Online Supplement S1: Search Reconciliation, Study Family Register, Primary-Source Forensic Audit, and Sensitivity Analyses

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

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

S1. Legacy Bibliographic Register Reconciliation and Multi-API Harvest

Two fully audited search streams anchored our systematic review across 1998–2026. Stream 1 audited 36 legacy export files holding 3,260 bibliographic entries across 1,870 unique APA identifiers (1,802 in expanded RIS batches and 68 in disjoint print batches), which included 20 anchor papers and 789 topic candidates slated for primary-text adjudication. Stream 2 ran programmatic API queries plus backward and forward citation snowballing from 19 seed articles across OpenAlex, Crossref, Semantic Scholar, ERIC, Europe PMC/PubMed, and arXiv through 31 March 2026 to recover 2,495 unique deduplicated records that absorbed the archival register and isolated 328 praise papers with 138 empirical child or adolescent candidates. Acquisition of open-access PDFs, PubMed Central XMLs, author manuscripts, and publisher supplements assembled 74 verified documents. All 28 paywalled candidates are accounted for. Table S1 documents every item: 20 were acquired and audited in full (including all 8/8 [100%] Priority-1 critical studies), while 8 were confirmed as non-blocking boundary citations.

Table S1

Complete Disposition of All 28 Paywalled Candidate Articles Identified During Systematic Screening.

ID	Priority	Authors (Year)	Journal	DOI	Final Status	Verification Basis / Boundary Rationale
S1.01	P1	Pomerantz, E. M., & Kempner, S. G. (2013)	Developmental Psychology	10.1037/a0031840	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.02	P1	Kamins, M. L., & Dweck, C. S. (1999)	Developmental Psychology	10.1037/0012-1649.35.3.835	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.03	P1	Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007)	Psychological Science	10.1111/j.1467-9280.2007.01896.x	Obtained & Integrated (Full Text)	Verified publisher full-text PDF and online supplement
S1.04	P1	Cimpian, A. (2010)	Developmental Psychology	10.1037/a0019665	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.05	P1	Stevens, C., Sidener, T. M., Reeve, S. A., & Sidener, D. W. (2011)	Research in Autism Spectrum Disorders	10.1016/j.rasd.2010.08.003	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.06	P1	Bennett-Pierre, G., Chernuta, T., Altamimi, R., & Gunderson, E. A. (2024)	Journal of Experimental Child Psychology	10.1016/j.jecp.2024.106032	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.07	P1	Liu, C.-H., & Yin, X.-R. (2026)	Instructional Science	10.1007/s11251-026-09789-0	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.08	P1	Li, K., Zhang, J., Wu, Z., & Zheng, Y. (2023)	Journal of Applied Developmental Psychology	10.1016/j.appdev.2023.101555	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.09	P2	Koestner, R., Zuckerman, M., & Koestner, J. (1989)	Personality and Social Psychology Bulletin	10.1177/0146167289151006	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary

ID	Priority	Authors (Year)	Journal	DOI	Final Status	Verification Basis / Boundary Rationale
S1.10	P2	Lam, S.-f., Yim, P.-s., & Ng, Y.-l. (2008)	Contemporary Educational Psychology	10.1016/j.cedpsych.2008.01.005	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.11	P2	Haimovitz, K., & Corpus, J. H. (2011)	Educational Psychology	10.1080/01443410.2011.585950	Non-Blocking Boundary (Not Required)	Covered by institutional report (undergraduate sample outside K–12 scope)
S1.12	P2	Brummelman, E., Thomaes, S., Orobio de Castro, B., Overbeek, G., & Bushman, B. J. (2014)	Psychological Science	10.1177/0956797613514251	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.13	P2	Dunsmore, J. C. (2015)	Social Development	10.1111/sode.12082	Obtained & Integrated (Full Text)	Verified PubMed Central full-text XML
S1.14	P2	Lee, H. I., Kim, Y.-H., Kesebir, P., & Han, D. E. (2017)	Social Psychological and Personality Science	10.1177/1948550616683020	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.15	P2	Brummelman, E., Nelemans, S. A., Thomaes, S., & Orobio de Castro, B. (2017)	Child Development	10.1111/cdev.12936	Obtained & Integrated (Full Text)	Verified publisher full-text PDF and online supplement
S1.16	P2	Seo, Y. J., Cheah, C. S. L., & Hart, C. H. (2017)	Parenting: Science and Practice	10.1080/15295192.2017.1304786	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary
S1.17	P2	Zhao, L., Heyman, G. D., Chen, L., & Lee, K. (2017)	Psychological Science	10.1177/0956797617721529	Obtained & Integrated (Full Text)	Verified publisher full-text PDF and online supplement
S1.18	P2	Ferrar, S. J., Stack, D. M., Dickson, D. J., Serbin, L. A., Ledingham, J. E., & Schwartzman, A. E. (2019)	Journal of Research in Childhood Education	10.1080/02568543.2019.1607787	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.19	P2	Glerum, J., Loyens, S. M. M., Wijnia, L., & Rikers, R. M. J. P. (2020)	Educational Psychology	10.1080/01443410.2019.1625306	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.20	P2	Zarrinabadi, N., Lou, N. M., & Darvishnezhad, Z. (2023)	Innovation in Language Learning and Teaching	10.1080/17501229.2021.1938079	Obtained; Isolated in DiD+DEFF Sensitivity	Verified publisher full-text PDF
S1.21	P2	Zhao, L., Li, Y., Qin, W., Amemiya, J., Fang, F., Compton, B. J., & Heyman, G. D. (2022)	Child Development	10.1111/cdev.13829	Obtained & Integrated (Full Text)	Verified publisher full-text PDF and online supplement
S1.22	P2	Chui, W., Andrus, B. M., Uttal, D., & Horn, M. (2026)	International Journal of Child-Computer Interaction	10.1016/j.ijcci.2026.100820	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary
S1.23	P3	Droe, K. L. (2013)	Journal of Music Teacher Education	10.1177/1057083712458592	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.24	P3	Amemiya, J. L., & Wang, M.-T. (2018)	Child Development Perspectives	10.1111/cdep.12284	Obtained & Integrated (Full Text)	Verified publisher full-text PDF
S1.25	P3	Zhao, L., Chen, L., Sun, W., Compton, B. J., Lee, K., & Heyman, G. D. (2019)	Developmental Science	10.1111/desc.12930	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary
S1.26	P3	Schneider, M. M., Hulac, D. M., Mickelson, L. R., & Phillips, E. K. (2020)	Psychology in the Schools	10.1002/pits.22411	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary
S1.27	P3	Graf-König, N., & Puca, R. M. (2024)	Educational Psychology	10.1080/01443410.2024.2396422	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary
S1.28	P3	Xing, H., Yao, M., & Liu, H. (2024)	Journal of Youth and Adolescence	10.1007/s10964-024-02125-0	Non-Blocking Boundary (Not Required)	Excluded on scope/population boundary

