Reviewing systematic reviews: meta-analysis of What Works Clearinghouse computer-assisted interventions.

November 2011

American Evaluation Association Annual Meeting, Anaheim

Andrei Streke • Tsze Chan

Session Title: Advanced Analytic Techniques in Educational Evaluation
Multipaper Session 244
Thursday, Nov 3

Presentation Overview

- What Works Clearinghouse systematic reviews
- Meta-analysis of computer-assisted programs across WWC topic areas, reading outcomes
- Meta-analysis of computer-assisted programs within Beginning Reading topic area

WWC Systematic Review

- A clearly stated set of objectives with pre-defined eligibility criteria for studies
- An explicit reproducible methodology
- A systematic search that attempts to identify all studies that would meet the eligibility criteria
- An assessment of the validity of the findings of the included studies
- A systematic presentation, and synthesis, of the characteristics and findings of the studies

WWC Systematic Review

Normative documents (http://ies.ed.gov/ncee/wwc):

- **WWC Procedures and Standards Handbook**
- **WWC topic area review protocol**

WWC products:

- **Intervention reports**
  http://ies.ed.gov/ncee/wwc/publications_reviews.aspx
- **Practice guides**
- **Quick reviews**

Selection Criteria for Beginning Reading Topic Area

- Manuscript is written in English and published 1983 or later
- Both published and unpublished reports are included
- Eligible designs: RCT; QED with statistical controls for pretest and/or a comparison group matched on pretest; regression discontinuity; SCD
- At least one relevant quantitative outcome measure
- Manuscript focuses on beginning reading
- Focus is on students ages 5-8 and/or in grades K-3.
- Primary language of instruction is English

Examples of problematic study designs that do not meet WWC criteria

- **Designs that confound study condition and study site**
  - **Programs that were tested with only one treatment and one control classroom or school**

- **Non-comparable groups**
  - **Study designs that compared struggling readers to average or good readers to test a program's effectiveness**

WWC Intervention reports

- Program description
- Intervention rating
- Technical Appendices
  - Study characteristics
  - Outcomes characteristics
  - Study findings: effect sizes and improvement indices
    http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_accelreader_app_101408.pdf

Appendix A3.2 Summary of study findings included in the rating for reading fluency domain[1]

			Authors' findings from the study		WWC calculations			
			Mean outcome[2] (standard deviation)[3]					
Outcome measure	**Study sample**	**Sample size (clusters/ students)**	***Success Maker* group**	**Comparison group**	**Mean difference[4] (*Success Maker* – comparison)**	**Effect size[5]**	**Statistical significance[6] (at $\alpha = 0.05$)**	**Improvement index[7]**
Beattie, 2000 (randomized controlled trial with attrition)[8]								
Gray Oral Reading test (GORT-3)	11-16 yrs old	26	83.18 (12.72)	79.50 (17.76)	3.68	0.23	ns	+9
Average for reading fluency (Beattie, 2000)[9]						**0.23**	**ns**	**+9**

[1] This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading fluency domain.
[2] The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
[3] The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
[4] Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
[5] For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
[6] Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
[7] The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
[8] The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Beattie (2000), no correction for clustering and multiple comparisons were needed.
[9] This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Meta-Analysis procedures

