled picture tomorrow (Cimpian et al., 2007; Kamins & Dweck, 1999; Zentall & Morris, 2010, 2012). That immediate verbal choice shows how wording colors acute error appraisal, yet nobody waited until Tuesday to see whether the child voluntarily reopened a hard task. Comparator logic poses a second trap. When a brief trial pits effort praise solely against intelligence praise, a positive gap cannot tell us whether effort wording lifted resilience above neutral informative feedback or whether ability labels triggered contingent self-worth fragility after failure (Mueller & Dweck, 1998; Skipper & Douglas, 2012; Zhao et al., 2017). The reverse confound strikes Corpus and Lepper (2007, Study 2), whose delayed classroom analysis combined person, product, and process arms against neutral feedback and thus showed post-praise engagement without isolating phrasing differences. Naturalistic family talk introduces a third boundary. Correlating toddler home transcripts or nightly parent diaries with mindsets and exam scores months or years later (Ferrar et al., 2019; Francis et al., 2025; Gunderson et al., 2013, 2018b; Li et al., 2023; Pomerantz & Kempner, 2013) tracks ongoing socialization rather than withdrawn experimental treatments, because evocative child traits, warmth, and unrecorded parental speech persist throughout follow-up.

1.2. Why Undifferentiated Meta-Analytic Pooling Fails: Three Structural Stratification Axes

Why avoid a single grand-mean effect size? Averaging puppet vignettes, immediate tangram counts, free-play classroom observations, and seven-year home-visit correlations into one summary number violates causal and psychometric commensurability. A standardized mean difference captured three minutes after an insoluble puzzle estimates an acute framing shock, whereas a correlation between maternal process-praise share at age two and fourth-grade math achievement reflects observational sorting across years of family life. Equating those two estimands erases the exact boundary conditions that parents, clinicians, and teachers need.

We therefore structured the synthesis around three pre-specified stratification axes. Subgroup A separates temporal horizon and ecological setting: Level 1 covers same-session laboratory responses recorded right after staged setbacks or successes; Level 2 covers delayed behavioral or preference observations collected days or weeks after experimental feedback ended; and Level 3 covers naturalistic caregiver praise in homes, daily diaries, or dyadic tasks under continuous relational exposure (splitting prospective multi-wave cohorts from concurrent dyadic or retrospective designs). What about control conditions? Subgroup B indexes comparator precision: direct process or effort versus person or ability praise (B1), episode-specific nongeneric versus category-wide generic evaluation (B2), process or effort praise versus neutral, objective, or non-attributional feedback (B3), and combined or non-inflated praise versus neutral or baseline exposure (B4). Subgroup C orders outcome realism across four tiers: hypothetical puppet or story choices and verbal preference rankings

(C1), directly observed behavioral persistence or task performance (C2), implicit-theory or growth mindset questionnaires (C3), and standardized school achievement or cognitive test scores (C4).

1.3. Objectives of the Present Review and Stratified Meta-Analysis

Prior narrative summaries of classroom and parental praise have leaned heavily on a handful of early demonstrations while leaving large bibliographic exports, multi-experiment tables, and online appendices unexamined (including Pomerantz & Kempner, 2013; Online Appendices B and C of Gunderson, Donnellan, et al., 2018a; Brummelman et al., 2014a, 2014b; Stevens et al., 2011; and preregistered replications through 31 March 2026). This PRISMA 2020 systematic review and stratified meta-analysis audits the complete 1998–2026 record. Four concrete questions guide our analysis. First, how much of the immediate Level 1 laboratory contrast survives small-study asymmetry adjustments, preregistered replications, and neutral-control comparisons (Level 1b)? Second, does Level 2 experimental evidence show lasting behavioral advantages of process over person praise once feedback ceases? Third, across Level 3 prospective home and diary cohorts, does everyday caregiver process praise predict later child mindset, persistence, or school achievement, and do within-family contrasts Var(A − B) reflect process-praise gains or person-praise harms? Fourth, do primary-source numerical errors, post-randomization exclusions, and single-classroom confounding change headline conclusions?

2. Methods

2.1. Search Strategy, Two-Stream Register Reconciliation, and PRISMA 2020 Selection Flow

Reporting adheres to the PRISMA 2020 statement (Page et al., 2021). No prospective protocol registration is claimed. Two fully audited bibliographic streams formed our evidence base (Figure 1). Stream 1 reconciled 36 legacy project export files holding 3,260 raw bibliographic entries across 1,870 unique APA record identifiers (1,802 from RIS batches and 68 from disjoint print batches), which contained 20 foundational anchor papers and 789 topic-relevant candidates slated for full-text adjudication. Stream 2 extended coverage from Mueller and Dweck's (1998) initial trials through 31 March 2026 via programmatic Boolean queries paired with backward reference and forward citation snowballing from 19 core praise papers across six open scholarly indices: OpenAlex, Crossref, Semantic Scholar, ERIC, Europe PMC/PubMed, and arXiv. Boolean syntax crossed praise and compliment terms ('process praise', 'person praise', 'effort praise', 'ability praise', 'intelligence praise', 'generic praise', 'inflated praise') with developmental populations (infant, toddler, preschool, child, pupil, adolescent, parent, mother, teacher) and motivational, mindset, persistence, or academic endpoints.

How did those two streams converge? Across the six APIs and citation graphs, 2,495 unique deduplicated records emerged and were matched against all 789 archival candidates by DOI, normalized title, and first-author/year keys. Title and abstract screening removed 2,167 off-topic, adult-workplace, or non-evaluative hits and retained 328 praise-specific records. Secondary eligibility checks excluded 190 papers that tracked routine classroom behavior-specific praise (BSP) compliance without attributional or genericness contrasts (n = 98) or sampled only university adults, commentaries, and book chapters lacking child data (n = 92), which left 138 empirical child and adolescent candidates. Full-text acquisition assembled 74 verified primary documents and online supplements: 47 open-access PDFs and PubMed Central XMLs, 20 newly retrieved paywalled journal articles, 4 publisher supplements, and 3 auxiliary files. From our 28-item paywalled target list, we secured and audited 20 full texts in entirety (including all 8 Priority-1 critical studies). Complete metadata checks confirmed that the 8 unretrieved citations were peripheral boundary items outside child process-versus-person synthesis, such as pre-1998 publications, college-student samples already captured in an institutional technical report (Haimovitz & Corpus, 2011), adult teacher-talk counts, or third-party trait remarks duplicated by full texts from the same laboratory (see Supplement S1, Table S1).

Figure 1

PRISMA 2020 Flow Diagram Separating Bibliographic Records, Verified Documents, Study Families, and Independent Participant Cohorts.

PRISMA 2020 Stage	Stream 1 (Legacy Register) & Primary Synthesis Flow	Stream 2 (6-Database Harvest) & Exclusion / Boundary Audit
1. Identification	Legacy Search Exports (36 Export Files): 3,260 raw bibliographic occurrences -> n = 1,870 deduplicated APA record identifiers (1,802 RIS + 68 print), including 20 foundational anchor sources (10 primary, 8 contextual/boundary, 2 narrative reviews).	6 Open Databases + Bidirectional Snowballing (1998–2026; 19 Seeds): OpenAlex, Crossref, Semantic Scholar, ERIC, Europe PMC/PubMed, and arXiv -> n = 2,495 unique deduplicated records retrieved; n = 2,167 automated off-topic/adult/non-praise records excluded; n = 328 topic-relevant praise records retained.
2. Screening	Legacy Candidate Queue Re-Assessed: n = 789 topic-relevant candidate records adjudicated at primary-text stage and cross-matched via DOI, normalized title, and first-author/year keys against the 6-database harvest.	Title & Abstract Screening (n = 328): 137 open-access + 191 paywalled candidates -> n = 138 empirical child/adolescent praise contrast hits retained; n = 190 excluded (n = 98 classroom BSP compliance without attributional contrast; n = 92 adult/college samples, reviews, or commentaries).
3. Retrieval	Consolidated Full-Text Archive (n = 74 Verified Reports & Supplements): 47 deduplicated open-access PDFs & PMC XMLs, 20 acquired paywalled primary journal texts, 4 publisher online supplements, and 3 auxiliary records (all legacy priority primary texts and Online Appendices B & C verified in full).	Paywalled Registry Audit (n = 28): 20 / 28 obtained and integrated in full text (including 8 / 8 [100%] Priority-1 critical studies); n = 8 unretrieved citations verified via complete metadata as non-blocking boundary works outside child process-vs.-person contrasts (Supplement S1, Table S1).
4. Eligibility & Linkage	Unit-of-Analysis Linkage & Primary-Source Audit (74 Verified Documents): Linked 4 shared longitudinal/video study families (FAM_GUNDERSON_TODDLER, N = 53; FAM_SOMMERVILLE_INFANT, N = 29; FAM_GUNDERSON_PERCEIVED, 58 cross-stratum overlap; FAM_BRUMMELMAN_DUTCH_HOME, N = 116) and audited 22 primary-source statistical discrepancies (Tables S2 & S4).	Non-Pooled Contextual & Cluster-Sensitivity Disposition: n = 25 contextual/non-pooled full texts synthesized qualitatively (6 reviews/outlier reanalyses; 19 DPICS/teacher/boundary reports without child contrasts; Table S3); n = 1 intact-classroom study (Zarrinabadi et al., 2023; N = 63, 1 class/arm) isolated in Pre–Post DiD + DEFF sensitivity (Table S6).
5. Included (Stratified)	Verified Primary Quantitative Corpus: 42 quantitative study units across 37 study families -> k = 41 independent cohorts; N = 5,088 analytic cohort participants [5,030 unique children] (133 primary effect sizes + 4 DiD sensitivity rows = 137 audited rows in	Stratified REML Synthesis (Never Naively Pooled): Level 1a Immediate Lab Process/Nongeneric vs. Person/Generic (k = 23, N = 2,345, g = +0.68; Egger p = .050, PEESE g = +0.26); Level 1b

	T04_EFFECT_SIZES_AUDIT.csv).	Immediate Process vs. Neutral (k = 16, N = 2,229, g = +0.11, p = .160, ns); Level 2a Delayed 10–21d Post-Withdrawal (k = 2, N = 97, g = +0.46, p = .058, ns); Level 3a All Caregiver Process Praise (k = 10, N = 1,295, r = +0.10, K–H 95% CI [−0.001, +0.204]); Prospective Home/Diary Cohorts (k = 4, N = 525, r = +0.05, p = .289, ns); Comparator Q_M(2) = 11.39, p = .0034, R^2 = 23.7%; Everyday Mindset (k = 4, N = 722, r = +0.01, p = .775, ns); School Achievement & IQ (k = 3, N = 310, r = +0.04, p = .548, ns).