Note. Every Priority-1 critical paper (8/8, 100%; S1.01–S1.08) was acquired in full text and checked directly against primary tables. Why were the remaining 8 citations non-blocking? Items S1.09, 11, 16, 22, and 25–28 either predated the 1998–2026 window (Koestner et al., 1989), sampled university undergraduates already captured in an institutional technical report (Haimovitz & Corpus, 2011), coded adult teacher speech or child loudness ratings without motivational endpoints, or lacked a process-versus-person contrast.

S2. Complete Register of Independent Quantitative Cohorts and Contextual Primary Sources

Cohorts were never double-counted across papers. Unit-of-analysis linkage mapped 42 primary quantitative study units across 37 study families onto $k = 41$ strictly independent participant cohorts ($N = 5{,}088$; Table S2), while qualitative boundary studies lacking child process-versus-person contrasts appear in Table S3.

Table S2

Complete Study-Family and Independent-Cohort Linkage Register Across All 42 Primary Quantitative Study Units.

Study Family ID	Independent Cohort ID	Linked Primary Report(s)	Design	N	Mean Age	Max Lag	Axes (A | B | C)
FAM_MD98_S1	COH_MD98_S1	Mueller & Dweck (1998, Study 1)	experimental between	123	10.7 yr	0 d	A1 | B1,B3 | C1,C2
FAM_MD98_S2	COH_MD98_S2	Mueller & Dweck (1998, Study 2)	experimental binary 2x2	51	10.7 yr	0 d	A1 | B1 | C1
FAM_MD98_S3	COH_MD98_S3	Mueller & Dweck (1998, Study 3)	experimental between	88	10.7 yr	0 d	A1 | B1,B3 | C1,C2
FAM_MD98_S4	COH_MD98_S4	Mueller & Dweck (1998, Study 4)	experimental between	50	10.7 yr	0 d	A1 | B1,B3 | C3
FAM_MD98_S5	COH_MD98_S5	Mueller & Dweck (1998, Study 5)	experimental between	46	10.7 yr	0 d	A1 | B1,B3 | C1,C2
FAM_MD98_S6	COH_MD98_S6	Mueller & Dweck (1998, Study 6)	experimental between	48	10.7 yr	0 d	A1 | B1,B3 | C1,C2,C3
FAM_KD99_S2	COH_KD99_S2	Kamins & Dweck (1999, Study 2)	experimental between	64	5.5 yr	0 d	A1 | B1 | C1,C3
FAM_CIMP07	COH_CIMP07	Cimpian, Arce, Markman, & Dweck (2007)	experimental contrast t	24	4.4 yr	0 d	A1 | B2 | C1
FAM_CIMP10	COH_CIMP10_45YO	Cimpian (2010 — 4-5-year-olds)	experimental contrast t	49	5.1 yr	0 d	A1 | B2 | C1
FAM_CIMP10	COH_CIMP10_67YO	Cimpian (2010 — 6-7-year-olds)	experimental contrast t	48	6.9 yr	0 d	A1 | B2 | C1
FAM_ZM10	COH_ZM10	Zentall & Morris (2010)	experimental contrast t	135	5.7 yr	0 d	A1 | B2 | C1
FAM_ZM12	COH_ZM12	Zentall & Morris (2012)	experimental between	30	5.5 yr	0 d	A1 | B2 | C1
FAM_MZ14	COH_MZ14	Morris & Zentall (2014)	experimental between	95	5.7 yr	0 d	A1 | B1,B3 | C1
FAM_SD12	COH_SD12_E1	Skipper & Douglas (2012, Exp. 1)	experimental contrast t	145	10.0 yr	0 d	A1 | B1,B3 | C1
FAM_DROE13	COH_DROE13	Droe (2013)	experimental binary 2x2	87	9.5 yr	0 d	A1 | B1,B3 | C1,C2
FAM_BRUM14	COH_BRUM14_S2	Brummelman et al. (2014b, JEP:G Study 2)	experimental contrast t	311	10.6 yr	0 d	A1 | B1,B3 | C1
FAM_BRUM14_PS	COH_BRUM14_PS_S3	Brummelman et al. (2014a, Psych Sci Study 3)	experimental contrast t	240	9.9 yr	0 d	A1 | B3 | C2
FAM_DUNS15	COH_DUNS15	Dunsmore (2015)	experimental contrast t	143	10.2 yr	0 d	A1 | B1,B3 | C2