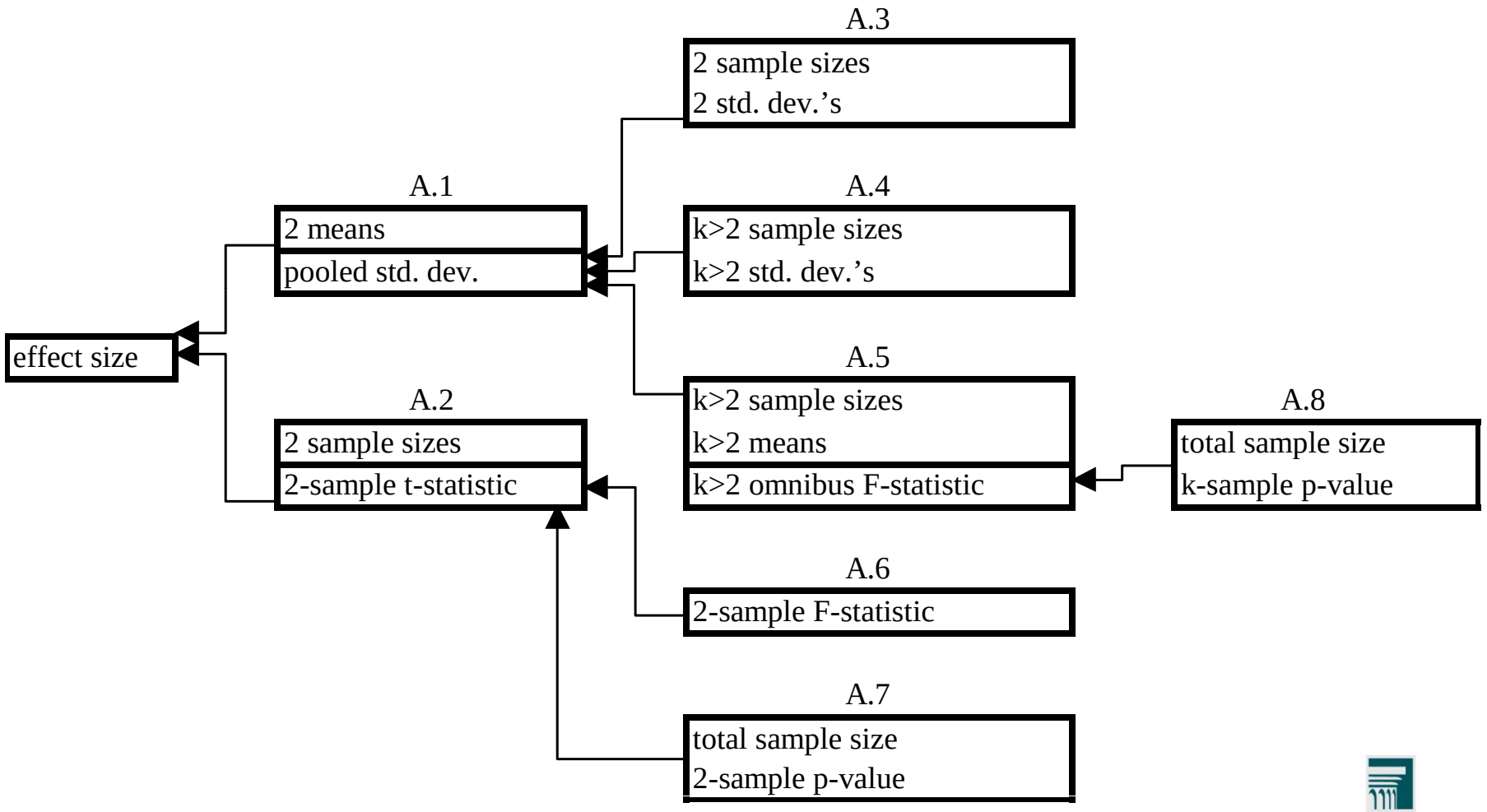
- Effect Sizes
- Aggregation Method
- Testing for Homogeneity
- Fixed and Random Effects Models
- Moderator Analysis
  - ANOVA type
  - Regression type

Effect Size

(1) Effect size (Hedges & Olkin, 1985):

$$d = \frac{\bar{x}_E - \bar{x}_C}{\sqrt{\frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2}}}$$

Flowchart for calculation of effect size (Tobler et al., 2000)



Aggregation of Effect Sizes

(1) Effect size (Hedges):

$$d = \frac{\bar{x}_E - \bar{x}_C}{\sqrt{\frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2}}}$$

(2) Effect size variance:

$$\sigma^2 = \frac{1}{n_E} + \frac{1}{n_C} + \frac{d^2}{2(n_E + n_C)}$$

$$w = \frac{1}{SE^2}$$

Weight (w)= (Variance)$^{-1}$

(3) Weighted average effect size:

$$\text{WES} = \sum(w_i d_i) \Big/ \sum w_i$$

(4) Weighted average effect size variance:

$$\text{var}[\text{WES}] = \frac{1}{\sum w_i}$$

Meta-analysis of computer-assisted programs across WWC topic areas, reading outcomes

- Does the evidence in WWC reports indicate that computer-assisted programs increase student reading achievement?