Note. Counts separate raw bibliographic occurrences (*n* = 3,260), deduplicated legacy APA identifiers (*n* = 1,870), six-database deduplicated harvest records (*n* = 2,495), verified full-text reports and online supplements (*n* = 74), quantitative study units (*n* = 42), study families (*n* = 37), and strictly independent participant cohorts (*k* = 41 cohorts; *N* = 5,088 analytic cohort participants [5,030 unique children] after subtracting 58 children shared across non-overlapping strata in Gunderson, Donnellan, et al., 2018a, Studies 1 and 2).

2.2. Eligibility Criteria and Strict Unit-of-Analysis Linkage

Three inclusion rules governed primary quantitative synthesis. Eligible studies had to (a) sample infants, children, or adolescents with mean age below 18 years, (b) manipulate or naturalistically code process/effort/nongeneric praise, person/ability/generic praise, or defined neutral and non-inflated praise comparators, and (c) report sufficient statistics on at least one motivational, persistence, self-evaluative, mindset, behavioral, or academic outcome to derive a standardized effect size and sampling variance. What happened to adjacent child studies that lacked an attributional contrast? Papers examining clinical tact maintenance under token economies, belief priming within effort praise alone, overheard peer remarks, or descriptive caregiver talk baselines (Pepper et al., n.d.; Swenson et al., 2016) entered structured qualitative boundary synthesis (Section 3.4; Supplement S1, Table S3) rather than distorting incompatible meta-analytic pools. One quasi-experimental EFL report (Zarrinabadi et al., 2023; $N = 63$) assigned a single intact classroom of 21 teenagers per condition under continuous teacher feedback amid a massive pretest gap; we quarantined its data in dedicated difference-in-differences (DiD) and cluster design-effect sensitivity models.

Publications are not cohorts. We never equated bibliographic citations, multi-study papers, and independent participant samples. Forensic cross-checking across all full texts uncovered four shared-participant families spanning multiple papers: (1) FAM_GUNDERSON_TODDLER joined Gunderson et al. (2013; $N = 53$) with Gunderson, Sorhagen, et al. (2018b; $N = 50$) as one longitudinal Chicago home-visit cohort (COH_CHI_TODDLER_53); (2) FAM_SOMMERVILLE_INFANT joined Lucca et al. (2019, Study 2; $N = 29$) with Radovanovic et al. (2023, Experiment 1; $N = 29$), because the latter recoded the exact same 29 infant laboratory videos (COH_UW_INFANT_S2_29); (3) FAM_GUNDERSON_PERCEIVED linked Studies 1 ($N = 307$) and 2 ($N = 267$) of Gunderson, Donnellan, et al. (2018a), whose 58 shared elementary pupils were placed in disjoint comparator strata (B4 vs. B1) and never combined in one estimate; and (4) FAM_BRUMMELMAN_DUTCH_HOME joined Brummelman et al. (2014a, Study 2; $N = 114$) with Brummelman et al. (2017; $N = 120/116$) on the same four-wave Dutch family cohort (COH_NL_HOME_116). Across 42 quantitative study units and 37 study families, exactly $k = 41$ strictly independent cohorts ($N = 5{,}088$ analytic participants; 5,030 unique children and adolescents; 133 primary effect sizes plus 4 sensitivity rows = 137 audited rows; Table 1) entered quantitative synthesis.

Table 1
Independent Primary Participant Cohorts Included in the Stratified Quantitative Synthesis ($k = 41$ Cohorts Across 42 Study Units, $N = 5{,}088$).

Cohort ID	Primary Report(s) / Cohort Label	N	Mean Age (yr)	Max Lag	Sub. A	Sub. B	Sub. C	m
MD98_S1	Mueller & Dweck (1998, Study 1)	123	10.7	Same session	A1	B1,B3	C1,C2	6*
MD98_S2	Mueller & Dweck (1998, Study 2)	51	10.7	Same session	A1	B1	C1	1
MD98_S3	Mueller & Dweck (1998, Study 3)	88	10.7	Same session	A1	B1,B3	C1,C2	5*
MD98_S4	Mueller & Dweck (1998, Study 4)	50	10.7	Same session	A1	B1,B3	C3	2*
MD98_S5	Mueller & Dweck (1998, Study 5)	46	10.7	Same session	A1	B1,B3	C1,C2	5*
MD98_S6	Mueller & Dweck (1998, Study 6)	48	10.7	Same session	A1	B1,B3	C1,C2,C3	6
KD99_S2	Kamins & Dweck (1999, Study 2)	64	5.5	Same session	A1	B1	C1,C3	5*
CIMP07	Cimpian, Arce, Markman, & Dweck (2007)	24	4.4	Same session	A1	B2	C1	2*
CIMP10_45YO	Cimpian (2010 — 4-5-year-olds)	49	5.1	Same session	A1	B2	C1	1*
CIMP10_67YO	Cimpian (2010 — 6-7-year-olds)	48	6.9	Same session	A1	B2	C1	2
ZM10	Zentall & Morris (2010)	135	5.7	Same session	A1	B2	C1	2*
ZM12	Zentall & Morris (2012)	30	5.5	Same session	A1	B2	C1	2*
MZ14	Morris & Zentall (2014)	95	5.7	Same session	A1	B1,B3	C1	4*
SD12_E1	Skipper & Douglas (2012, Exp. 1)	145	10.0	Same session	A1	B1,B3	C1	4*

Cohort ID	Primary Report(s) / Cohort Label	N	Mean Age (yr)	Max Lag	Sub. A	Sub. B	Sub. C	m
DROE13	Droe (2013)	87	9.5	Same session	A1	B1,B3	C1,C2	9*
BRUM14_S2	Brummelman et al. (2014b, JEP:G Study 2)	311	10.6	Same session	A1	B1,B3	C1	2
BRUM14_PS_S3	Brummelman et al. (2014a, Psych Sci Study 3)	240	9.9	Same session	A1	B3	C2	1
DUNS15	Dunsmore (2015)	143	10.2	Same session	A1	B1,B3	C2	6*
ZHAO17	Zhao, Heyman, Chen, & Lee (2017)	300	4.5	Same session	A1	B2,B3	C2	2*
XING18	Xing et al. (2018)	103	10.8	Same session	A1	B1,B3	C1,C2	6*
GLERUM20	Glerum, Loyens, Wijnia, & Rikers (2020)	108	17.2	Same session	A1	B1,B3	C1,C2,C3	8*
BRUM22_VR	Brummelman, Grapsas, & van der Kooij (2022)	192	9.9	Same session	A1	B3	C2	2
BP24	Bennett-Pierre, Chernuta, Altamimi, & Gunderson (2024)	150	8.0	Same session	A1	B1,B3	C1,C2	6*
CL07_S1	Corpus & Lepper (2007, S1)	93	10.3	21 d	A1,A2	B1,B3	C1,C2	5
CL07_S2	Corpus & Lepper (2007, Study 2)	54	4.8	10 d	A1,A2	B1,B3,B4	C2	6*
ISHI26	Ishibashi et al. (2026)	58	1.8	Same session	A1	B4	C2	1
LIUYIN26	Liu & Yin (2026)	146	12.9	Same session	A1	B4	C2	2
CHI_TODDLER_53	Gunderson et al. (2013/2018b)	53	7.5	7.0 yr	A3	B1	C3,C4	5*
POM13_110	Pomerantz & Kempner (2013)	110	10.2	0.5 yr	A3	B1	C1,C3	4*
FERRAR19_156	Ferrar et al. (2019)	156	9.3	5.7 yr	A3	B1	C4	2*
FRANCIS25_250	Francis, Boullion, & Yates (2025)	250	8.0	4.0 yr	A3	B1,B3	C2	3*
CA_ELEM_S1	Gunderson, Donnellan, et al. (2018a, Study 1)	307	10.1	Same session	A3	B4	C3	1
CA_ELEM_S2	Gunderson, Donnellan, et al. (2018a, Study 2)	267	11.4	Same session	A3	B1	C1,C3	2*
LI23_292	Li, Zhang, Wu, & Zheng (2023)	292	9.3	Same session	A3	B1,B4	C3,C4	5*
UW_INFANT_S1_29	Lucca, Horton, & Sommerville (2019, Study 1)	29	1.5	Same session	A3	B1	C2	1
UW_INFANT_S2_29	Lucca (2019 S2)/Radovanovic (E1)	29	1.5	Same session	A3	B1	C2	1*
TORONTO_ZOOM_60	Radovanovic, Soldovieri, & Sommerville (2023, Exp. 2)	60	2.0	Same session	A3	B1	C2	1*
NAHM17_KR_49	Nahm & Park (2017)	49	5.0	Same session	A3	B1	C2	1*
NL_HOME_116	Brummelman et al. (2014a S2/2017)	116	8.8	1.5 yr	A3	B4	C1	1*
LEE17_298	Lee, Kim, Kesebir, & Han (2017)	298	9.5	Same session	A3	B4	C4	1*
CARVER22_91	Carver et al. (2022)	91	4.5	1.0 yr	A3	B4	C4	2

Note. *N* = maximum verified analyzed cohort sample size without double-counting shared participants across linked papers (COH_CHI_TODDLER_53 links Gunderson et al., 2013 & 2018b; COH_UW_INFANT_S2_29 links Lucca et al., 2019 S2 & Radovanovic et al., 2023 E1; COH_NL_HOME_116 links Brummelman et al., 2014a S2 & 2017; COH_CA_ELEM_S1 and S2 share 58 participants and never enter the same stratum; cohort-summed *N* = 5,088 across 41 rows equals 5,030 unique children once that 58-child cross-stratum overlap is subtracted). Sub. A: A1 = Immediate Lab, A2 = Delayed Withdrawal, A3 = Naturalistic Caregiver Speech. Sub. B: B1 = Process vs. Person, B2 = Nongeneric vs. Generic, B3 = Process vs. Neutral/Objective, B4 = Pooled/Non-inflated vs. Neutral. Sub. C: C1 = Hypothetical/Preference, C2 = Observed Behavior, C3 = Mindset Questionnaire, C4 = Academic Achievement/IQ. *m* = extracted primary effect sizes (* flags ≥ 1 audited source discrepancy cataloged in Table S4).