Study Family ID	Independent Cohort ID	Linked Primary Report(s)	Design	N	Mean Age	Max Lag	Axes (A | B | C)
FAM_ZHAO17	COH_ZHAO17	Zhao, Heyman, Chen, & Lee (2017)	experimental binary 2x2	300	4.5 yr	0 d	A1 | B2,B3 | C2
FAM_XING18	COH_XING18	Xing et al. (2018)	experimental between	103	10.8 yr	0 d	A1 | B1,B3 | C1,C2
FAM_GLERUM20	COH_GLERUM20	Glerum, Loyens, Wijnia, & Rikers (2020)	experimental binary 2x2	108	17.2 yr	0 d	A1 | B1,B3 | C1,C2,C3
FAM_BRUM22	COH_BRUM22_VR	Brummelman, Grapsas, & van der Kooij (2022)	experimental contrast t	192	9.9 yr	0 d	A1 | B3 | C2
FAM_BP24	COH_BP24	Bennett-Pierre, Chernuta, Altamimi, & Gunderson (2024)	experimental between	150	8.0 yr	0 d	A1 | B1,B3 | C1,C2
FAM_CL07_S1	COH_CL07_S1	Corpus & Lepper (2007, Study 1); Corpus & Lepper (2007, Study 1 — Girls); Corpus & Lepper (2007, Study 1 — Boys)	experimental contrast t	93	10.3 yr	21 d	A1,A2 | B1,B3 | C1,C2
FAM_CL07_S2	COH_CL07_S2	Corpus & Lepper (2007, Study 2)	experimental contrast t	54	4.8 yr	10 d	A1,A2 | B1,B3,B4 | C2
FAM_ISHI26	COH_ISHI26	Ishibashi et al. (2026)	experimental contrast t	58	1.8 yr	0 d	A1 | B4 | C2
FAM_LIUYIN26	COH_LIUYIN26	Liu & Yin (2026)	experimental between	146	12.9 yr	0 d	A1 | B4 | C2
FAM_GUNDERSON_TODDLER	COH_CHI_TODDLER_53	Gunderson et al. (2013); Gunderson, Sorhagen, et al. (2018b)	naturalistic observational	53	7.5 yr	2555 d	A3 | B1 | C3,C4
FAM_POMERANTZ13	COH_POM13_110	Pomerantz & Kempner (2013)	naturalistic observational	110	10.2 yr	180 d	A3 | B1 | C1,C3
FAM_CONCORDIA	COH_FERRAR19_156	Ferrar, Stack, Dickson, Serbin, Ledingham, & Schwartzman (2019)	naturalistic observational	156	9.3 yr	2070 d	A3 | B1 | C4
FAM_FRANCIS25	COH_FRANCIS25_250	Francis, Boullion, & Yates (2025)	naturalistic observational	250	8.0 yr	1460 d	A3 | B1,B3 | C2
FAM_GUNDERSON_PERCEIVED	COH_CA_ELEM_S1	Gunderson, Donnellan, et al. (2018a, Study 1)	naturalistic observational	307	10.1 yr	0 d	A3 | B4 | C3
FAM_GUNDERSON_PERCEIVED	COH_CA_ELEM_S2	Gunderson, Donnellan, et al. (2018a, Study 2)	naturalistic observational	267	11.4 yr	0 d	A3 | B1 | C1,C3
FAM_LI23	COH_LI23_292	Li, Zhang, Wu, & Zheng (2023)	naturalistic observational	292	9.3 yr	0 d	A3 | B1,B4 | C3,C4
FAM_SOMMERVILLE_INFANT	COH_UW_INFANT_S1_29	Lucca, Horton, & Sommerville (2019, Study 1)	naturalistic observational	29	1.5 yr	0 d	A3 | B1 | C2
FAM_SOMMERVILLE_INFANT	COH_UW_INFANT_S2_29	Lucca et al. (2019, Study 2) & Radovanovic et al. (2023, Exp. 1)	naturalistic observational	29	1.5 yr	0 d	A3 | B1 | C2
FAM_SOMMERVILLE_INFANT	COH_TORONTO_ZOOM_60	Radovanovic, Soldovieri, & Sommerville (2023, Exp. 2)	naturalistic observational	60	2.0 yr	0 d	A3 | B1 | C2
FAM_NAHM17	COH_NAHM17_KR_49	Nahm & Park (2017)	naturalistic observational	49	5.0 yr	0 d	A3 | B1 | C2
FAM_BRUMMELMAN_DUTCH_HOME	COH_NL_HOME_116	Brummelman et al. (2014a, Study 2) & Brummelman, Nelemans, Thomaes, & Orobio de Castro (2017)	naturalistic observational	116	8.8 yr	548 d	A3 | B4 | C1