Computer-assisted interventions

WWC Topic	Intervention	# of studies
Adolescent Literacy	Accelerated Reader	5
	Fast ForWord®	8
	Read 180	14
	Reading Plus®	1
	SuccessMaker®	3
Beginning Reading	Accelerated Reader/Reading Renaissance	2
	Auditory Discrimination in Depth®	2
	DaisyQuest	6
	Earobics	4
	Failure Free Reading	1
	Fast ForWord®	6
	Lexia Reading	5
	Read Naturally	3
	Read, Write & Type!™	1
	Voyager Universal Literacy System®	2
	Waterford Early Reading Program	1
English Language	Fast ForWord® Language	2
Learners	Read Naturally	1
Early Childhood	DaisyQuest	1
Education	Ready, Set, Leap!®	2
	Waterford Early Reading Level One™	1
	Words and Concepts	2
Total	**22**	**73**

Examples of computer-assisted programs

- *Earobics*® is interactive software that provides students in pre-K through third grade with individual, systematic instruction in early literacy skills as students interact with animated characters. The program builds children's skills in phonemic awareness, auditory processing, and phonics, as well as the cognitive and language skills required for comprehension.

Examples of computer-assisted programs

- ***Lexia Reading*** **is a computerized reading program that provides phonics instruction and gives students independent practice in basic reading skills.** ***Lexia Reading*** **is designed to supplement regular classroom instruction. It is designed to support skill development in the five areas of reading instruction identified by the National Reading Panel.**

Number of students and effect sizes by topic area

Topic Area	total #	n_exp	n_cntrl	n_effct
Adolescent Literacy	26970	12717	14253	59
Beginning Reading	2636	1339	1297	151
Early Childhood Education	910	447	463	39
English Language Learners	308	173	135	6
Total	30824	14676	16148	255

Computer-assisted programs, fixed effects

Topic Area	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Adolescent Literacy	31	0.09	0.01	0.07	0.11	7.34	0.00
Beginning Reading	33	0.26	0.04	0.18	0.34	6.52	0.00
Early Childhood Education	6	0.12	0.07	-0.01	0.25	1.74	0.14
English Language Learners	3	0.24	0.12	-0.02	0.50	2.03	0.18

Homogeneity Testing

- Homogeneity analysis tests whether the assumption that all of the effect sizes are estimating the same population mean is a reasonable assumption.
- If homogeneity is rejected, the distribution of effect sizes is assumed to be heterogeneous.

Tests for Homogeneity of Weighted Effect Sizes by Topic Area

Computer-assisted programs					
Topic	n	M	Q_{within}	$Q_{critical}$[a]	Homogeneity
Adolescent Literacy	31	0.09	75.63	43.77	rejected
Beginning Reading	33	0.26	61.07	46.19	rejected
Early Childhood Education	6	0.12	1.21	11.07	not rejected
English Language Learners	3	0.24	9.28	5.99	rejected
[a] p=0.05 significance level					

Random versus Fixed Effects Models

- **Fixed effects model assume:**

  (1) there is one true population effect that all studies are estimating

  (2) all of the variability between effect sizes is due to sampling error

- **Random effects model assume:**

  (1) there are multiple (i.e., a distribution) of population effects that the studies are estimating

  (2) variability between effect sizes is due to sampling error + variability in the population of effects (Lipsey and Wilson, 2001)

Random Effects Model weights

- **Fixed effects model weights each study by the inverse of the sampling variance.**

$$w_i = \frac{1}{se_i^2}$$

- **Random effects model weights each study by the inverse of the sampling variance plus a constant that represents the variability across the population effects (Lipsey & Wilson, 2001).**

$$w_i = \frac{1}{se_i^2 + \hat{v}_\theta}$$

This is the random effects variance component.

