Workshop on Statistical Applications in Meta-Analysis

Robert M. Bernard & Phil C. Abrami

Centre for the Study of Learning and Performance
and CanKnow
Concordia University

May 16, 2007

CSLP/CEAP
CENTRE FOR THE STUDY OF LEARNING AND PERFORMANCE
CENTRE D'ÉTUDES SUR L'APPRENTISSAGE ET LA PERFORMANCE

Canadian Centre for Knowledge Utilization

Two Main Purposes of a Meta-Analysis

- Estimate the population central tendency and variability of effect sizes between an intervention (treatment) condition and a control condition.
- Explore unexplained variability through the analysis of methodological and substantive coded study features.

10 Steps in Planning and Conducting a Systematic Review/Meta-Analysis

1. Determine the research question
2. Develop terms and definitions related to the question
3. Develop a search strategy for identification of relevant studies
4. Establish criteria for inclusion and exclusion of studies
5. Select studies based on abstract review (agreement)
6. Select studies based on full-text review (agreement)
7. Extract effect sizes (agreement)
8. Develop codebook of study features
9. Code studies (agreement)
10. Conduct statistical analysis and interpretation

Outcomes of Three Designs

Type of Research Design	*Pretest O*	*X or No X*	*Posttest O*
One group Pretest-posttest design			
Experimental Group	*70*	*X*	*90*
Pre-test Post-test Control Group Design			
Experimental Group (random assignment)	*70*	*X*	*90*
Control Group (random assignment)	*70*		*80*
Nonequivalent pretest posttest control group design			
Experimental Group	*70*	*X*	*90*
Control Group	*65*		*80*

10 Steps in a Meta-Analysis

7. Extract effect sizes

Effect sizes extraction involves converting descriptive or other statistical information contained in studies into a standard metric by which studies can be compared.

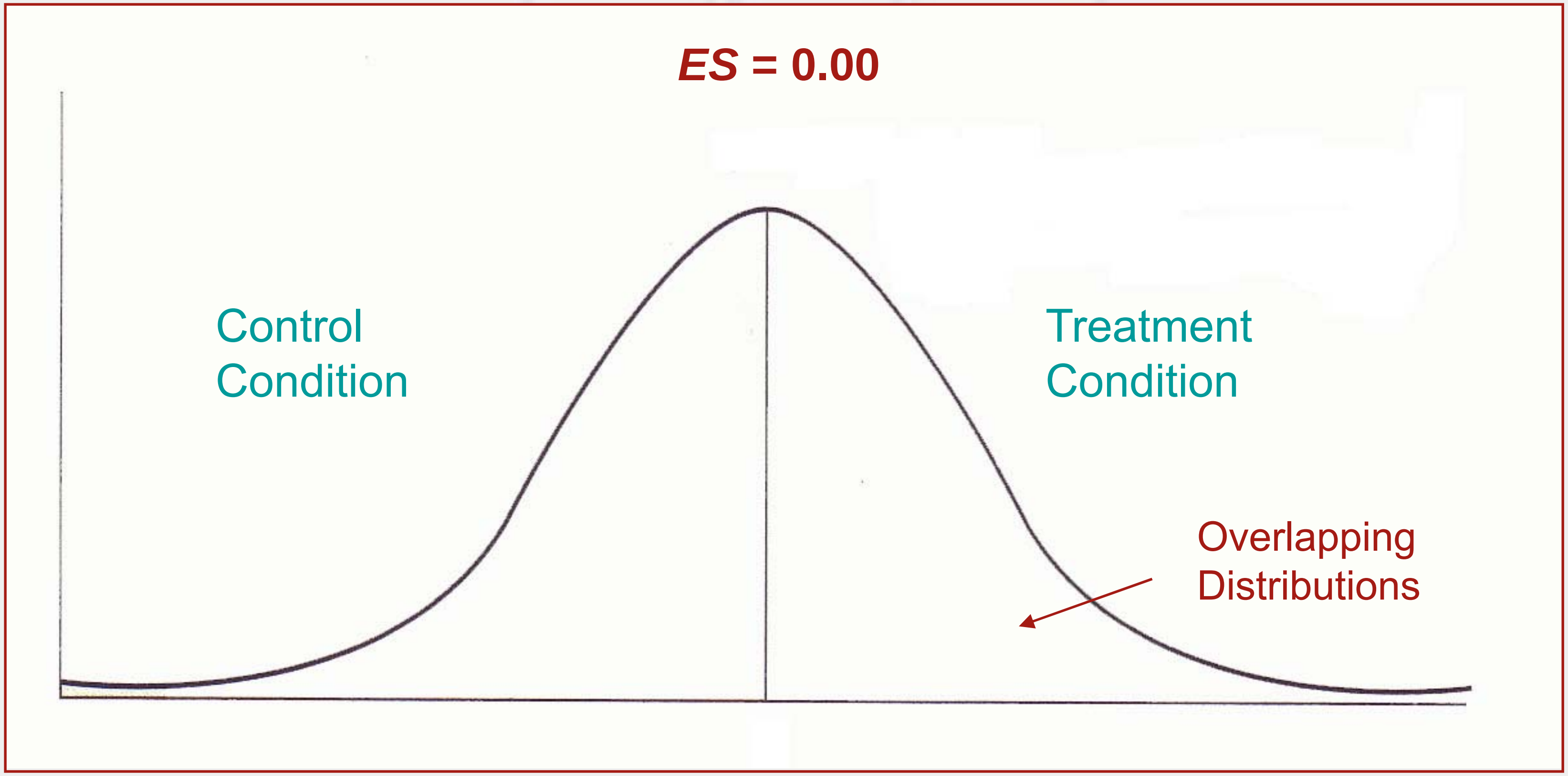
What is an Effect size?