2.3. Effect-Size Standardization, Covariance Modeling, Cluster Adjustments, and Critical Appraisal

Experimental contrasts (Subgroups A1, A2, and B1–B4) were standardized onto unbiased Hedges' g via the exact small-sample factor $J(df) = 1 - 3/(4 \cdot df - 1)$ (Hedges & Olkin, 1985). Positive signs ($+g$) always denote more adaptive motivation, persistence, self-regard, honesty, or incremental beliefs in the process, effort, or nongeneric arm relative to its comparator. Binary endpoints such as cheating counts in Zhao et al. (2017) or puzzle approach in Corpus and Lepper (2007) were transformed from exact 2 × 2 log-odds ratios through $d = \ln(OR) \cdot \sqrt{3}/\pi$ and scaled by $J(df)$. Naturalistic cohorts at Level 3 were analyzed on Pearson's r, stabilized via Fisher's $z = 0.5 \cdot \ln[(1 + r)/(1 - r)]$ with variance $\mathrm{Var}(z) = 1/(N - 3)$, and converted back through $r = \tanh(z)$ for display; correlations were mapped onto Hedges' g via $d = 2r/\sqrt{(1 - r^2)}$ solely in cross-level meta-regression MR-1.

Our Python pipeline (run_meta_analysis.py) enforced four statistical guardrails. First, whenever a cohort reported multiple correlated measures inside one stratum ($m > 1$), we formed a single cohort composite prior to pooling via the exact variance equation $\mathrm{Var}(\bar{y}) = [\Sigma \mathrm{Var}(y_i) + 2 \cdot \rho_w \cdot \Sigma_{i<j} \sqrt{(\mathrm{Var}(y_i)\mathrm{Var}(y_j))}]/m^2$, anchored at $\rho_w = .35$ and checked at $\rho_w = .20$ and .50. Second, classroom-clustered sampling variances were inflated by $\mathrm{DEFF} = 1 + (\bar{m} - 1) \cdot \mathrm{ICC}$ across $\mathrm{ICC} \in \{.00, .02, .05, .10\}$. Third, we never compared separate p-values for maternal process praise (r_{1y}) and person praise (r_{2y}) inside the same families. Instead, Steiger's (1980) asymptotic covariance equations gave exact paired difference variances $\mathrm{Var}(z_1 - z_2) = \mathrm{Var}(z_1) + \mathrm{Var}(z_2) - 2 \cdot \mathrm{Cov}(z_1, z_2)$ at each cohort's empirical praise intercorrelation r_{12} and across $r_{12} \in \{.20, .35, .50\}$. Fourth, random-effects models used REML τ^2 estimation (Viechtbauer, 2005) paired with DerSimonian–Laird (1986) benchmarks, Knapp–Hartung t-intervals (Knapp & Hartung, 2003), 95% prediction intervals (PI) whenever $k \geq 3$, mixed-effects meta-regressions (MR-1 to MR-3), Egger's (1997) asymmetry regression, and PET-PEESE precision corrections (Stanley & Doucouliagos, 2014). Risk of bias was rated at the result level with Cochrane RoB 2 (Higgins et al., 2019, 2021) for randomized trials and JBI instruments (JBI, 2024a, 2024b) for quasi-experimental and cohort studies, while all 22 primary-source numerical anomalies were preserved verbatim (Supplement S1, Table S4).

3. Results

3.1. Subgroup A (Level 1) and Subgroup B: Immediate Laboratory Contrasts, Neutral-Control Nulls, and Replication Boundaries

Across $k = 23$ independent laboratory cohorts ($N = 2{,}345$; $m = 56$ endpoints) pitting process, effort, or nongeneric praise directly against person, ability, or generic praise in the same session (Subgroup A, Level 1a), REML pooling yielded $g = +0.678$ (95% Wald CI [+0.430, +0.926], $p < .0001$; Knapp–Hartung 95% CI [+0.412, +0.943]; Table 2 and Figure 2, Panel A). Dispersion was massive ($\tau^2 = 0.302$, $I^2 = 85.0\%$). Because the 95% prediction interval stretches from −0.49 to +1.85, no single-session replication can count on a positive effect without specifying task framing, child age, and verbal script. Small-study bias compounded that spread. Egger's test detected significant funnel asymmetry across Level 1a cohorts (intercept $\beta_0 = +3.48$, SE = 1.67, $t(21) = 2.08$, $p = .050$), as small early trials (Zentall & Morris, 2012, $n = 30$, $g = +1.91$; Morris & Zentall, 2014, $n = 95$, $g = +1.68$; Mueller & Dweck, 1998, Studies 1–6, $g = +0.94$ to +1.52) towered over larger recent replications. Precision adjustments shrank the immediate contrast to $g = -0.165$ under PET and $g = +0.264$ under PEESE.

Comparator stratification (Subgroup B) exposed the active ingredient behind Level 1a. What happens when process praise meets a neutral baseline? Across $k = 16$ independent laboratory cohorts ($N = 2{,}229$; $m = 35$ endpoints) containing a neutral, objective-score, or non-attributional control arm alongside process praise (Level 1b), the immediate process-praise advantage vanished: $g = +0.111$ (95% Wald CI [−0.044, +0.266], $p = .160$; Knapp–Hartung 95% CI [−0.061, +0.283], $I^2 = 60.3\%$). Mixed-effects meta-regression MR-2 confirmed strong comparator moderation across experimental cohorts, $Q_M(2) = 11.39$, $p = .003$, which explained $R^2 = 23.7\%$ of between-study variance as pure process-versus-person trials (B1: $k = 17$, $g = +0.660$,

$p < .0001$) and nongeneric-versus-generic puppet scripts (B2: $k = 6$, $g = +0.792$, $p < .0001$) far outpaced process-versus-neutral comparisons ($\Delta\beta = +0.48$, $p = .0045$, and $\Delta\beta = +0.64$, $p = .0075$).

Within-trial three-arm decompositions ($k = 12$ cohorts with Process, Person, and Neutral/Outcome groups; Table 3) pinpoint why. Person praise pulls children down; process praise leaves them at baseline. In Zhao et al.'s (2017; $N = 300$) temptation-to-cheat paradigm, preschoolers' honesty after process praise (59% non-cheating; 41/100 cheated) matched baseline controls (60% non-cheating; 40/100 cheated; $g = -0.02$, $p = .885$), whereas ability praise ('you are so smart') drove up cheating (40% non-cheating; 60/100 cheated; $g = +0.45$ vs. baseline; 105.4% harm share). Brummelman et al. (2014b, Study 2; $N = 311$) found the exact same split after a competitive loss: process praise matched no praise on self-esteem and shame ($g = +0.03$), while person praise intensified shame among low-self-esteem youths (92.4% harm share on composite self-feelings; 165.4% on reversed shame, Table 3). In Kamins and Dweck (1999, Study 2; $N = 64$), process praise beat person praise on puppet vignettes ($g = +0.91$), yet all four post-setback contrasts against simple outcome approval ('that's the right way to do it') were statistically null. After repairing SE columns mislabeled as SD in Xing et al. (2018; $N = 103$, Tables 1 and 3) and checking Morris and Zentall (2014; $N = 95$), effort and nongeneric praise actually trailed objective-score or 'high-five' gestural controls slightly ($g = -0.34$ and -0.29) so the ability-praise harm share exceeded 100%. Even in Skipper and Douglas (2012, Experiment 1; $N = 145$), small gaps after an initial failure disappeared completely after a second failure, $F(2, 142) = 0.02$, exact $p = .980$ (misprinted as $p = .092$).

Recent full-text replications tighten Level 1 boundaries further. Consider preregistration first. In the sole preregistered multi-arm replication of the classic puzzle-setback paradigm (Bennett-Pierre et al., 2024; $N = 150$ ages 7–11), process praise separated neither from person praise ($g = +0.05$, 95% CI [−0.31, +0.41]) nor from no-praise control ($g = +0.17$, 95% CI [−0.25, +0.59]) on behavioral persistence or self-appraisal once inverted Cohen's d signs on page 9 were corrected. Ecological delivery showed similar nulls. When real parents delivered scripted process praise inside a virtual-reality exploration task (Brummelman et al., 2022; $N = 192$), challenge-seeking and exploration duration were unchanged ($g = -0.04$). Non-inflated praise matched no-praise controls while inflated praise curbed challenge-seeking among low-self-esteem children (Brummelman et al., 2014a, Study 3; $N = 240$, $g = 0.00$), an interaction that missed significance in half of multiverse outlier specifications (Verkoeijen et al., 2018); overheard peer effort praise failed across four preschool experiments (Zhao et al., 2022; $N = 330$); and priming inverse effort-ability beliefs among Hong Kong early adolescents turned effort praise into lower self-evaluation ($d = -0.36$) and lower intrinsic motivation ($d = -0.47$; Lam et al., 2008; $N = 160$). Can process praise actively backfire? Two youth trials say yes: Dunsmore (2015; $N = 143$ ages 9–12) observed less voluntary helping persistence after process feedback than after person ($g = -0.45$) or neutral feedback ($g = -0.50$), and Glerum et al. (2020; $N = 108$ vocational adolescents) saw effort praise trail intelligence praise slightly on problem-solving ($g = -0.19$) while triggering low-ability attributions after failure ($g = -0.54$, $p = .043$).

Table 2
Stratified Random-Effects (REML) Meta-Analysis Across Temporal/Ecological Distance (Subgroup A), Comparator Precision (Subgroup B), and Outcome Type (Subgroup C).