Computer-assisted programs, random effects

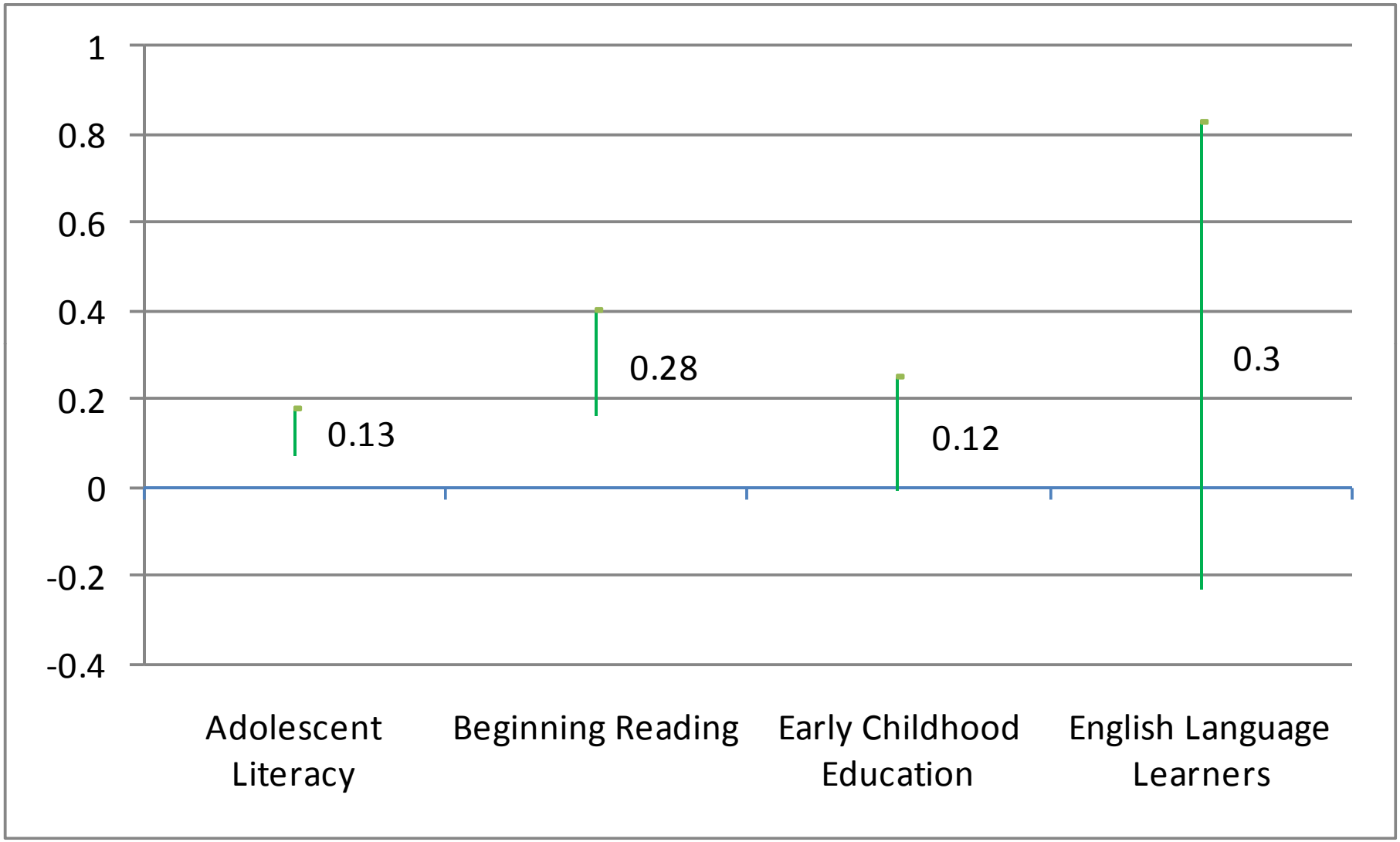
		Computer-Assisted Programs					
Topic Area	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Adolescent Literacy	31	0.13	0.03	0.07	0.18	4.56	0.00
Beginning Reading	33	0.28	0.06	0.16	0.40	4.71	0.00
English Language Learners	3	0.30	0.27	-0.23	0.83	1.11	0.38