- A descriptive metric that characterizes the standardized difference (in *SD* units) between the mean of a control group and the mean of a treatment group (educational intervention)
- Can also be calculated from correlational data derived from pre-experimental designs or from repeated measures designs

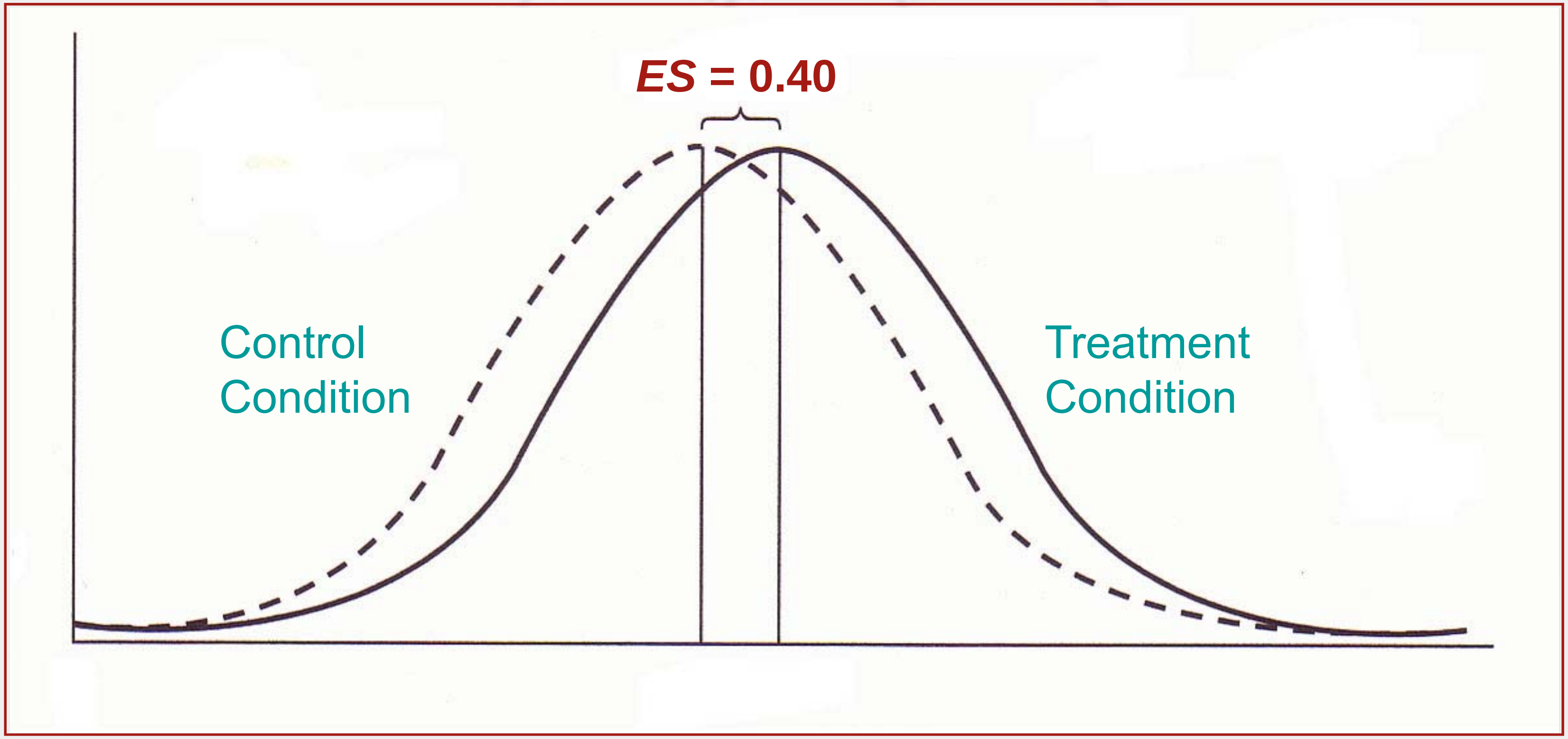
Characteristics of Effect Sizes

- Can be positive or negative
- Interpreted as a *z*-score, in SD units, although individual effect sizes are not part of a *z*-score distribution
- Can be aggregated with other effect sizes and subjected to other statistical procedures such as ANOVA and multiple regression
- Magnitude interpretation: ≤ 0.20 is a small effect size, 0.50 is a moderate effect size and ≥ 0.80 is a large effect size (Cohen, 1992)

