

Do Teacher Expectations Raise Students' IQ?

A frozen, data-only baseline for comparing an ordinary PDF workflow with the WebMCP-native Living Evidence document.

18 experiments / 19 effect-size records

Baseline version: 2026-09-04.v1

Analytic dataset: metadat::dat.raudenbush1985

Synthesis DOI: 10.1037/0022-0663.76.1.85

Purpose

This file is the control artifact in a descriptive A/B evaluation. It contains the same study-level numbers and methods context as the web document, but no executable statistics, registered browser tools, or interactive audit trail. It is not evidence that either condition performs better.

Traceability limit

The y_i and v_i values are secondary-dataset transcriptions. Primary reports and effect-size derivations were not independently checked, and no structured risk-of-bias assessment is included. The comparison measures task execution on this frozen artifact, not scientific truth.

Evidence summary and model

The source reports 18 experiments. The table contains 19 effect-size records because the Pellegrini and Hicks (1972) experiment contributes separate aware-tester and blind-tester conditions (s04 and s05). The historical reference analysis treats the 19 rows as independent and does not model their within-experiment covariance.

Full-corpus reference result

A random-effects REML fit over all 19 records gives pooled SMD = 0.0837, 95% CI [-0.0175, 0.1849], two-sided p = 0.1051. These printed values make the full-corpus task an extraction control. Results for exclusion and Egger tasks are intentionally not printed.

Statistical specification

Each observed standardized mean difference y_i has known sampling variance v_i . The random-effects model estimates a pooled mean and between-record variance tau-squared by restricted maximum likelihood (REML). Confidence intervals use the standard normal 1.959964 critical value, and the pooled-effect p value is a two-sided normal test against zero.

Egger's diagnostic regresses standardized effect ($y_i / \sqrt{v_i}$) on precision ($1 / \sqrt{v_i}$); the intercept is tested with a two-sided t distribution with $k - 2$ degrees of freedom. The document-registered classification rule is: rule passed when $p \geq 0.10$, rule failed when $p < 0.05$, and inconclusive otherwise. This classification is not a judgment of publication bias or scientific validity.

Effect direction

Positive SMD means the experimentally designated expectancy group had higher measured IQ than the control group. weeks is teacher-student contact before expectancy induction. setting is group or individual testing; tester records whether the tester knew the assignment.

Study-level data (1/2)

All rows are secondary-dataset transcriptions. Source locator: corresponding row in metadat::dat.raudenbush1985. Primary-source checked: no. Risk of bias: not assessed.

ID	Experiment	Author	Year	Weeks	Set	Tester	n1	n2	yi	vi
s01	s01	Rosenthal et al.	1974	2	group	aware	77	339	0.03	0.0156
s02	s02	Conn et al.	1968	21	group	aware	60	198	0.12	0.0216
s03	s03	Jose & Cody	1971	19	group	aware	72	72	-0.14	0.0279
s04	P&H-1972	Pellegrini & Hicks	1972	0	group	aware	11	22	1.18	0.1391
s05	P&H-1972	Pellegrini & Hicks	1972	0	group	blind	11	22	0.26	0.1362
s06	s06	Evans & Rosenthal	1969	3	group	aware	129	348	-0.06	0.0106
s07	s07	Fielder et al.	1971	17	group	blind	110	636	-0.02	0.0106
s08	s08	Claiborn	1969	24	group	aware	26	99	-0.32	0.0484
s09	s09	Kester	1969	0	group	aware	75	74	0.27	0.0269
s10	s10	Maxwell	1970	1	indiv	blind	32	32	0.80	0.0630

s04 and s05 share one experiment identifier. Excluding s04 removes only the aware-tester record; s05 remains in the requested sensitivity fit.

Study-level data (2/2)

All rows are secondary-dataset transcriptions. Source locator: corresponding row in `metadat::dat.raudenbush1985`. Primary-source checked: no. Risk of bias: not assessed.

ID	Experiment	Author	Year	Weeks	Set	Tester	n1	n2	yi	vi
s11	s11	Carter	1970	0	group	blind	22	22	0.54	0.0912
s12	s12	Flowers	1966	0	group	blind	43	38	0.18	0.0497
s13	s13	Keshock	1970	1	indiv	blind	24	24	-0.02	0.0835
s14	s14	Henrikson	1970	2	indiv	blind	19	32	0.23	0.0841
s15	s15	Fine	1972	17	group	aware	80	79	-0.18	0.0253
s16	s16	Grieger	1970	5	group	blind	72	72	-0.06	0.0279
s17	s17	Rosenthal & Jacobson	1968	1	group	aware	65	255	0.30	0.0193
s18	s18	Fleming & Anttonen	1971	2	group	blind	233	224	0.07	0.0088
s19	s19	Ginsburg	1970	7	group	aware	65	67	-0.07	0.0303

Provenance and limitations

Analytic values: Viechtbauer's open metadat distribution, `dat.raudenbush1985`. Synthesis: Raudenbush, S. W. (1984), Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments, *Journal of Educational Psychology* 76(1), 85-97, DOI 10.1037/0022-0663.76.1.85.

This baseline does not contain primary-paper page and table locators for each contributing record, verbatim primary quotations, independently reconstructed effect-size derivations, or risk-of-bias judgments. Missingness is explicit so that absence is not mistaken for a negative assessment.

The corpus is one historical meta-analytic dataset. A score on this artifact does not establish general performance on other papers, effect measures, study designs, or scientific domains.