Study Family ID	Independent Cohort ID	Linked Primary Report(s)	Design	N	Mean Age	Max Lag	Axes (A | B | C)
FAM_LEE17	COH_LEE17_298	Lee, Kim, Kesebir, & Han (2017)	naturalistic observational	298	9.5 yr	0 d	A3 | B4 | C4
FAM_CARVER22	COH_CARVER22_91	Carver et al. (2022)	naturalistic observational	91	4.5 yr	365 d	A3 | B4 | C4

Note. Four multi-paper families prevented participant double-counting: FAM_GUNDERSON_TODDLER (Gunderson et al., 2013 + 2018b), FAM_SOMMERVILLE_INFANT (Lucca et al., 2019 S2 + Radovanovic et al., 2023 E1), FAM_GUNDERSON_PERCEIVED (Gunderson, Donnellan et al., 2018a S1 & S2; 58 shared pupils placed in disjoint strata and never pooled together), and FAM_BRUMMELMAN_DUTCH_HOME (Brummelman et al., 2014a S2 + 2017).

Table S3

Contextual, Boundary, and Prior Review Sources Examined in Full Text and Synthesized Qualitatively.

Source	Verified Document Version	Design Scope & Reason Not Pooled Into Primary Child Process-vs.-Person Contrasts
Stevens et al. (2011)	Verified Elsevier PDF + Figs. 1–2	Single-case alternating tact training (N = 2 boys with ASD, ages 15;8 & 6;6) under token economy; ≤ 3 sessions acquisition difference and virtually identical 1–6 week skill retention across specific praise, general praise, and tokens alone.
Lam, Yim, & Ng (2008)	Verified Elsevier PDF (Studies 1–2)	Hong Kong 6th–7th graders (N = 28 & N = 132) all received effort praise; priming inverse effort–ability beliefs lowered self-evaluation (d = −0.36) and intrinsic motivation (d = −0.47).
Zhao et al. (2022)	Verified Child Dev PDF + App. S1	Four experiments (N = 330, ages 4–5): overhearing a peer receive effort praise never increased children's effort valuation above baseline control (Study 1: 40.0% vs. 63.3% baseline, p = .073).
Zarrinabadi et al. (2023)	Verified ILLT PDF	Cluster-confounded EFL classroom study (N = 63, 1 intact class of 21/arm, ongoing 6-week praise); pretest competence imbalance (+93.76 pts) reverses true DiD (−7.09); isolated in Table S6 sensitivity.
Liu & Yin (2026)	Verified Springer PDF	Taiwanese 7th graders (N = 146 enrolled, n = 132 analyzed): personal affirmation vs. no praise under task/ego contexts (included solely in Subgroup B4 benchmark; 1-week lag preceded praise delivery; no process-praise arm).
Conroy & Coatsworth (2007); Coatsworth & Conroy (2009)	Verified journal PDFs	Youth summer-league swimmers (N = 165 & N = 119; potential participant overlap unverified): perceived autonomy-supportive coaching praise linked to coach relatedness and competence; no process-vs.-person contrast.
Nagaoka et al. (2022; 2023 Corrigendum)	Verified Frontiers PDFs	Tokyo Teen Cohort: parental praise for prosocial behaviors longitudinally associated with lower depressive symptoms; non-academic prosocial praise domain.
Swenson et al. (2016)	Verified PMC4685017 XML	Baseline parent report vs. brief play DPICS coding pooling labeled and unlabeled praise in clinical preschool sample; no child process-vs.-person outcome.
Schoneveld & Brummelman (2023)	Verified npj Sci. Learn. PDF	Experimental studies of teachers' inflated praise toward low-SES pupils and children's third-party ability/effort inferences.
Amemiya & Wang (2018); Brummelman et al. (2016); Corpus & Good (2021); Henderlong & Lepper (2002)	Verified PDFs / author MS	Theoretical and narrative review benchmarks on conditional praise effects, self-esteem paradoxes, and adolescent pragmatic inference.

Source	Verified Document Version	Design Scope & Reason Not Pooled Into Primary Child Process-vs.-Person Contrasts
Pepper et al. (n.d.)	Research poster PDF	Descriptive structured parent–child motivational talk in first-graders without a child motivational or behavioral outcome.