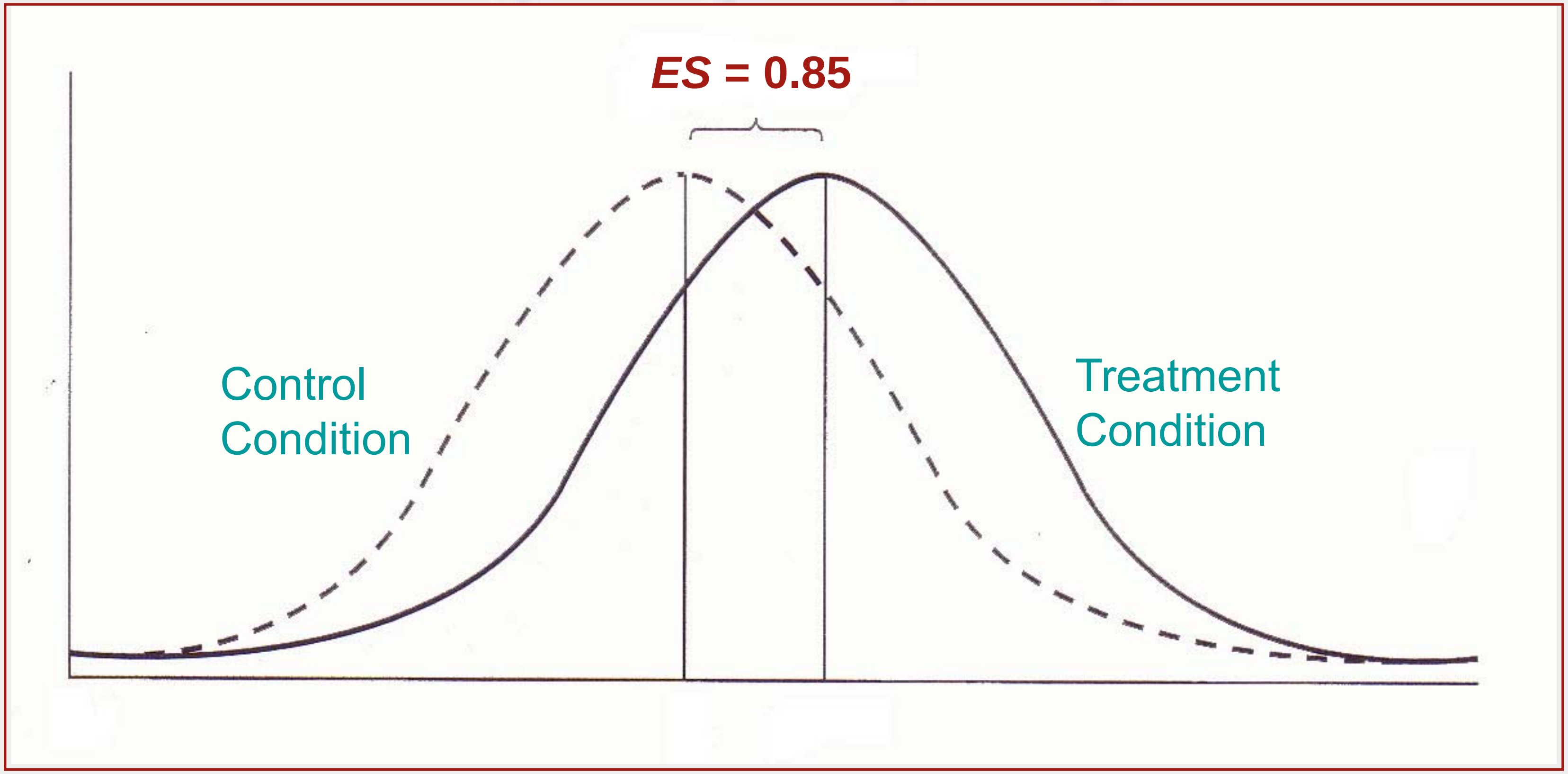
Zero Effect Size



Moderate Effect Size



Large Effect Size



ES Calculation: Descriptive Statistics

$$\Delta_{Glass} = \frac{\bar{Y}_{Experimental} - \bar{Y}_{Control}}{SD_{Control}}$$

$$d_{Cohen} = \frac{\bar{Y}_{Experimental} - \bar{Y}_{Control}}{SD_{Pooled}}$$

$$SD_{pooled} = \sqrt{((N_E - 1)SD^2{}_E + (N_C - 1)SD^2{}_C)) / (N_{Total} - 2)}$$

Note: this equation is the same as adding two SSs and dividing by df_{Total}

Adjustment for Small Samples: Hedges' *g*

- Cohen's *d* is inaccurate for small samples ($N < 20$), so Hedges' *g* was developed (Hedges & Olkin, 1985)

$$g_{Hedges} = \frac{\bar{Y}_{Experimental} - \bar{Y}_{Control}}{\sqrt{((N_E - 1) \cdot SD^2{}_E + (N_C - 1)SD^2{}_C)) / (N_{Tot} - 2)}} \cdot \left(1 - \frac{3}{4(N_E + N_C) - 9}\right)$$

g = Cohen's *d* times a multiplier based on sample size

Example of *ES* Extraction with Descriptive Statistics

Study reports:	Treatment mean = 42.8	Control Mean = 32.5
	Treatment *SD* = 8.6	Control *SD* = 7.4
	n = 26	*n* = 31

Procedure: Calculate SD_{pooled} Calculate *d* and *g*

$$SD_{pooled} = \sqrt{((26-1)8.6^2) + (31-1)7.4^2)) / (57-2)}$$

$$SD_{pooled} = \sqrt{(1849 + 1642.8) / 55} = \sqrt{3491.8 / 55} = \sqrt{63.49} = \boxed{7.97}$$

$$d = \frac{42.8 - 32.5}{\boxed{7.97}} = \frac{10.3}{7.97} = \boxed{1.29}$$

$$g = d\left(1 - \frac{3}{(4(N_E + N_C)) - 9}\right) = 1.29\left(1 - \frac{3}{4(26+31) - 9}\right) = 1.29\left(1 - \frac{3}{219}\right) = \boxed{1.27}$$

ES Extraction Exercise

Materials:

- EXCEL SE Calculator
- 5 studies from which to extract effect sizes

Alternative Methods of *ES* Extraction: Exact Statistics

- Study Reports: $t\ (60) = 2.66,\ p < .05$

$$d = \frac{2t}{\sqrt{df}} = \frac{2(2.66)}{\sqrt{60}} = \frac{5.32}{7.46} = 0.687$$

- Study Reports: $F\ (1, 61) = 7.08,\ p < .05$

Convert *F* to *t* and apply the above equation:

$$t = \sqrt{F} = 2.66; df = 60$$

$$d = \frac{2t}{\sqrt{df}} = \frac{2(2.66)}{\sqrt{60}} = \frac{2(2.66)}{7.46} = \frac{5.32}{7.46} = 0.687$$

Alternative Methods of *ES* Extraction: Exact *p*-value

- Study Reports: t (60) is sig. $p = 0.013$

Look up t-value for $p = 0.013$

$t = 2.68$

$$d \geq t\sqrt{\frac{1}{N_E} + \frac{1}{N_C}}$$

$$d \geq 2.68\sqrt{\frac{1}{31} + \frac{1}{31}} = 2.68(0.254) = 0.681$$

Statistics Related to Effect Size

Standard Error: $$\hat{\sigma}_g = \sqrt{\frac{1}{n_e} + \frac{1}{n_c} + \frac{g^2}{2(n_e + n_c)}}\left(1 - \frac{3}{4(n_e + n_c) - 9}\right)$$