Stratum	Metric	k	N	m	REML Est.	95% Wald CI	95% K–H CI	p (Wald)	I^2 & [95% PI]
Subgroup A: Temporal & Ecological Distance									
Level 1a: Immediate Lab (Process/Nongeneric vs. Person/Generic)	g	23	2,345	56	+0.678	[+0.430, +0.926]	[+0.412, +0.943]	< .0001	85.0% [-0.49, +1.85]
Level 1b: Immediate Lab (Process vs. Neutral / Objective Control)	g	16	2,229	35	+0.111	[-0.044, +0.266]	[-0.061, +0.283]	.160	60.3% [-0.44, +0.66]
Level 2a: Delayed 10–21d Post-Withdrawal (Process vs. Person)	g	2	97	4	+0.461	[-0.015, +0.938]	[-2.627, +3.550]	.058	0.0% (k < 3)

Stratum	Metric	k	N	m	REML Est.	95% Wald CI	95% K–H CI	p (Wald)	I^2 & [95% PI]
Level 2b: Delayed 10–21d Post-Withdrawal (Process vs. Neutral)	g	2	97	3	+0.708	[-0.086, +1.502]	[-4.440, +5.856]	.081	45.4% (k < 3)
Level 2c: Delayed 10d Classroom (Pooled Praise vs. Neutral)	g	1	48	2	+0.999	[+0.367, +1.631]	—	.002	— (k = 1)
Level 3a (All): Naturalistic Caregiver Process Praise	r	10	1,295	23	+0.103	[+0.027, +0.177]	[−0.001, +0.204]	.008	57.6% [-0.10, +0.30]
Level 3a (Prospective): Multi-Wave Home/Diary Cohorts (6m–7yr)	r	4	525	12	**+0.048**	[-0.040, +0.135]	[-0.095, +0.189]	.289	29.0% [-0.23, +0.32]
Level 3a (Concurrent): Dyadic / Short-Window / Cross-Sectional	r	7	976	11	+0.145	[+0.030, +0.256]	[-0.023, +0.305]	.014	62.4% [-0.18, +0.44]
Level 3b: Naturalistic Overall / Non-Inflated Caregiver Praise	r	5	1,104	6	+0.075	[+0.017, +0.132]	[-0.008, +0.156]	.012	0.0% [-0.02, +0.17]
Subgroup B: Experimental Comparator Precision									
B1: Pure Process/Effort vs. Person/Ability Praise	g	17	1,759	50	+0.660	[+0.364, +0.956]	[+0.340, +0.981]	< .0001	87.1% [-0.60, +1.92]
B2: Nongeneric Episode vs. Generic Trait Evaluation	g	6	586	10	+0.792	[+0.428, +1.155]	[+0.271, +1.313]	< .0001	68.3% [-0.36, +1.95]
B3: Process/Effort vs. Neutral / Objective / Ambiguous	g	18	2,326	38	+0.142	[-0.012, +0.297]	[-0.033, +0.318]	.071	60.7% [-0.42, +0.71]
B4: Pooled / Non-Attributional Praise vs. Neutral	g	3	252	5	+0.436	[-0.185, +1.058]	[-0.929, +1.801]	.169	83.0% [-7.00, +7.88]
Subgroup C: Outcome Realism & Domain									
Exp. (Proc vs. Pers/Gen) — C1: Hypothetical Puppet / Verbal Rating	g	19	1,754	38	+0.767	[+0.503, +1.032]	[+0.483, +1.052]	< .0001	81.7% [-0.37, +1.90]
Exp. (Proc vs. Pers/Gen) — C2: Observed Task Engagement / Solved	g	12	1,343	18	+0.474	[+0.161, +0.786]	[+0.123, +0.825]	.003	81.5% [-0.67, +1.62]
Exp. (Proc vs. Pers/Gen) — C3: Mindset / Attribution Questionnaire	g	4	270	4	+0.612	[-0.209, +1.434]	[-0.721, +1.946]	.144	88.8% [-3.18, +4.40]
Exp. (Proc vs. Neutral) — C1: Hypothetical / Verbal Rating	g	12	1,353	18	+0.110	[-0.131, +0.351]	[-0.166, +0.385]	.372	64.9% [-0.71, +0.93]
Exp. (Proc vs. Neutral) — C2: Observed Task / Exploration / VR	g	13	1,676	17	+0.236	[+0.002, +0.470]	[-0.031, +0.503]	.048	69.8% [-0.58, +1.05]
Exp. (Proc vs. Neutral) — C3: Mindset / Attribution Questionnaire	g	3	206	3	+0.218	[-0.178, +0.615]	[-0.653, +1.090]	.281	22.4% [-3.21, +3.65]
Naturalistic (Level 3) — C1: Challenge Preference / Learning Goals	r	2	373	3	+0.256	[-0.086, +0.544]	[-0.964, +0.987]	.141	92.0%
Naturalistic (Level 3) — C2: Observed Task Persistence (All Dyadic)	r	5	417	6	+0.201	[-0.001, +0.388]	[-0.100, +0.469]	.051	62.6% [-0.44, +0.71]
Naturalistic (Level 3) — C2: 4-Year Prospective Persistence Only	r	1	206	1	+0.110	[-0.027, +0.243]	[-0.027, +0.243]	.116	0.0%
Naturalistic (Level 3) — C3: Child Mindset Questionnaire	r	4	722	6	+0.009	[-0.056, +0.075]	[-0.147, +0.166]	.775	54.7% [-0.13, +0.15]
Naturalistic (Level 3) — C4: School Academic Achievement & IQ	r	3	310	8	+0.038	[-0.085, +0.159]	[-0.228, +0.298]	.548	37.2% [-0.82, +0.84]

Note. Incompatible estimands are never combined into a single grand mean. Within-cohort correlated endpoints ($m > 1$) are composited at $\rho_w = .35$ before pooling so each cohort enters at most once per stratum (Francis et al., 2025 contributes its concurrent Age-4 wave [$N = 250$] to Concurrent Level 3a, its 4-year prospective Age-8 wave [$N = 206$] to Prospective Level 3a, and one cross-wave composite to All Level 3a [$k = 10$, $N = 1{,}295$]). K–H = Knapp–Hartung *t*-distribution interval; PI = 95% prediction interval ($k \geq 3$).

Table 3

Decomposition of Twelve Three-Arm Experimental Cohorts Into Process-Praise Gain Above Neutral Feedback Versus Person-Praise Harm Below Neutral Feedback.

Three-Arm Experimental Cohort	g (Process vs. Person)	g (Process vs. Neutral)	g (Neutral vs. Person: Harm)	Share Driven by Person Harm
Mueller & Dweck (1998, S1) Persistence	+1.028	+0.190	+0.761	81.5%
Mueller & Dweck (1998, S3) Persistence	+2.101	+0.788	+0.946	62.5%
Mueller & Dweck (1998, S5) Persistence	+0.714	+0.040	+0.757	94.4%
Mueller & Dweck (1998, S6) Persistence	+0.721	+0.549	+0.220	23.8%
Kamins & Dweck (1999, S2) Persistence [vs Outcome]	+0.876	+0.497	+0.263	43.2%
Droe (2013) Post-Failure Persistence	+0.568	+0.000	+0.533	100.0%
Morris & Zentall (2014) Persistence	+1.672	-0.087	+1.715	105.2%
Brummelman et al. (2014b, JEP:G S2) Shame (rev)	+0.338	-0.221	+0.497	165.4%
Zhao et al. (2017) Behavioral Honesty (2x2)	+0.423	-0.023	+0.445	105.4%
Xing et al. (2018) Raven Working Time	+1.154	-0.055	+1.199	104.8%
Xing et al. (2018) Raven Score Change	+0.752	-0.076	+0.813	110.1%
Bennett-Pierre et al. (2024) Self-Evaluation	-0.251	-0.273	+0.026	N/A (Process ≤ Person)

Note. Positive *g* (Neutral vs. Person) means neutral or objective feedback beat person or ability praise (showing active impairment below baseline). Harm shares above 100% arise whenever process praise scored slightly below neutral control while person praise dropped much further.

3.2. Subgroup A (Level 2): Delayed Observation After Praise Withdrawal and Classroom Cluster Confounding

How much post-withdrawal durability evidence exists after 28 years of research since Mueller and Dweck (1998)? Just one primary paper (Corpus & Lepper, 2007, Studies 1 and 2) tracked children 10 to 21 days after experimental praise ended under individual randomization. Pooled across those two cohorts ($N = 97$ in the process-versus-person arms), the delayed contrast (Level 2a) fell short of statistical significance: $g = +0.461$ (95% Wald CI [−0.015, +0.938], $p = .058$; Knapp–Hartung 95% CI [−2.627, +3.550]; Table 2). Delayed process praise versus neutral feedback (Level 2b) also missed significance, $g = +0.708$ (95% Wald CI [−0.086, +1.502], $p = .081$).

Inspecting Corpus and Lepper (2007) at full text shows why neither experiment proves lasting superiority of process over person wording. Study 1 retested fourth- and fifth-graders at a 3-week median lag using verbal rankings of six gift options rather than actual tangram play, and boys showed zero process-over-person benefit. Study 2 observed preschoolers during free classroom play at a 10-day median lag via masked behavioral coding. Two design flaws undercut that preschool contrast. First, a post-randomization praise-recall screen was imposed exclusively on the three praised arms and shrank person, product, and process cells from 20/20/20 down to 11/11/16 (discarding 45% of person- and product-praised children before follow-up attrition left 11/10/13), whereas 0 of 16 neutral controls faced memory screening (14 observed; total $N = 48$ of 76 randomized). Second, the published delayed classroom test pooled person, product, and process praise together against neutral feedback (Level 2c: $g = +0.999$, 95% CI [+0.367, +1.631], $p = .0019$; mean puzzle time was 735 s in Process, 627 s in Person, and 592 s in Product vs. 426 s in Neutral). Restoring all 28

excluded preschoolers via conservative zero-difference intention-to-treat imputation shrinks the unpooled Study 2 process-versus-person gap from g = +0.56 (n = 24) to g = +0.35 (n = 40, p = .27, *ns*).

Two other multi-week studies clarify classroom boundaries without offering unconfounded post-withdrawal contrasts. Digitizing session plots in Stevens et al. (2011; N = 2 boys with autism spectrum disorder) showed that pairing tokens with behavior-specific praise, general praise, or tokens alone produced minimal acquisition differences (≤ 3 sessions to mastery) and identical retention at 1-, 2-, 4-, and 6-week probes. Zarrinabadi et al. (2023; N = 63 Iranian EFL pupils, ages 14–16) delivered six weeks of teacher ability praise, effort praise, or no praise and tested willingness to communicate, oral anxiety, enjoyment, and perceived competence two days later. Forensic checks uncovered five severe defects (Table S4): (a) one intact classroom (n = 21) per arm, which confounds treatment with teacher and peer-group effects; (b) continuous six-week delivery rather than post-withdrawal follow-up; (c) 'effort' notes containing the trait label 'you are hardworking'; (d) a +93.76-point baseline advantage on perceived competence in the Effort class (pretest M = 430.80 vs. 337.04 in Ability) that flips the true Pre–Post DiD sign (Effort +20.15 vs. Ability +27.24; true DiD = −7.09, g = −0.06), paired with an impossible paired t-ratio in Table 3 (printed t = 2.86, p = .010 vs. exact paired $t(20)$ = 1.22, p = .235, *ns* at r = .61); and (e) contradictory sample totals (Table 1 N = 44 vs. prose N = 63). Even if Zarrinabadi et al.'s four-endpoint Pre–Post DiD composite (g = +1.22) is merged with Corpus and Lepper (2007) under modest classroom clustering (ICC = .05, DEFF = 2.00; k = 3), the Knapp–Hartung 95% CI still crosses zero ([−0.461, +1.899], p_{HK} = .120, *ns*; Table S6).