S3. Complete Forensic Register of Primary-Source Discrepancies (22 Audited Items)

Primary tables do not always agree with article prose. Whenever a report printed conflicting sample sizes, flipped effect signs, swapped regression headers, or incompatible test statistics and *p*-values across text, tables, and online appendices, we logged the exact anomaly in Table S4 rather than picking a convenient number silently.

Table S4

Forensic Register of Twenty-Two Statistical, Sign, Denominator, and Methodological Discrepancies Verified Across Primary Sources.

#	Source	Locator	Printed Discrepancy / Anomaly	Verification & Analytic Resolution
1	Pomerantz & Kempner (2013)	Table 3, p. 2044	In the Child preference for challenge block, maternal person praise has negative standardized coefficients (β = −.21* for frequency; β = −.17* for proportion) alongside positive unstandardized coefficients (B = 0.79, SE = 0.33; B = 1.64, SE = 0.83).	Minus signs omitted on printed B estimates; exact t-ratios (−0.79/0.33 = −2.39; −1.64/0.83 = −1.98) confirm negative slopes matching β and Table 2 zero-order correlations (r = −.18 to −.24).
2	Bennett-Pierre et al. (2024)	Results, p. 9 & Table 5	Authors printed positive d = .25 where post-failure self-evaluation was lower after process praise (3.78) than person praise (4.07), and negative d = −.22 where persistence was higher after process praise (0.64 vs. 0.59); baseline pre-praise puzzle time differed significantly across conditions, F(4, 145) = 3.93, p = .005; Table 5 prints '.1.38 (.62)'.	Authors computed (Person − Process) instead of (Process − Person); all effect sizes were recomputed directly from cell Ms and SDs under consistent (+ = favors Process) polarity.
3	Zarrinabadi, Lou, & Darvishnezhad (2023)	Tables 1–3 & p. 8	(a) Table 1 header states N = 44 and 8 interviews vs. text N = 63 and 12 interviews; (b) baseline perceived competence was +93.76 points higher in the single intact Effort class (430.80) than Ability class (337.04), reversing true Pre–Post DiD (+20.15 vs. +27.24, Δ = −7.09); (c) Table 3 paired t = 2.86, p = .010 is mathematically irreproducible from printed M1, SD1, M2, SD2, and r = .61 (exact paired t(20) = 1.22, p = .235, ns); (d) Table 2 prints F = 6.61, ηp² = .196 vs. text F = 6.68, ηp² = .192; (e) written 'effort' notes labeled students with trait adjective 'you are hardworking'.	Excluded from Level 2 post-withdrawal primary pool due to 1-cluster-per-arm teacher/class confounding and ongoing 6-week exposure; evaluated in dedicated Pre–Post DiD + DEFF sensitivity analyses.
4	Glerum et al. (2020)	Results, p. 1279	Low-ability failure attribution omnibus test printed as F(2, 105) = 2.27, p = .0423, ηp² = 0.059.	Mathematically incompatible triad: F(2, 105) = 2.27 yields exact p = .1084 and ηp² = .041, whereas ηp² = .059 corresponds to F = 3.29, p = .041. Pairwise Hedges' g computed directly from Table 2 Ms and SDs.