Variance: $$\hat{\sigma}_g^2 = (\hat{\sigma}_g)^2$$

Inverse Variance (*w*): $$w_i = \frac{1}{\hat{\sigma}^2}$$

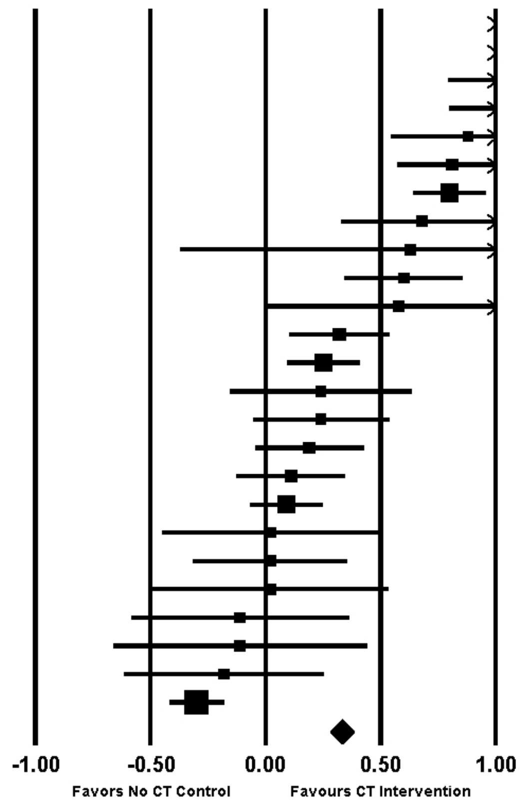
Weighted g (*g**w*): $$Weightedg = (w_i)(g_i)$$

HedgesÕ g	Standard Error ($\hat{\sigma}_g$)	Variance ($\hat{\sigma}_g^2$)	95th Lower Limit	95th Upper Limit	z-Value	p-Value	Weights (w_i)	Weighted g (w_i)(g_i)
2.44	0.22	0.05	2.00	2.88	10.89	0.00	19.94	48.65
2.31	0.17	0.03	1.98	2.64	13.59	0.00	34.60	79.93
1.38	0.30	0.09	0.79	1.97	4.60	0.00	11.11	15.33
1.17	0.19	0.04	0.80	1.54	6.16	0.00	27.70	32.41
0.88	0.17	0.03	0.55	1.21	5.18	0.00	34.60	30.45
0.81	0.12	0.01	0.57	1.05	6.75	0.00	69.44	56.25
0.80	0.08	0.01	0.64	0.96	10.00	0.00	156.25	125.00
0.68	0.18	0.03	0.33	1.03	3.78	0.00	30.86	20.99
0.63	0.51	0.26	-0.37	1.63	1.24	0.22	3.84	2.42
0.60	0.13	0.02	0.35	0.85	4.62	0.00	59.17	35.50
0.58	0.29	0.08	0.01	1.15	2.00	0.05	11.89	6.90
0.32	0.11	0.01	0.10	0.54	2.91	0.00	82.64	26.45
0.25	0.08	0.01	0.09	0.41	3.13	0.00	156.25	39.06
0.24	0.20	0.04	-0.15	0.63	1.20	0.23	25.00	6.00
0.24	0.15	0.02	-0.05	0.53	1.60	0.11	44.44	10.67
0.19	0.12	0.01	-0.05	0.43	1.58	0.11	69.44	13.19
0.11	0.12	0.01	-0.13	0.35	0.92	0.36	69.44	7.64
0.09	0.08	0.01	-0.07	0.25	1.13	0.26	156.25	14.06
0.02	0.24	0.06	-0.45	0.49	0.08	0.93	17.36	0.35
0.02	0.17	0.03	-0.31	0.35	0.12	0.91	34.60	0.69
0.02	0.26	0.07	-0.49	0.53	0.08	0.94	14.79	0.30
-0.11	0.24	0.06	-0.58	0.36	-0.46	0.65	17.36	-1.91
-0.11	0.28	0.08	-0.66	0.44	-0.39	0.69	12.76	-1.40
-0.18	0.22	0.05	-0.61	0.25	-0.82	0.41	20.66	-3.72
-0.30	0.06	0.00	-0.42	-0.18	-5.00	0.00	277.78	-83.33
0.330	0.03	0.00	0.28	0.38	12.62	0.00	1458.21*	481.87*

Meta Analysis of CT Studies

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1.000	2.440	0.224	0.050	2.001	2.879	10.895	0.000
2.000	2.310	0.170	0.029	1.977	2.643	13.588	0.000
3.000	1.380	0.300	0.090	0.792	1.968	4.600	0.000
4.000	1.170	0.190	0.036	0.798	1.542	6.158	0.000
5.000	0.880	0.170	0.029	0.547	1.213	5.176	0.000
6.000	0.810	0.120	0.014	0.575	1.045	6.750	0.000
7.000	0.800	0.080	0.006	0.643	0.957	10.000	0.000
8.000	0.680	0.180	0.032	0.327	1.033	3.778	0.000
9.000	0.630	0.510	0.260	-0.370	1.630	1.235	0.217
10.000	0.600	0.130	0.017	0.345	0.855	4.615	0.000
11.000	0.580	0.290	0.084	0.012	1.148	2.000	0.046
12.000	0.320	0.110	0.012	0.104	0.536	2.909	0.004
13.000	0.250	0.080	0.006	0.093	0.407	3.125	0.002
14.000	0.240	0.200	0.040	-0.152	0.632	1.200	0.230
15.000	0.240	0.150	0.023	-0.054	0.534	1.600	0.110
16.000	0.190	0.120	0.014	-0.045	0.425	1.583	0.113
17.000	0.110	0.120	0.014	-0.125	0.345	0.917	0.359
18.000	0.090	0.080	0.006	-0.067	0.247	1.125	0.261
20.000	0.020	0.240	0.058	-0.450	0.490	0.083	0.934
19.000	0.020	0.170	0.029	-0.313	0.353	0.118	0.906
21.000	0.020	0.260	0.068	-0.490	0.530	0.077	0.939
22.000	-0.110	0.240	0.058	-0.580	0.360	-0.458	0.647
23.000	-0.110	0.280	0.078	-0.659	0.439	-0.393	0.694
24.000	-0.180	0.220	0.048	-0.611	0.251	-0.818	0.413
25.000	-0.300	0.060	0.004	-0.418	-0.182	-5.000	0.000
	0.330	0.026	0.001	0.279	0.382	12.619	0.000