3.3. Subgroup A (Level 3) and Subgroup C: Everyday Caregiver Praise Across Homes, Diaries, and Schools

What happens in everyday homes and classrooms? Across all k = 10 independent naturalistic cohorts coding caregiver process praise (N = 1,295; m = 23 associations), the overall correlation with adaptive child outcomes was r = +0.103 (95% Wald CI [+0.027, +0.177], p = .008; I^2 = 57.6%), or 1.06% of shared variance. Under Knapp–Hartung inference, its 95% CI crossed zero ([−0.001, +0.204], p_{HK} = .051; Table 2 and Figure 2, Panel B). Meta-regression MR-1 confirmed steep attenuation from immediate Level 1 laboratory scripts down to Level 3 everyday speech (β = −0.397, 95% CI [−0.785, −0.009], p = .0449; Table 4).

Splitting concurrent single-session samples from prospective multi-wave family cohorts explained Level 3 dispersion. Two small concurrent infant gear-stacking samples from one laboratory (Lucca et al., 2019, Study 1, N = 29, r = +.51; linked Study 2 / Radovanovic et al., 2023, Experiment 1 cohort, N = 29, r = +.47) plus one cross-sectional child recall survey of perceived learning goals (Gunderson, Donnellan, et al., 2018a, Study 2, N = 267, composite r = +.20) accounted for nearly the entire positive signal. Prospective family cohorts told a different story. Across all four multi-wave longitudinal home and daily-diary studies tracking children over 6 months to 7 years (Gunderson et al., 2013/2018b; Pomerantz & Kempner, 2013; Ferrar et al., 2019; Francis et al., 2025; k = 4, N = 525), naturalistic caregiver process praise yielded a complete null: r = +0.048 (95% Wald CI [−0.040, +0.135], p = .289; Knapp–Hartung 95% CI [−0.095, +0.189], I^2 = 29.0%), equal to just 0.23% of variance.

Cohort-by-cohort full-text audits clarify why everyday parental process praise fails to predict later mindsets or school achievement:

- **Pomerantz and Kempner (2013):** Auditing Tables 1–3 (N = 120 at Wave 1; N = 110 longitudinal mother–child dyads over 6 months, mean age 10.23 years, 10-day phone diaries) showed that non-attributional praise dominated maternal speech (85–87%), whereas process and person utterances were rare (7–8% each; 10-day diary M = .10 and .12). Net of Wave 1 baseline scores and total praise, maternal process praise predicted neither 6-month incremental mindset (proportion r_p = +.082, p = .41; daily frequency r_p = −.091, p = .36) nor challenge preference (r_p = +.124, p = .21; r_p = +.036, p = .72). Person praise did the work instead. Maternal person praise predicted lower incremental mindset (frequency r_p = −.224, p = .023) and weaker challenge preference (frequency r_p = −.229, p = .020; proportion r_p = −.191,

p = .052; Table 3 on p. 2044 dropped minus signs on unstandardized *B* = 0.79 and 1.64 alongside negative β = −.21 *and −.17*).

- **Ferrar et al. (2019):** This Canadian cohort remains the largest multi-year direct home-observation study to date (*N* = 156 mother–child dyads coded at age 4 and retested 4 to 7 years later). Maternal process praise during home teaching tasks had zero longitudinal association with elementary Stanford-Binet IQ (GEE standardized β = −0.01, SE = 0.08, *p* = .88) and a slight negative non-significant relation with BQAL literacy scores (*N* = 125, β = −0.12, SE = 0.09, *p* = .19).
- **Li, Zhang, Wu, and Zheng (2023):** Across 292 Chinese elementary families (mean age 9.33 years; *n* = 103 linked to official Grade 4 school exams), binary PDF inspection recovered minus signs stripped by an Elsevier null-byte font bug in Table 1. Parent-reported process praise showed null links with child growth mindset (*r* = −.02, *p* = .73), reverse-scored fixed mindset (*r* = +.03, *p* = .61), official mathematics exam scores (*r* = +.11, *p* = .27), and official Chinese language exam scores (*r* = +.12, *p* = .23).
- **Francis, Boullion, and Yates (2025):** Across 250 caregiver–child dyads observed at age 4 and followed to age 8 (*n* = 206 retained), caregiver process praise correlated *r* = +.01 (*ns*) with concurrent age-4 puzzle persistence and *r* = +.11 (*p* = .116, *ns*; ambiguous praise *r* = +.09, *ns*) with age-8 persistence. Table 2 (p. 49) transposed column headers *B* and β across every predictor, and its 95% bias-corrected bootstrap interval for age-4 process praise ([−0.079, +0.327]) spanned zero.
- **Gunderson Cohorts and Home Contingency (Gunderson et al., 2013, 2018a, 2018b; Radovanovic et al., 2023):** In the Chicago toddler cohort (*N* = 53 across 2013 and 2018b), early process-praise share correlated *r* = .34–.35 with age 7–8 motivational frameworks and *r* = .22, .26, and −.01 with Grade 4 math, passage comprehension, and word decoding (composite *r* = +.14, 95% CI [−.05, +.33]). Larger follow-ups complicated that early picture. Online Appendices B and C (PMC5986600) of Gunderson, Donnellan, et al. (2018a, Study 2; *N* = 267) revealed a zero bivariate correlation between child-perceived parental process praise and incremental intelligence beliefs (*r* = −.04, Table B.5) alongside a negative structural path once shared praise variance was partialed out (β = −0.22, *p* = .026). In Radovanovic et al. (2023, Experiment 2; *N* = 60 home Zoom dyads), total maternal process praise showed virtually zero association with infant persistence (*r* = +.02) because non-contingent 'random' process praise made up 39% of home utterances and depressed persistence (*t* = −2.35, *p* = .02).

Outcome stratification across Level 3 (Subgroup C; Table 2 and Figure 2, Panel B2) confirmed zero transfer to mindsets or school grades. Across all four cohorts administering child mindset scales (*N* = 722), pooled *r* = +0.009 (95% CI [−0.056, +0.075], *p* = .775, I^2 = 54.7%); across all three cohorts with standardized achievement or IQ (*N* = 310), pooled *r* = +0.038 (95% CI [−0.085, +0.159], *p* = .548, I^2 = 37.2%). General non-inflated or total parental praise (Level 3b, Subgroup B4: Brummelman et al., 2014a S2/2017; Gunderson, Donnellan, et al., 2018a S1; Lee et al., 2017; Carver et al., 2022; Li et al., 2023; *k* = 5, *N* = 1,104) was similarly near zero (*r* = +0.075, Knapp–Hartung 95% CI [−0.008, +0.156], I^2 = 0.0%), whereas Brummelman et al. (2017) documented longitudinal self-esteem erosion after inflated praise (β = −.11*) and Lee et al. (2017;* N = *298) found curvilinear praise-accuracy effects on middle-school GPA. What about within-family contrasts between process and person talk? Across 19 Steiger dependent tests* Var(z~SUB~Process~ENDSUB~ − z~SUB~Person~ENDSUB~) *in nine naturalistic cohorts (Table 5; sensitivity across* r~SUB~12~ENDSUB~ ∈ *{.20, .35, .50} in Table S5), just 4 of 19 reached* p < *.05 at empirical* r~SUB~12~ENDSUB~*, every one of those four reflected negative person-praise links (Pomerantz & Kempner, 2013 mindset and challenge proportions; Gunderson et al., 2018b Grade 4 math; Francis et al., 2025 age-8 persistence), and Gunderson et al.'s (2013) signature age 7–8 mindset contrast missed significance under paired testing (*p = *.062 at PMC XML* r~SUB~person~ENDSUB~ = −.05;* p = *.160 at printed* r^*_{person} = +.05).

Table 4

Mixed-Effects REML Meta-Regression Models Testing Temporal/Ecological Attenuation (MR-1), Comparator Precision (MR-2), and Outcome Realism/Age (MR-3).

Meta-Regression Model	k	Omnibus Moderator Test	R^2 (%)	Predictor Term	β	SE	95% CI	p
MR-1: Ecological & Temporal Attenuation (Levels 1–3 on common g)	35	Q_M(2) = 4.06, p = .132	9.1%	Intercept (Level 1: Immediate Lab)	+0.669	0.111	[+0.451, +0.886]	< .001
				Level 2 (Delayed Withdrawal vs Level 1)	-0.204	0.425	[-1.037, +0.629]	.632
				Level 3 (Naturalistic Everyday vs Level 1)	-0.397	0.198	[-0.785, -0.009]	.045
MR-2: Experimental Comparator Precision (Subgroup B)	41	Q_M(2) = 11.39, p = .003	23.7 %	Intercept (B3: Process vs Neutral/Ambiguous)	+0.168	0.116	[-0.059, +0.395]	.146
				B1 (Pure Process vs Person)	+0.475	0.167	[+0.147, +0.804]	.004
				B2 (Nongeneric vs Generic)	+0.636	0.238	[+0.170, +1.102]	.007
MR-3: Experimental Outcome Realism & Child Age Moderation	31	Q_M(2) = 3.26, p = .196	4.3%	Intercept (C1: Hypothetical/Verbal at Age 9)	+0.763	0.131	[+0.506, +1.020]	< .001
				C2 (Real Behavioral Task vs Hypothetical)	-0.258	0.212	[-0.672, +0.157]	.224
				Age_Centered_9yr (per year)	-0.038	0.033	[-0.102, +0.026]	.248

Note. Small-study funnel asymmetry in Level 1a (k = 23): Egger intercept β_0 = +3.48 (SE = 1.67, $t(21)$ = 2.08, two-tailed p = .050); PET intercept g = −0.165; PEESE precision-adjusted estimate g = +0.264.

Table 5

Within-Cohort Dependent-Correlation Contrasts Δz (Process Praise Minus Person Praise) Using Steiger's Exact Covariance Formula Var(A − B) = Var(A) + Var(B) − 2·Cov(A, B).