#	Source	Locator	Printed Discrepancy / Anomaly	Verification & Analytic Resolution
5	Droe (2013)	Results, pp. 72, 74 & Table 1	Page 72 states learning-goal choosers had significantly higher enjoyment (M = 3.34) than performance-goal choosers (M = 3.35), with both means lower than every condition mean in Table 1 (3.61–3.72); p. 74 claims Effort had highest post-failure enjoyment despite Table 1 showing Talent 3.72 > Effort 3.63.	Effect sizes extracted strictly from Table 1 condition Ms, SDs, and cell ns; prose contradiction documented.
6	Dunsmore (2015)	Table 1 & p. 65 (PMC4326064)	Gender-weighted mean helping time across girls (n = 70, M = 486.8 s) and boys (n = 73, M = 370.7 s) equals 427.53 s vs. printed overall M = 425.80 s; condition cell ns after excluding 16 children not reported.	Evaluated balanced allocation across N = 143 with exact Welch sensitivity; negative direction (process backfire) invariant.
7	Kamins & Dweck (1999, Study 2)	Results, p. 842	Pre-setback affect printed as Ms = 4.82, 4.83, and 3.95, F(2, 61) = 0.21, ns (3.95 duplicated from self-assessment row; F would exceed 5.0 if true); pairwise test printed as t(21) = −1.78 for n = 12 vs. n = 19 (true df = 29).	Post-setback endpoints extracted from verified Table/Figure values with exact group sizes (n = 19, 33, 12).
8	Cimpian et al. (2007) & Cimpian (2010)	2007 p. 315; 2010 p. 1336	In 2007, post-mistake self-evaluation t(22) = 1.94 yields two-tailed p = .065 (ns), obscured via prep = .858; in 2010, pre-mistake 4–5-year-old ANCOVA F(1, 41) = 0.76 is printed with p = .785 (exact p = .3884).	Exact two-tailed p-values and Hedges' g computed from test statistics and degrees of freedom.
9	Zhao, Heyman, Chen, & Lee (2017)	Main text p. 1869 vs. Supplement p. 6	Baseline vs. Ability cheating contrast reported as Wald $\chi^2 = 7.93$, p = .005 in main text and Table S2, but p = .007 in Supplement prose.	Log-odds ratios converted to exact Hedges' g directly from raw cell cheating counts (40/100 baseline, 41/100 process, and 60/100 ability).
10	Li, Zhang, Wu, & Zheng (2023)	Table 1 & p. 6	Publisher PDF encodes minus glyph in font /JRLWZW+CharisSIL as null byte \x00 (causing plain-text copy-paste to strip all negative signs); Model 2 SEM prints TLI = 1.06 (> 1.00) and compresses four praise-to-achievement paths into two numbers (Bs = −0.06 and 0.08, ns).	Binary glyph inspection recovered all exact negative signs in Table 1; verified zero-order correlations extracted for N = 292 and n = 103.
11	Ferrar et al. (2019)	Tables 4 & 6	Table 4 prints positive β = 0.05 alongside negative t = −0.588 for Parental distress → Maternal discouragement; Table 6 prints '−.1.36'.	Longitudinal child IQ (N = 156) and reading BQAL (N = 125) GEE paths verified across Tables 4–6.
12	Lee, Kim, Kesebir, & Han (2017)	Tables 1, 4, 5 & pp. 682–684	Table 5 prints SE = 0.36 and 0.27 alongside B = 0.93 (t = 25.85) and B = 0.98 (t = 36.38) for Previous GPA (decimal-place error for 0.036 and 0.027); prose prints SD = 0.87 vs. Table 1 SD = 0.89.	Zero-order maternal praise × school GPA correlation extracted directly from Table 2 (N = 298).
13	Brummelman et al. (2014a, S2; 2017)	2014a p. 731; 2017 p. 1801, 1806	Unflagged reuse of the same 4-wave Dutch home-visit family cohort (N = 114 with video at W1 in 2014a S2; N = 120/116 across W1–W4 in 2017); 2017 p. 1806 prints 'neither inflated nor inflated praise was effective' (for noninflated).	Linked into single shared family FAM_BRUMMELMAN_DUTCH_HOME (COH_NL_HOME_116) to prevent double-counting.

#	Source	Locator	Printed Discrepancy / Anomaly	Verification & Analytic Resolution
14	Francis, Boullion, & Yates (2025)	Table 2, p. 49	Column headers B and β are transposed across the entire regression table; 95% bias-corrected bootstrap CIs for Process praise [−0.079, 0.327] and Ambiguous praise [−0.039, 0.286] span zero despite bolded p-values.	Bivariate correlations extracted from verified Table 1 matrix (Age 4 N = 250; Age 8 N = 206); bootstrap CI crossing zero documented.
15	Gunderson, Donnellan et al. (2018a)	Main text p. 126 vs. Online Appendix B.5	All four prose statements describing grade-level differences in praise and criticism on p. 126 have the opposite sign to Online Appendix Table B.5; Study 2 SEM shows perceived process praise negatively predicting incremental mindset (β = −0.22, p = .026).	Verified directly from PMC5986600 Appendix B.4–B.5 matrices; 58 overlapping children between S1 and S2 never pooled in the same stratum.
16	Gunderson et al. (2013; 2018b)	2013 pp. 1536–1537 vs. PMC3655123 & 2018b	Process praise × mindset printed as r = .35 in 2013 journal text vs. r = .34 in PMC3655123 Table 2 and 2018b Table 2; person praise sign flips from r = +.05 (2013 text) to r = −.05 (PMC XML); recruited families reported as 64 vs. 63.	Linked as one cohort (COH_CHI_TODDLER_53); both r_person = −.05 and +.05 tested explicitly in Steiger dependent-difference sensitivity.
17	Lucca et al. (2019, S2) & Radovanovic et al. (2023, E1)	2019 S2 vs. 2023 E1 Method/Tables	Secondary recoding of the exact same 29 infant laboratory video recordings; 2023 tables label unstandardized B coefficients as β.	Linked into single shared cohort COH_UW_INFANT_S2_29 (FAM_SOMMERVILLE_INFANT) so N = 29 is counted once.
18	Xing et al. (2018)	Tables 1 & 3, pp. 4–5	Values of 0.12–0.18 printed under column header SD across three groups (n ≈ 34/arm) where F(2, 100) = 7.45 requires true SD ≈ 0.85.	Confirmed standard errors of the mean (SE = SD/√n) mislabeled as SD; exact Hedges' g computed from verified F/t statistics and recovered SDs.
19	Skipper & Douglas (2012, E1)	Results, p. 332	Omnibus test after Failure 2 printed as F(2, 142) = 0.02, p = .092.	Typographical error for exact p = .9802 (complete collapse of praise differences by second setback).
20	Zentall & Morris (2010; 2012) & Morris & Zentall (2014)	2010 p. 159; 2012 p. 1076; 2014 p. 4	Incompatible $\chi^2(12, N = 135) = 2.3$, $p < .001$ and Methods/Appendix item mismatch in 2010; pooled regression df printed as F(1, 29) with N = 30 in 2012; omnibus df printed as F(4, 94) with N = 95 (true df2 = 90) in 2014.	Pairwise Hedges' g computed from condition Ms, SDs, and cell counts; unrepaired χ^2 documented.
21	Mueller & Dweck (1998)	Studies 1, 3, 5, & 6	Study 1 Tables 1–2 cell counts sum to 128 vs. prose N = 123 and t(75); Study 3 prints t(51) for two-group contrast inside 3-arm N = 88; Study 5 prints F(2, 45) for N = 46 (true df2 = 43); Study 6 allocation mechanism unspecified.	Preserved conservative analyzed N = 123 in Study 1 and exact group Ms/SDs across all six studies.
22	Nahm & Park (2017) & Corpus & Lepper (2007, S2)	Nahm Table 4; Corpus p. 499	Nahm & Park Table 4 prints single-predictor standardized β = .48 alongside bivariate r = .66 (mathematically β must equal r in simple OLS); Corpus & Lepper S2 applied post-assignment praise-memory exclusion to praise arms (removing 22/60 = 36.7%, including 45% of Person/Product) but 0/16 of Neutral.	Used verified bivariate matrix for Nahm & Park; ran explicit randomized ITT (N = 76) vs. per-protocol (N = 48) sensitivity bounds for Corpus & Lepper S2.