Hedges's g and 95% CI



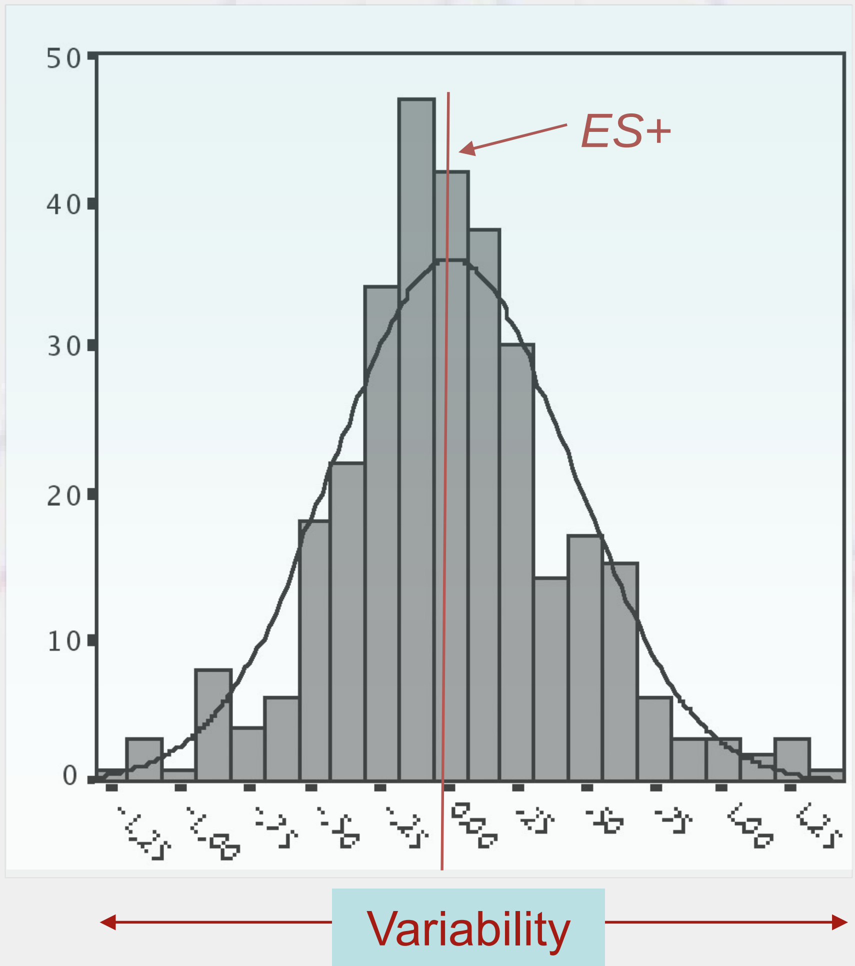
Meta Analysis

10 Steps in a Meta-Analysis

10: Analysis and interpretation

Analysis involves invoking a range of standard statistical tests to examine average effect sizes, variability and the relationship between study features and effect size. Interpretation is drawing conclusion from these analyses.

Mean and Variability



Note: Results from Bernard, Abrami, Lou, et al. (2004) *RER*

Mean Effect Size

$g+$	$g+ = \dfrac{\sum_{i=1}^{k} (w_i)(g_i)}{\sum_{i=1}^{k} w_i}$	$g+ = \dfrac{481.87}{1458.21} = 0.330$
Var	$\hat{\sigma}^2_{g+} = \left(\sum_{i=1}^{k} \dfrac{1}{\hat{\sigma}^2} \right)^{-1} = \dfrac{1}{\sum_{i=1}^{k} \dfrac{1}{\hat{\sigma}^2}} = \dfrac{1}{\sum_{i=1}^{k} w_i}$	$\hat{\sigma}^2_{g+} = \dfrac{1}{1458.21} = 0.0007$
SE	$\hat{\sigma}_{g+} = \sqrt{\hat{\sigma}^2}$	$\hat{\sigma}_{g+} = \sqrt{0.00} = 0.0265$
z	$z_{g+} = \dfrac{g+}{\hat{\sigma}_{g+}}$	$z_{g+} = \dfrac{0.330}{0.0265} \cong 12.62$

Conclusion: Mean *g* = 0.33 and it is significant.

VARIABILITY (*Q* STATISTIC)

Question: How much variability surrounds g+ and is it significant? Are the effect sizes heterogeneous or homogeneous?

$$Q = \sum_{i=1}^{k} \frac{(g - g+)^2}{\hat{\sigma}_g^2}$$

$$Q_{Total} = \frac{(2.44 - 0.330)^2)}{0.11} + \frac{(2.31 - 0.330)^2}{0.03} + \ldots \frac{(-0.18 - 0.330)^2}{0.22} + \frac{(-0.30 - 0.330)^2}{0.07} = 469.54$$

Q-value	df (Q)	P-value
469.54	24	0.000

Tested with the χ^2 distribution.

Conclusion: Effect sizes are heterogeneous.

Homogeneity vs. Heterogeneity of Effect Size

- If homogeneity of effect size is established, then the studies in the meta-analysis can be thought of as sharing the same effect size (i.e., the mean)
- If homogeneity of effect size is violated (heterogeneity of effect size), then no single effect size is representative of the collection of studies (i.e., the "true" mean effect size remains unknown)

Statistics in Comprehensive Meta-Analysis™

Effect size and 95% confidence interval						Test of null (2-Tail)	
Number Studies	Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-value	P-value
25	0.33	0.03	0.00	0.28	0.38	12.62	0.00

Heterogeneity		
Q-value	df (Q)	P-value
469.54	24	0.00

Interpretation: Moderate *ES* for all outcomes (*g+* = 0.33) in favor of the intervention condition.