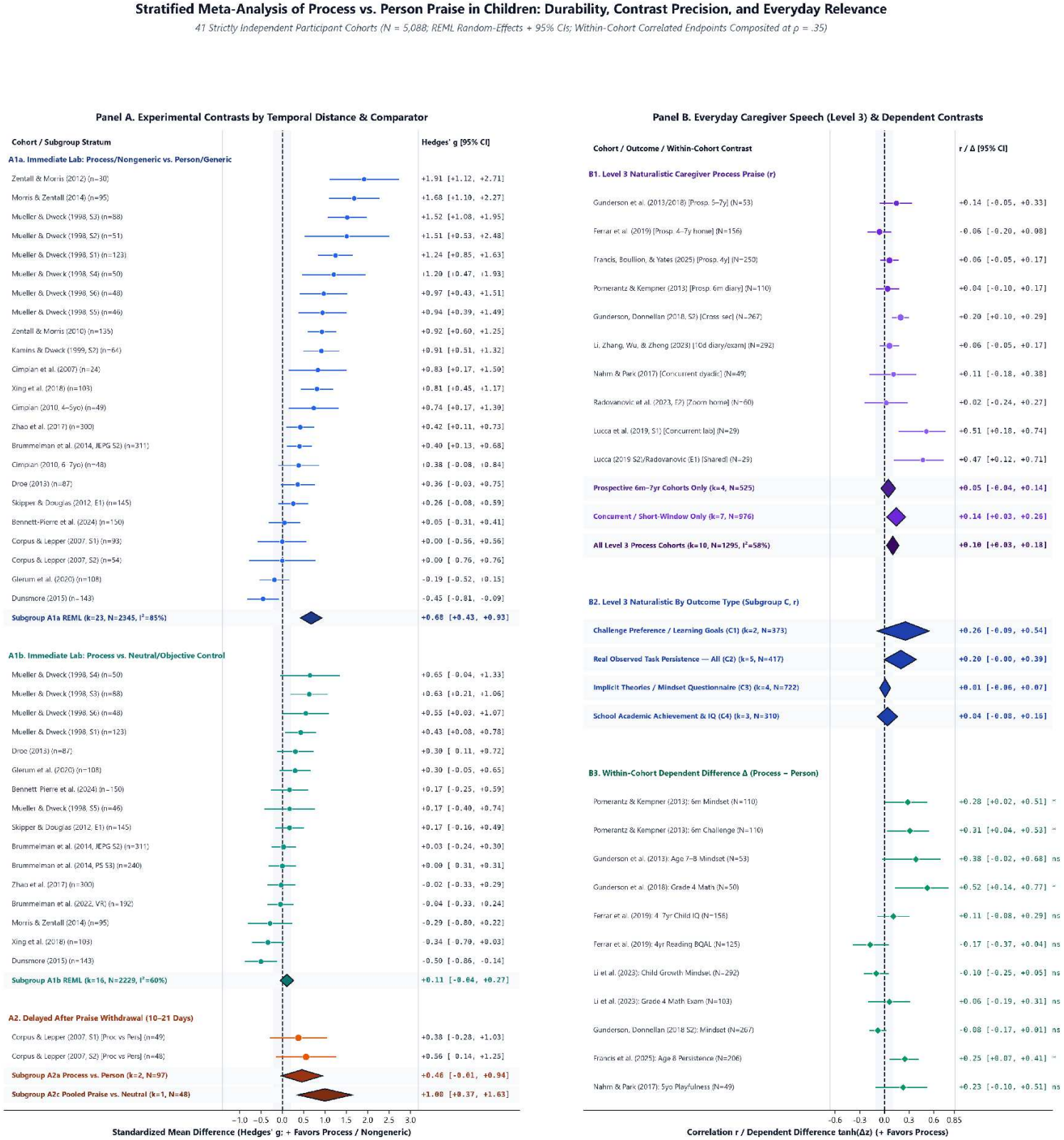
Naturalistic Cohort & Endpoint	N	r (Process)	r (Person)	Emp. r_12	Δz [95% CI] (Emp. r_12)	p (Emp.)	p (r_12 = .35)
Pomerantz & Kempner (2013) — 6m Mindset (Proportion, r12=-.05)	110	+0.082	-0.207	-0.05	+0.292 [+0.02, +0.57]	.038	.009
Pomerantz & Kempner (2013) — 6m Mindset (Frequency, r12=+.39)	110	-0.091	-0.217	+0.39	+0.129 [-0.08, +0.34]	.231	.245
Pomerantz & Kempner (2013) — 6m Challenge Pref (Proportion, r12=-.05)	110	+0.124	-0.191	-0.05	+0.318 [+0.04, +0.59]	.024	.004
Gunderson et al. (2013) — Age 7-8 Mindset (PMC XML r_person=-.05)	53	+0.340	-0.050	-0.14	+0.404 [-0.02, +0.83]	.062	.014
Gunderson et al. (2013) — Age 7-8 Mindset (2013 Text r_person=+.05)	53	+0.340	+0.050	-0.14	+0.304 [-0.12, +0.73]	.160	.065
Gunderson, Sorhagen, et al. (2018b) — Grade 4 Math	50	+0.330	-0.230	-0.14	+0.577 [+0.14, +1.02]	.010	.0007
Gunderson, Sorhagen, et al. (2018b) — Grade 4 Reading Comprehension	50	-0.020	-0.240	-0.14	+0.225 [-0.21, +0.66]	.310	.181
Ferrar et al. (2019) — 4-7yr Longitudinal IQ (GEE)	156	+0.039	-0.070	+0.25	+0.109 [-0.09, +0.30]	.271	.237
Ferrar et al. (2019) — 4yr Longitudinal Reading BQAL	125	-0.163	+0.011	+0.25	-0.175 [-0.39, +0.04]	.115	.091
Li, Zhang, Wu, & Zheng (2023) — Growth Mindset (r12=+.15)	292	-0.020	+0.080	+0.15	-0.100 [-0.25, +0.05]	.192	.136
Li, Zhang, Wu, & Zheng (2023) — Grade 4 Math Exam (r12=+.15)	103	+0.110	+0.050	+0.15	+0.060 [-0.20, +0.32]	.644	.597
Gunderson, Donnellan, et al. (2018a, S2) — Incremental Mindset (r12=+.69)	267	-0.040	+0.040	+0.69	-0.080 [-0.18, +0.01]	.099	.254
Gunderson, Donnellan, et al. (2018a, S2) — Learning Goals (r12=+.69)	263	+0.410	+0.360	+0.69	+0.059 [-0.04, +0.16]	.261	.430

Naturalistic Cohort & Endpoint	N	r (Process)	r (Person)	Emp. r_12	Δz [95% CI] (Emp. r_12)	p (Emp.)	p (r_12 = .35)
Francis, Boullion, & Yates (2025) — Age 8 Persistence (r12=+.11)	206	+0.110	-0.140	+0.11	+0.251 [+0.07, +0.44]	.008	.002
Nahm & Park (2017) — 5yo Playfulness Disposition	49	+0.110	-0.120	+0.35	+0.231 [-0.10, +0.56]	.171	.171

Note. Complete outputs across all 19 dependent contrasts and sensitivity correlations $r_{12} \in \{.20, .35, .50\}$ appear in Supplement S1, Table S5. Across every significant contrast ($p < .05$), *r* (Person) is negative while *r* (Process) alone does not differ from zero.

Figure 2

Stratified Subgroup Forest Plot Across Experimental Comparators/Time Windows (Panel A, Hedges' *g*) and Everyday Caregiver Speech Plus Within-Cohort Dependent Contrasts (Panel B, *r* / tanh(Δ*z*)).



Note. Squares denote cohort point estimates scaled by random-effects weight; horizontal bars mark 95% confidence intervals; diamonds show REML subgroup summaries across strictly independent cohorts (k = 41, N = 5,088).

3.4. Related Boundary Exposures and Forensic Audit of Primary-Source Discrepancies

Adjacent full-text studies clarify related social-contextual processes outside direct process-versus-person contrasts (Supplement S1, Table S3). In youth competitive swimming, perceived autonomy-supportive coach praise predicted one-week coach relatedness (Conroy & Coatsworth, 2007; $N = 165$) and competence/initiative (Coatsworth & Conroy, 2009; $N = 119$), though neither paper contrasted process against person wording or established sample independence. Liu and Yin (2026; $N = 146$ enrolled, $n = 132$ analyzed Taiwanese seventh-graders; Subgroup B4 only) tested personal affirmation against no praise under task- versus ego-involving expectations, where the one-week lag preceded praise delivery and no process arm existed. Ishibashi et al. (2026; $N = 58$ enrolled, $n = 49$ analyzed toddlers; Subgroup B4 only) observed longer same-session persistence during praised blocks from a human or social robot (within-subject $g = +0.43$) without attributional contrasts or delayed probes. In the Tokyo Teen Cohort, Nagaoka et al. (2022, plus 2023 Corrigendum) tied parental praise for prosocial acts rather than academic effort to lower adolescent depressive symptoms. Vignette experiments further showed that teachers and adults steer inflated and person-oriented praise toward pupils perceived as low in self-esteem or socioeconomic status, which signals low ability expectations to onlookers (Brummelman et al., 2014a, Study 1; Schoneveld & Brummelman, 2023).

Primary-source errors proved widespread. Across 74 verified texts, our forensic check logged 22 numerical, sign, denominator, or degrees-of-freedom discrepancies (Supplement S1, Table S4). Alongside inverted Cohen's *d* signs in Bennett-Pierre et al. (2024), dropped minus signs in Pomerantz and Kempner (2013, Table 3), swapped *B*/β headers and zero-crossing bootstrap CIs in Francis et al. (2025), SE columns mislabeled as SD in Xing et al. (2018), reversed prose grade trends against Online Appendix Table B.5 in Gunderson, Donnellan, et al. (2018a), and pretest reversal plus irreproducible paired *t*-tests in Zarrinabadi et al. (2023), we documented impossible chi-square probabilities in Zentall and Morris (2010; $\chi^2(12) = 2.3$ printed as $p < .001$), incompatible F–p–η_p^2 values in Glerum et al. (2020; $F(2, 105) = 2.27$ printed as $p = .0423$), contradicted enjoyment means in Droe (2013), unnoted cohort reuse in Brummelman et al. (2014a S2/2017) and Lucca et al. (2019 S2)/Radovanovic et al. (2023 E1), and null-byte minus-sign stripping in Li et al. (2023). Perturbing $\rho_w \in \{.20, .35, .50\}$ and classroom ICC $\in \{.00, .02, .05, .10\}$ left every substantive conclusion intact (Supplement S1, Table S6).