S4. Complete Covariance (Var(A − B)) and Multi-Parameter Sensitivity Analyses

Table S5

All Nineteen Within-Cohort Steiger Dependent Contrasts Testing Process Praise Minus Person Praise Across Empirical and Plausible Cross-Praise Correlations.

Cohort & Endpoint	N	r_Proc	r_Pers	r_12	Δz [95% CI] (Emp.)	p (Emp.)	p (r=.20)	p (r=.35)	p (r=.50)
Pomerantz & Kempner (2013) — 6m Mindset (Proportion, r12=-.05)	110	+0.082	-0.207	-0.05	+0.292 [+0.02, +0.57]	.038	.018	.009	.003
Pomerantz & Kempner (2013) — 6m Mindset (Frequency, r12=+.39)	110	-0.091	-0.217	+0.39	+0.129 [-0.08, +0.34]	.231	.295	.245	.186
Pomerantz & Kempner (2013) — 6m Challenge Pref (Proportion, r12=-.05)	110	+0.124	-0.191	-0.05	+0.318 [+0.04, +0.59]	.024	.010	.004	.001
Pomerantz & Kempner (2013) — 6m Challenge Pref (Frequency, r12=+.39)	110	+0.036	-0.229	+0.39	+0.269 [+0.06, +0.48]	.013	.029	.016	.006
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	.026	.014	.005
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	.095	.065	.036
Gunderson, Sorhagen, et al. (2018b) — Grade 4 Math	50	+0.330	-0.230	-0.14	+0.577 [+0.14, +1.02]	.010	.002	.0007	.0001
Gunderson, Sorhagen, et al. (2018b) — Grade 4 Reading Comprehension	50	-0.020	-0.240	-0.14	+0.225 [-0.21, +0.66]	.310	.227	.181	.127
Gunderson, Sorhagen, et al. (2018b) — Grade 4 Reading Decoding	50	-0.070	+0.000	-0.14	-0.070 [-0.50, +0.36]	.750	.704	.674	.631
Ferrar et al. (2019) — 4-7yr Longitudinal IQ (GEE)	156	+0.039	-0.070	+0.25	+0.109 [-0.09, +0.30]	.271	.286	.237	.178
Ferrar et al. (2019) — 4yr Longitudinal Reading BQAL	125	-0.163	+0.011	+0.25	-0.175 [-0.39, +0.04]	.115	.127	.091	.054
Li, Zhang, Wu, & Zheng (2023) — Growth Mindset (r12=+.15)	292	-0.020	+0.080	+0.15	-0.100 [-0.25, +0.05]	.192	.179	.136	.089
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	.634	.597	.547
Li, Zhang, Wu, & Zheng (2023) — Grade 4 Chinese Exam (r12=+.15)	103	+0.120	-0.100	+0.15	+0.221 [-0.04, +0.48]	.091	.082	.054	.028
Gunderson, Donnellan, et al. (2018a, S2) — Incremental Mindset (r12=+.69)	267	-0.040	+0.040	+0.69	-0.080 [-0.18, +0.01]	.099	.304	.254	.194
Gunderson, Donnellan, et al. (2018a, S2) — Learning Goals (r12=+.69)	263	+0.410	+0.360	+0.69	+0.059 [-0.04, +0.16]	.261	.474	.430	.372
Francis, Boullion, & Yates (2025) — Age 8 Persistence (r12=+.11)	206	+0.110	-0.140	+0.11	+0.251 [+0.07, +0.44]	.008	.005	.002	.0004
Lucca, Horton, & Sommerville (2019, S1) — 18m Gear Stacking	29	+0.510	+0.090	+0.15	+0.472 [-0.05, +0.99]	.075	.067	.044	.022
Nahm & Park (2017) — 5yo Playfulness Disposition	49	+0.110	-0.120	+0.35	+0.231 [-0.10, +0.56]	.171	.217	.171	.119

Table S6

Sensitivity Analyses Across Within-Cohort Outcome Correlations (ρ_w), Cluster Design Effects (ICC/DEFF), and Forced Inclusion of Cluster-Confounded Classroom Data.