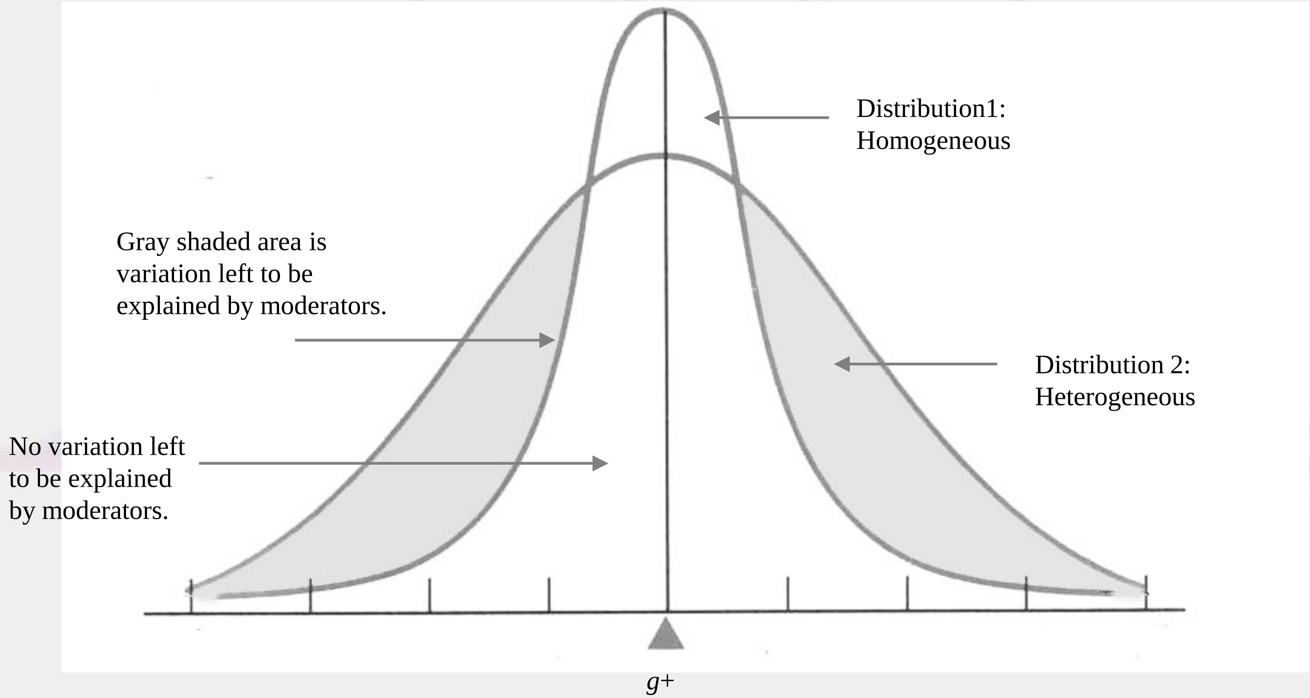
Homogeneity of ES is violated. *Q-value* is significant (i.e., there is too much variability for *g+* to represent a true average in the population).