4. Discussion

4.1. Reframing the Literature: Person-Praise Vulnerability Rather Than Process-Praise Potency

Why do dramatic laboratory gaps coexist with flat longitudinal trajectories? Synthesizing 41 independent cohorts ($N = 5{,}088$) resolves that apparent paradox. When children hit an engineered setback moments after hearing fixed ability or generic labels ('you are so smart'; 'you are a good drawer'), they voice more helpless attributions, report sharper shame, and cheat more often to defend that newly awarded status than peers given effort or episode-specific remarks ($g = +0.68$ unadjusted; PEESE $g = +0.26$). Compare process praise against neutral feedback instead. Across 16 laboratory cohorts ($N = 2{,}229$), process praise matched neutral, objective, or non-attributional feedback on immediate persistence, performance, exploration, and honesty ($g = +0.11$, $p = .160$). Twelve three-arm decompositions confirm the asymmetry: ability and generic labels drag children below baseline during setbacks, whereas process praise holds them near baseline. Popular guidance mistook the absence of ability-praise harm for an active motivational boost.

4.2. Temporal Durability Versus Ongoing Socialization Exposure

Post-withdrawal durability (Level 2) and naturalistic family correlations (Level 3) answer separate causal questions. Neither rescues the strong process-praise thesis. At Level 2, nearly thirty years of experimentation have yielded just two small unconfounded post-withdrawal contrasts in a single paper (Corpus & Lepper, 2007; $N = 97$, $g = +0.46$, $p = .058$, *ns*), further weakened by asymmetric memory screening among preschoolers and gender-bound gift rankings among preadolescents. At Level 3, synthesis of full-text tables from Pomerantz and Kempner (2013), Ferrar et al. (2019), and Li et al. (2023) alongside Gunderson et al.

(2013, 2018b) and Francis et al. (2025) overturns claims built on one 53-family toddler sample. Across all four multi-wave prospective home and diary cohorts (N = 525), everyday caregiver process praise showed a null prospective correlation (r = +0.048, p = .289), alongside zero links to child growth mindset (r = +0.009, N = 722) and standardized achievement or IQ (r = +0.038, N = 310). Whenever within-family gaps between process and person praise reached significance (Table 5), Steiger covariance models traced them to negative correlates of parental person praise and person criticism.

4.3. Developmental Pragmatics, Contingency, and Backfire in Older Children

Why does everyday process praise fall flat and sometimes backfire? Three developmental mechanisms emerge across verified full texts. Contingency breaks down first: nearly two-fifths of home maternal process praise occurs randomly without prior effortful progress, and non-contingent process praise depresses infant persistence (Radovanovic et al., 2023, Experiment 2). Attribution pragmatics shift next. By middle childhood and adolescence, pupils infer inverse effort-ability rules whereby heavy praise for effort on routine tasks implies low underlying talent (Amemiya & Wang, 2018; Good & Shaw, 2022; Lam et al., 2008). Praising effort in vocational classrooms or helping tasks can therefore trigger low-ability inferences (Glerum et al., 2020) and curb voluntary prosocial persistence (Dunsmore, 2015). Hyperbole harms as well, because inflated or inaccurate praise erects evaluative bars that anxious children dread missing (Brummelman et al., 2014a, 2017; Lee et al., 2017).

4.4. Practical Implications and Design Standards for Future Durability Trials

Practical guidance for parents, teachers, and clinicians becomes simpler and more honest. Drop diagnostic trait labels ('you're a natural', 'you're so smart') and exaggerated flattery when children face frustration or self-doubt. Offer contingent, task-specific information instead. Adults should not expect routine effort slogans ('you worked so hard!') to implant growth mindsets, lift standardized test scores, or sustain voluntary persistence weeks later, and school leaders should retire classroom process-praise quotas extrapolated from five-minute puppet scripts.

What would establish genuine durability? Future randomized trials must meet five design standards rarely satisfied together: (1) preregister three-arm designs (Process vs. Person vs. Neutral informative feedback) equated on warmth, specificity, and contingency; (2) randomize individual learners or at least 20 classrooms per arm with explicit ICC modeling rather than one-class-per-arm layouts; (3) verify feedback cessation and observe voluntary behavioral engagement days and weeks later without asymmetric memory exclusions; (4) measure real task persistence alongside verbal preferences; and (5) publish exact cell denominators, pre-post correlations, and open raw data.

5. Limitations and Conclusion

Four limitations bound our conclusions. First, although all eight Priority-1 paywalled studies and 74 primary texts were audited first-hand, eight peripheral paywalled citations were unavailable in full layout; complete abstract and metadata checks verified that none tested an eligible unrepresented child process-versus-person contrast (Table S1). Second, unconfounded Level 2 post-withdrawal trials remain scarce (k = 2 cohorts), which restricts statistical power for small delayed differences. Third, Level 3 cohorts coded caregiver praise variously as raw counts, shares of total speech, shares of praise, or child perceptions; testing both frequency and proportion specifications (as in Pomerantz & Kempner, 2013) yielded matching nulls, yet observational cohorts cannot rule out evocative child-to-parent effects. Fourth, extraction, scripted verification, and RoB 2 / JBI ratings were executed by one methodologist with programmatic table checks rather than dual independent coders.

Distinguishing acute laboratory scripts, post-withdrawal follow-ups, and everyday family speech reshapes developmental theory on praise. Process and nongeneric phrasing spares children the immediate setback fragility provoked by person and generic trait labels, yet same-session process praise adds nothing above neutral informative feedback, post-withdrawal classroom durability remains unproven, and prospective

longitudinal cohorts show null links between parental process praise and later child mindsets or achievement. Both motivational theory and parenting advice should move past sweeping claims about process praise toward exact accounting of comparator harm, contingency, developmental pragmatics, and follow-up horizon.

Declarations

Supplementary Material: Online Supplement S1 documents the multi-database harvest and legacy register reconciliation, the full-text disposition of all 28 paywalled candidates (Table S1), the complete cohort and contextual study registers (Tables S2–S3), our 22-item forensic register of primary-source statistical discrepancies (Table S4), full Steiger dependent-contrast and sensitivity matrices (Tables S5–S6), and result-level RoB 2 and JBI appraisals.

Data and Code Availability: All 137 audited effect-size rows (T04_EFFECT_SIZES_AUDIT.csv), complete random-effects and meta-regression outputs (T04_META_RESULTS.json), and reproducible Python scripts (run_meta_analysis.py) are archived in the project repository.

Funding and Competing Interests: This research received no external grant from any public, commercial, or not-for-profit agency. The author declares no competing financial or non-financial interests.

Ethics Approval and Consent to Participate: Not applicable; this study is a PRISMA 2020 systematic review and meta-analysis of published aggregate study data and involved no direct recruitment of human participants.

Copyright Notice: © 2026 Alexander Kuznetsov. Personal reading, local archival download, and scholarly citation are permitted. Redistribution, commercial reuse, and circulation of adapted versions require written permission from the copyright holder except where applicable preprint or journal publishing licences or law permit otherwise.

References

References marked with an asterisk (*) indicate primary reports included in the quantitative meta-analysis.

Amemiya, J., & Wang, M.-T. (2018). Why effort praise can backfire in adolescence. *Child Development Perspectives, 12*(3), 199–203. https://doi.org/10.1111/cdep.12284

*Bennett-Pierre, G., Chernuta, T., Altamimi, R., & Gunderson, E. A. (2024). Effects of praise and “easy” feedback on children’s persistence and self-evaluations. *Journal of Experimental Child Psychology, 247*, 106032. https://doi.org/10.1016/j.jecp.2024.106032

Brummelman, E., Crocker, J., & Bushman, B. J. (2016). The praise paradox: When and why praise backfires in children with low self-esteem. *Child Development Perspectives, 10*(2), 111–115. https://doi.org/10.1111/cdep.12171

*Brummelman, E., Grapsas, S., & van der Kooij, K. (2022). Parental praise and children’s exploration: A virtual reality experiment. *Scientific Reports, 12*, 4967. https://doi.org/10.1038/s41598-022-08226-9

*Brummelman, E., Nelemans, S. A., Thomaes, S., & Orobio de Castro, B. (2017). When parents’ praise inflates, children’s self-esteem deflates. *Child Development, 88*(6), 1799–1809. https://doi.org/10.1111/cdev.12936

*Brummelman, E., Thomaes, S., Orobio de Castro, B., Overbeek, G., & Bushman, B. J. (2014a). “That’s not just beautiful—that’s incredibly beautiful!”: The adverse impact of inflated praise on children with low self-esteem. *Psychological Science, 25*(3), 728–735. https://doi.org/10.1177/0956797613514251

*Brummelman, E., Thomaes, S., Overbeek, G., Orobio de Castro, B., van den Hout, M. A., & Bushman, B. J. (2014b). On feeding those hungry for praise: Person praise backfires in children with low self-esteem. *Journal of Experimental Psychology: General, 143*(1), 9–14. https://doi.org/10.1037/a0031917

*Carver, C. E., Duong, S., Bachman, H. J., Votruba-Drzal, E., & Libertus, M. E. (2022). Examining relations between parental feedback types and preschool-aged children’s academic skills. *International Journal of Psychological Studies, 14*(4), 1–16. https://doi.org/10.5539/ijps.v14n4p1

*Cimpian, A. (2010). The impact of generic language about ability on children’s achievement motivation. *Developmental Psychology, 46*(5), 1333–1340. https://doi.org/10.1037/a0019665

*Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science, 18*(4), 314–316. https://doi.org/10.1111/j.1467-9280.2007.01896.x

Coatsworth, J. D., & Conroy, D. E. (2009). The effects of autonomy-supportive coaching, need satisfaction, and self-perceptions on initiative and identity in youth swimmers. *Developmental Psychology, 45*(2), 320–328. https://doi.org/10.1037/a0014027

Conroy, D. E., & Coatsworth, J. D. (2007). Assessing autonomy-supportive coaching strategies in youth sport. *Psychology of Sport and Exercise, 8*(5), 671–684. https://doi.org/10.1016/j.psychsport.2006.12.001

Corpus, J. H., & Good, K. A. (2021). The effects of praise on children's intrinsic motivation revisited. In E. Brummelman (Ed.), *Psychological perspectives on praise* (pp. 39–46). Routledge.