Sensitivity Scenario	Primary Estimate / Stratum 1	Comparator / Stratum 2	Naturalistic / Knapp–Hartung Inference
Within-cohort outcome correlation ρ_w = 0.20 (ICC = .00)	g = +0.681 [+0.432, +0.930]	g = +0.113 [-0.044, +0.271] (ns)	r = +0.102 [+0.028, +0.174]
Within-cohort outcome correlation ρ_w = 0.35 (ICC = .00)	g = +0.678 [+0.430, +0.926]	g = +0.111 [-0.044, +0.266] (ns)	r = +0.103 [+0.027, +0.177]
Within-cohort outcome correlation ρ_w = 0.50 (ICC = .00)	g = +0.675 [+0.427, +0.922]	g = +0.108 [-0.043, +0.260] (ns)	r = +0.104 [+0.027, +0.179]
Cluster design-effect inflation ICC_0.02 (ρ_w = .35)	g = +0.678 [+0.429, +0.928]	Remains null (p > .15)	r = +0.101 [+0.025, +0.176]

Sensitivity Scenario	Primary Estimate / Stratum 1	Comparator / Stratum 2	Naturalistic / Knapp–Hartung Inference
Cluster design-effect inflation ICC_0.05 (ρ_w = .35)	g = +0.679 [+0.428, +0.930]	Remains null (p > .15)	r = +0.100 [+0.022, +0.176]
Cluster design-effect inflation ICC_0.10 (ρ_w = .35)	g = +0.680 [+0.427, +0.934]	Remains null (p > .15)	r = +0.098 [+0.019, +0.176]
Level 2a + Zarrinabadi (2023) Pre–Post DiD (ICC_0.00_DEFF_1.00)	Zarrinabadi 4-scale DiD g = +1.220 [+0.719, +1.722]	Pooled k = 3 REML g = +0.758	K–H 95% CI [-0.428, +1.943], p_HK = .111 (ns)
Level 2a + Zarrinabadi (2023) Pre–Post DiD (ICC_0.05_DEFF_2.00)	Zarrinabadi 4-scale DiD g = +1.220 [+0.510, +1.930]	Pooled k = 3 REML g = +0.719	K–H 95% CI [-0.461, +1.899], p_HK = .120 (ns)
Level 2a + Zarrinabadi (2023) Pre–Post DiD (ICC_0.10_DEFF_3.00)	Zarrinabadi 4-scale DiD g = +1.220 [+0.351, +2.090]	Pooled k = 3 REML g = +0.673	K–H 95% CI [-0.492, +1.837], p_HK = .131 (ns)

S5. Result-Specific Risk-of-Bias Summary (RoB 2 and JBI Checklists)

- **Immediate Laboratory Trials (RoB 2):** Early scripted experiments (Mueller & Dweck, 1998; Kamins & Dweck, 1999; Cimpian et al., 2007; Zentall & Morris, 2010, 2012) typically drew 'some concerns' in RoB 2 Domain 1 (unreported allocation concealment) and Domain 5 (absent preregistration), while self-reported persistence ratings inevitably reflected awareness of spoken feedback. Preregistration changed that profile. Bennett-Pierre et al. (2024; preregistered protocol, automated timing, N = 150) achieved low risk on measurement and reporting (with 'some concerns' in Domain 1 from baseline puzzle-time imbalance, F = 3.93, p = .005) and found g = +0.05 (*ns*).
- **Delayed & Multi-Week Classroom Studies (RoB 2 / JBI Quasi-Experimental):** Corpus and Lepper (2007, Study 2) paired masked classroom behavioral coding (Domain 4: low risk) with high risk in Domains 2 and 3 because post-randomization memory exclusions targeted praised arms exclusively (dropping 22 of 60 praised preschoolers vs. 0 of 16 neutral controls) and primary tests pooled praise arms together. Zarrinabadi et al. (2023) drew critical/high risk on JBI quasi-experimental criteria due to one classroom per arm, a +93.76-point baseline imbalance, and irreproducible paired t-tests.
- **Naturalistic Caregiver Speech Cohorts (JBI Cohort Checklist):** Prospective home and diary studies (Gunderson et al., 2013, 2018b; Pomerantz & Kempner, 2013; Ferrar et al., 2019; Francis et al., 2025) achieved strong inter-coder reliability on parent speech and independent follow-up testing, although unmeasured evocative child traits, unrecorded intervening parental talk across multi-year lags, and covariate specification choices remain inherent observational constraints.