Comprehensive Meta-Analysis 2.0.027 is a trademark of BioStat®

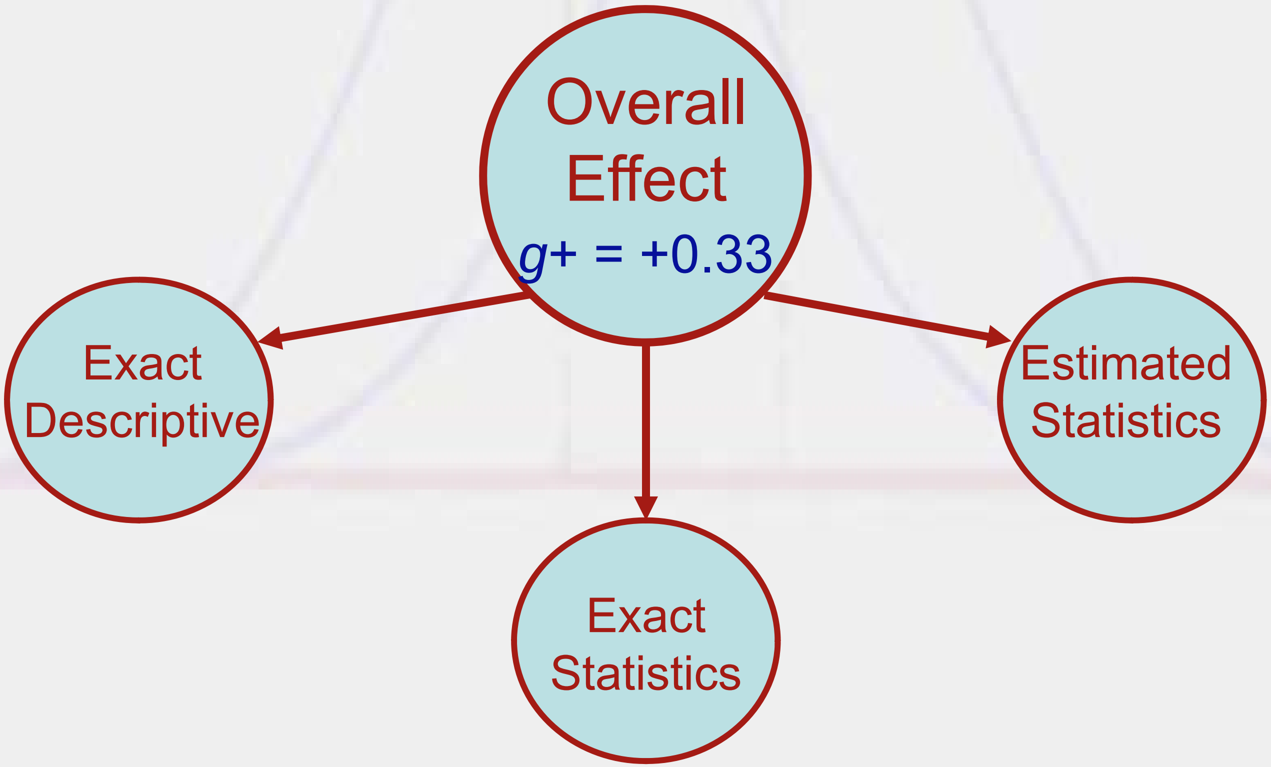
Back to ES Calculator

1. Interpretation of Mean Effect Size
2. Interpretation of Q-Statistic

Homogeneity versus Heterogeneity of Effect Size



Examining the Study Feature "Method of ES Extraction"



Tests of Levels of "Method of ES Extraction"

Groups		Effect size and 95% confidence interval				Heterogeneity		
Group	**N of Studies**	**Point estimate**	**Standard error**	**Lower limit**	**Upper limit**	**Q-value**	**df (Q)**	**P-value**
Descriptive	15	0..29	0.03	0.22	0.35	402.56	14	0.00
Statistics	3	0.21	0.06	0.09	0.33	0.97	2	0.62
Est. Statistics	7	0.63	0.06	0.50	0.75	37.00	6	0.00
Total within						442.50	22	0.00
Total between						27.04	2	0.00
Overall	25	0.33	0.03	0.28	0.38	469.54	24	0.00

Interpretation: Small to Moderate *ESs* for all categories in favor of the intervention condition.

Homogeneity of ES is violated. *Q-value* is significant for all categories (i.e., "Method of ES Extraction" does not explain enough variability to reach homogeneity).

Meta-Regression

Seeks to determine if “Method of ES Extraction” predicts effect size.

	Point Estimate	**Standard Error**	**Lower limit**	**Upper limit**	***z*-value**	***p*-value**
Extraction Method (β)	0.14	0.03	0.07	0.20	3.94	0.00
Intercept	0.26	0.03	0.21	0.32	8.34	0.00

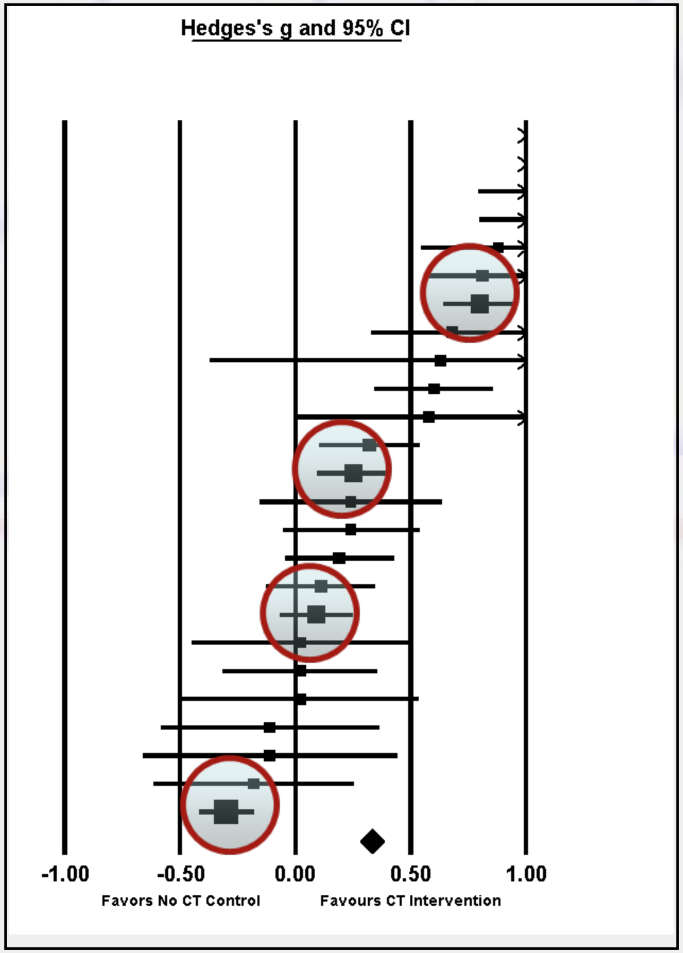
	Q	***df***	***p*-value**
Model	15.50	1	0.00
Residual	454.04	23	0.00
Total	469.54	24	0.00

Conclusion: “Method of Extraction” design is a significant predictor of *ES* but *ES* is still heterogeneous.