*Corpus, J. H., & Lepper, M. R. (2007). The effects of person versus performance praise on children's motivation: Gender and age as moderating factors. *Educational Psychology, 27*(4), 487–508. https://doi.org/10.1080/01443410601159852

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

*Droe, K. L. (2013). Effect of verbal praise on achievement goal orientation, motivation, and performance attribution. *Journal of Music Teacher Education, 23*(1), 63–78. https://doi.org/10.1177/1057083712458592

*Dunsmore, J. C. (2015). Effects of person- and process-focused feedback on prosocial behavior in middle childhood. *Social Development, 24*(1), 57–75. https://doi.org/10.1111/sode.12082

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Ferrar, S. J., Stack, D. M., Dickson, D. J., Serbin, L. A., Ledingham, J. E., & Schwartzman, A. E. (2019). Maternal socialization responses to preschoolers' success and struggle: Links to contextual factors and academic and cognitive outcomes. *Journal of Research in Childhood Education, 33*(3), 363–381. https://doi.org/10.1080/02568543.2019.1607787

*Francis, R. S., Boullion, A., & Yates, T. M. (2025). Evaluating the impact of parental praise on children's problem-solving persistence. *UC Riverside Undergraduate Research Journal, 19*(1), 45–53. https://doi.org/10.5070/RJ5.39984

*Glerum, J., Loyens, S. M. M., Wijnia, L., & Rikers, R. M. J. P. (2020). The effects of praise for effort versus praise for intelligence on vocational education students. *Educational Psychology, 40*(10), 1270–1286. https://doi.org/10.1080/01443410.2019.1625306

Good, K., & Shaw, A. (2022). Being versus appearing smart: Children's developing intuitions about how reputational motives guide behavior. *Child Development, 93*(2), 418–436. https://doi.org/10.1111/cdev.13711

*Gunderson, E. A., Donnellan, M. B., Robins, R. W., & Trzesniewski, K. H. (2018a). The specificity of parenting effects: Differential relations of parent praise and criticism to children's theories of intelligence and learning goals. *Journal of Experimental Child Psychology, 173*, 116–135. https://doi.org/10.1016/j.jecp.2018.03.015

*Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent praise to 1- to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development, 84*(5), 1526–1541. https://doi.org/10.1111/cdev.12064

*Gunderson, E. A., Sorhagen, N. S., Gripshover, S. J., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2018b). Parent praise to toddlers predicts fourth grade academic achievement via children's incremental mindsets. *Developmental Psychology, 54*(3), 397–409. https://doi.org/10.1037/dev0000444

Haimovitz, K., & Corpus, J. H. (2011). Effects of person versus process praise on student motivation: Stability and change in emerging adulthood. *Educational Psychology, 31*(5), 595–609. https://doi.org/10.1080/01443410.2011.585950

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin, 128*(5), 774–795. https://doi.org/10.1037/0033-2909.128.5.774

Higgins, J. P. T., Li, T., & Sterne, J. (Eds.). (2021). *Revised Cochrane risk of bias tool for randomized trials (RoB 2): Additional considerations for crossover trials*. https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-crossover-trials

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. C. (2019). Assessing risk of bias in a randomized trial. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), Cochrane handbook for systematic reviews of interventions (2nd ed., pp. 205–228). Wiley. https://doi.org/10.1002/9781119536604.ch8

*Ishibashi, M., Shinya, Y., Yoshikawa, Y., Ishiguro, H., & Itakura, S. (2026). Effects of praise from a social robot on task persistence in 18- to 24-month-old children. *Frontiers in Robotics and AI, 13*, 1782839. https://doi.org/10.3389/frobt.2026.1782839

JBI. (2024a). Appendix 4.3: JBI critical appraisal checklist for quasi-experimental studies. In *JBI manual for evidence synthesis*. https://jbi-global.atlassian.net/wiki/spaces/MANUAL/pages/355829117

JBI. (2024b). Appendix 7.1: Critical appraisal checklist for cohort studies. In *JBI manual for evidence synthesis*. https://jbi-global.atlassian.net/wiki/spaces/MANUAL/pages/355599324

*Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology, 35*(3), 835–847. https://doi.org/10.1037/0012-1649.35.3.835

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Lam, S.-f., Yim, P.-s., & Ng, Y.-l. (2008). Is effort praise motivational? The role of beliefs in the effort–ability relationship. *Contemporary Educational Psychology, 33*(4), 694–710. https://doi.org/10.1016/j.cedpsych.2008.01.005

*Lee, H. I., Kim, Y.-H., Kesebir, P., & Han, D. E. (2017). Understanding when parental praise leads to optimal child outcomes: Role of perceived praise accuracy. *Social Psychological and Personality Science, 8*(6), 679–688. https://doi.org/10.1177/1948550616683020

*Li, K., Zhang, J., Wu, Z., & Zheng, Y. (2023). Intelligence mindset in Chinese children: The role of parental praise and autonomy support. *Journal of Applied Developmental Psychology, 87*, 101555. https://doi.org/10.1016/j.appdev.2023.101555

*Liu, C.-H., & Yin, X.-R. (2026). Effects of person praise on students' motivation and performance under ego-involving and task-involving feedback conditions. *Instructional Science, 54*, 34. https://doi.org/10.1007/s11251-026-09789-0

*Lucca, K., Horton, R., & Sommerville, J. A. (2019). Keep trying!: Parental language predicts infants' persistence. *Cognition, 193*, 104025. https://doi.org/10.1016/j.cognition.2019.104025

*Morris, B. J., & Zentall, S. R. (2014). High fives motivate: The effects of gestural and ambiguous verbal praise on motivation. *Frontiers in Psychology, 5*, 928. https://doi.org/10.3389/fpsyg.2014.00928

*Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*(1), 33–52. https://doi.org/10.1037/0022-3514.75.1.33

Nagaoka, D., Tomoshige, N., Ando, S., Morita, M., Kiyono, T., Kanata, S., Fujikawa, S., Endo, K., Yamasaki, S., Fukuda, M., Nishida, A., Hiraiwa-Hasegawa, M., & Kasai, K. (2022). Being praised for prosocial behaviors longitudinally reduces depressive symptoms in early adolescents: A population-based cohort study. *Frontiers in Psychiatry, 13*, 865907. https://doi.org/10.3389/fpsyt.2022.865907

Nagaoka, D., Tomoshige, N., Ando, S., Morita, M., Kiyono, T., Kanata, S., Fujikawa, S., Endo, K., Yamasaki, S., Fukuda, M., Nishida, A., Hiraiwa-Hasegawa, M., & Kasai, K. (2023). Corrigendum: Being praised for prosocial behaviors longitudinally reduces depressive symptoms in early adolescents: A population-based cohort study. *Frontiers in Psychiatry, 14*, 1232335. https://doi.org/10.3389/fpsyt.2023.1232335

*Nahm, E. Y., & Park, S. E. (2017). The mediating effects of maternal praise types on the relationship between mother and child emotional reactions in task interaction. *Korean Journal of Child Studies, 38*(4), 79–89. https://doi.org/10.5723/kjcs.2017.38.4.79

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, Article n71*. https://doi.org/10.1136/bmj.n71

Pepper, S., Ren, K., Bartek, V., & Gunderson, E. A. (n.d.). *Mothers' and fathers' motivational talk to first-graders: Praise, enjoyment and high expectations* [Research poster]. Temple University. https://mgol.net/wp-content/uploads/2019/03/Priase-Enjoyment-high-expectations_-Sarah-Pepper.pdf

*Pomerantz, E. M., & Kempner, S. G. (2013). Mothers' daily person and process praise: Implications for children's theory of intelligence and motivation. *Developmental Psychology, 49*(11), 2040–2046. https://doi.org/10.1037/a0031840

*Radovanovic, M., Soldovieri, A., & Sommerville, J. A. (2023). It takes two: Process praise linking trying and success is associated with greater infant persistence. *Developmental Psychology, 59*(9), 1668–1675. https://doi.org/10.1037/dev0001584

Schoneveld, E., & Brummelman, E. (2023). "You did incredibly well!": Teachers' inflated praise can make children from low-SES backgrounds seem less smart (but more hardworking). *npj Science of Learning, 8*(1), 31. https://doi.org/10.1038/s41539-023-00183-w

*Skipper, Y., & Douglas, K. (2012). Is no praise good praise? Effects of positive feedback on children's and university students' responses to subsequent failures. *British Journal of Educational Psychology, 82*(2), 327–339. https://doi.org/10.1111/j.2044-8279.2011.02028.x

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

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Stevens, C., Sidener, T. M., Reeve, S. A., & Sidener, D. W. (2011). Effects of behavior-specific and general praise, on acquisition of tacts in children with pervasive developmental disorders. *Research in Autism Spectrum Disorders, 5*(1), 666–669. https://doi.org/10.1016/j.rasd.2010.08.003

Swenson, S., Ho, G. W. K., Budhathoki, C., Belcher, H. M. E., Tucker, S., Miller, K., & Gross, D. (2016). Parents' use of praise and criticism in a sample of young children seeking mental health services. *Journal of Pediatric Health Care, 30*(1), 49–56. https://doi.org/10.1016/j.pedhc.2015.09.010

Verkoeijen, P. P. J. L., Polak, M. G., & Bouwmeester, S. (2018). A practical illustration of methods to deal with potential outliers: A multiverse outlier analysis of Study 3 from Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014). *Collabra: Psychology, 4*(1), 37. https://doi.org/10.1525/collabra.118

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30*(3), 261–293. https://doi.org/10.3102/10769986030003261

*Xing, S., Gao, X., Jiang, Y., Archer, M., & Liu, X. (2018). Effects of ability and effort praise on children's failure attribution, self-handicapping, and performance. *Frontiers in Psychology, 9*, 1883. https://doi.org/10.3389/fpsyg.2018.01883

Zarrinabadi, N., Lou, N. M., & Darvishnezhad, Z. (2023). To praise or not to praise? Examining the effects of ability vs. effort praise on speaking anxiety and willingness to communicate in EFL classrooms. *Innovation in Language Learning and Teaching, 17*(1), 88–101. https://doi.org/10.1080/17501229.2021.1938079

*Zentall, S. R., & Morris, B. J. (2010). "Good job, you're so smart": The effects of inconsistency of praise type on young children's motivation. *Journal of Experimental Child Psychology, 107*(2), 155–163. https://doi.org/10.1016/j.jecp.2010.04.015

*Zentall, S. R., & Morris, B. J. (2012). A critical eye: Praise directed toward traits increases children's eye fixations on errors and decreases motivation. *Psychonomic Bulletin & Review, 19*(6), 1073–1077. https://doi.org/10.3758/s13423-012-0294-y

*Zhao, L., Heyman, G. D., Chen, L., & Lee, K. (2017). Praising young children for being smart promotes cheating. *Psychological Science, 28*(12), 1868–1870. https://doi.org/10.1177/0956797617721529

Zhao, L., Li, Y., Qin, W., Amemiya, J., Fang, F., Compton, B. J., & Heyman, G. D. (2022). Overheard evaluative comments: Implications for beliefs about effort and ability. *Child Development, 93*(6), 1889–1902. https://doi.org/10.1111/cdev.13829

Supplementary Files

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

- CleanednewT04S1Supplement.pdf