Sensitivity Analysis

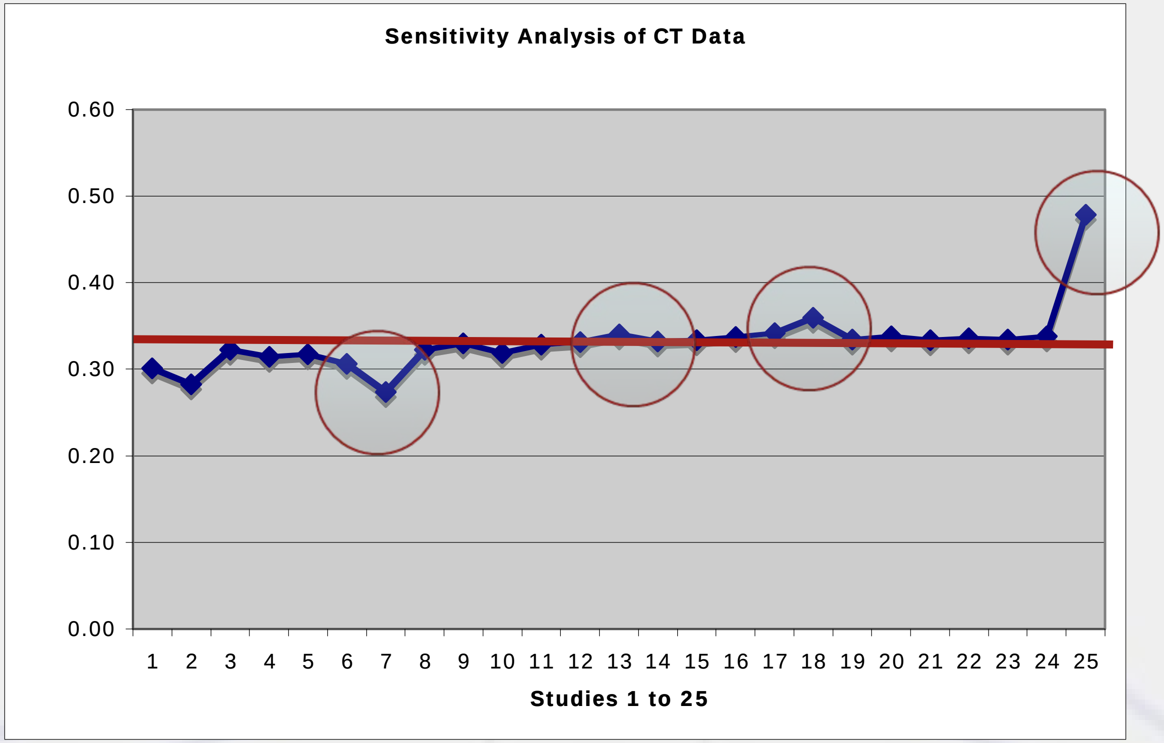
- Tests the robustness of the findings
- Asks the question: Will these results stand up when potentially distorting or deceptive elements, such as outliers, are removed?
- Particularly important to examine the robustness of the effect sizes of study features, as these are usually based on smaller numbers of outcomes

Sensitivity Analysis: Low Standard Error Samples



One Study Removed Analysis

Study	Point	SE	Variance	Lower limit	Upper limit	*z*-Value	*p*-Value
1.00	0.30	0.03	0.00	0.25	0.35	11.42	0.00
2.00	0.28	0.03	0.00	0.23	0.33	10.65	0.00
3.00	0.32	0.03	0.00	0.27	0.37	12.26	0.00
4.00	0.31	0.03	0.00	0.26	0.37	11.88	0.00
5.00	0.32	0.03	0.00	0.27	0.37	11.96	0.00
6.00	0.31	0.03	0.00	0.25	0.36	11.42	0.00
7.00	0.27	0.03	0.00	0.22	0.33	9.89	0.00
8.00	0.32	0.03	0.00	0.27	0.37	12.20	0.00
9.00	0.33	0.03	0.00	0.28	0.38	12.57	0.00
10.00	0.32	0.03	0.00	0.27	0.37	11.93	0.00
11.00	0.33	0.03	0.00	0.28	0.38	12.49	0.00
12.00	0.33	0.03	0.00	0.28	0.38	12.28	0.00
13.00	0.34	0.03	0.00	0.29	0.39	12.27	0.00
14.00	0.33	0.03	0.00	0.28	0.38	12.57	0.00
15.00	0.33	0.03	0.00	0.28	0.39	12.53	0.00
16.00	0.34	0.03	0.00	0.28	0.39	12.58	0.00
17.00	0.34	0.03	0.00	0.29	0.39	12.73	0.00
18.00	0.36	0.03	0.00	0.30	0.41	12.96	0.00
20.00	0.33	0.03	0.00	0.28	0.39	12.69	0.00
19.00	0.34	0.03	0.00	0.29	0.39	12.75	0.00
21.00	0.33	0.03	0.00	0.28	0.39	12.68	0.00
22.00	0.34	0.03	0.00	0.28	0.39	12.74	0.00
23.00	0.33	0.03	0.00	0.28	0.39	12.71	0.00
24.00	0.34	0.03	0.00	0.29	0.39	12.81	0.00
25.00	0.48	0.03	0.00	0.42	0.54	16.45	0.00
Total	**0.33**	**0.03**	**0.00**	**0.28**	**0.38**	**12.62**	**0.00**



Studies with High Weighted *g*+	*g*	*g*+	*g*+ with study removed	Difference	(*w*)	(*g*)(*w*)	%* Influence
Study 7	0.80	0.330	0.27	-0.06	156.25	125.00	25.9
Study 13	0.25	0.330	0.34	+0.04	156.25	39.09	8.1
Study 18	0.02	0.330	0.36	+0.06	156.25	14.06	2.9
Study 25	-0.30	0.330	0.48	+0.15	277.78	-83.33	17.41
Totals					746.53		54.31

*% Influence = (*g*)(*w*)/481.87 (100)

Steps in Controlling for Study Quality

- Step one: Are the effect sizes homogeneous?
- Step two: Does study quality explain the heterogeneity?
- Step three: Which qualities of studies matter?
- Step four: How do we deal with the differences?

Controlling Study Quality Using Dummy Coding in Meta-Regression

Categories of Study Quality	Dummy 1	Dummy 2	Dummy 3	Dummy 4
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0
5	0	0	0	1

Adjusting Effect Sizes

Categories	$g+$ Before Adjustment	$g+$ After Adjustment	Adjusted Heterogeneity		
			Q_{Within}	df	p
1	-0.185	-0.185	2.243	3	0.524
2	-0.218	-0.218	3.302	3	0.347
3	0.683	-0.065	3.252	3	0.354
4	0.565	-0.183	4.953	3	0.175
5	0.390	-0.358	1.985	3	0.576
Total	0.247	-0.202	15.734	15	0.400

Selected References

Bernard, R. M., Abrami, P. C., Lou, Y. Borokhovski, E., Wade, A., Wozney, L., Wallet, P.A., Fiset, M., & Huang, B. (2004). How does distance education compare to classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

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

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. [ERIC Document Reproduction Service No. ED 309 952].