Computer-assisted programs, random and fixed effects

		Computer-Assisted Programs					
Topic Area	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Adolescent Literacy	31	0.13	0.03	0.07	0.18	4.56	0.00
Beginning Reading	33	0.28	0.06	0.16	0.40	4.71	0.00
English Language Learners	3	0.30	0.27	-0.23	0.83	1.11	0.38

Topic Area	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Adolescent Literacy	31	0.09	0.01	0.07	0.11	7.34	0.00
Beginning Reading	33	0.26	0.04	0.18	0.34	6.52	0.00
Early Childhood Education	6	0.12	0.07	-0.01	0.25	1.74	0.14
English Language Learners	3	0.24	0.12	-0.02	0.50	2.03	0.18

Computer-assisted reading interventions, topic area effects and 95% CIs



Meta-analysis of computer-assisted programs within Beginning Reading topic area

- Are computer-assisted reading programs more effective than non-computer reading programs in improving student reading achievement?

Number of students and effect sizes by type of program: BR topic area

	Beginning Reading			
Type of Program	total #	n_exp	n_cntrl	n_effct
BR Computer Programs	2636	1339	1297	151
Other BR Programs	7591	4042	3549	224
Total Beginning Reading	10227	5381	4846	375

Beginning Reading Topic Area

Program type	Intervention	Number of studies
Computer-Assisted	Accelerated Reader/Reading Renaissance	2
Programs	Auditory Discrimination in Depth® / Lindamood Phonemic	2
	DaisyQuest	6
	Earobics	4
	Failure Free Reading	1
	Fast ForWord®	6
	Lexia Reading	5
	Read Naturally	3
	Read, Write & Type!™	1
	Voyager Universal Literacy System®	2
	Waterford Early Reading Program	1
Other BR Programs	Cooperative Integrated Reading and Composition©	2
	Corrective Reading	1
	Classwide Peer Tutoring©	1
	Early Intervention in Reading (EIR)®	1
	Fluency Formula™	1
	Kaplan Spell, Read, PAT	2
	Ladders to Literacy	3
	Little Books	3
	Peer-Assisted Learning Strategies (PALS)©	5
	Reading Recovery®	5
	Sound Partners	7
	Success for All	12
	Start Making a Reader Today® (SMART®)	1
	Stepping Stones to Literacy	2
	Wilson Reading	1
Total	**26**	**80**

Other reading programs

- ***Reading Recovery*® is a short-term tutoring intervention intended to serve the lowest-achieving first-grade students. The goals of *Reading Recovery*® are to promote literacy skills, reduce the number of first-grade students who are struggling to read, and prevent long-term reading difficulties. *Reading Recovery*® supplements classroom teaching with one-to-one tutoring sessions, generally conducted as pull-out sessions during the school day.**

Beginning Reading programs, fixed effects

Type of Program	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Computer-assisted programs	33	0.26	0.04	0.18	0.34	6.50	0.000
Othe BR programs	47	0.34	0.02	0.29	0.39	14.35	0.000
Beginning Reading Total	80	0.32	0.02	0.28	0.36	15.65	0.000

Tests for Homogeneity of Weighted Effect Sizes by Type of Program, BR

	Beginning Reading				
Type of Program	n	M	Q_{within}	$Q_{critical}$ [a]	Homogeneity
Beginning Reading, Total	80	0.31	166.23	101.90	rejected
BR Computer Programs	33	0.26	61.07	46.19	rejected
Other BR Programs	47	0.34	101.93	63.20	rejected
[a] p=0.05 significance level					

Beginning Reading programs, random effects

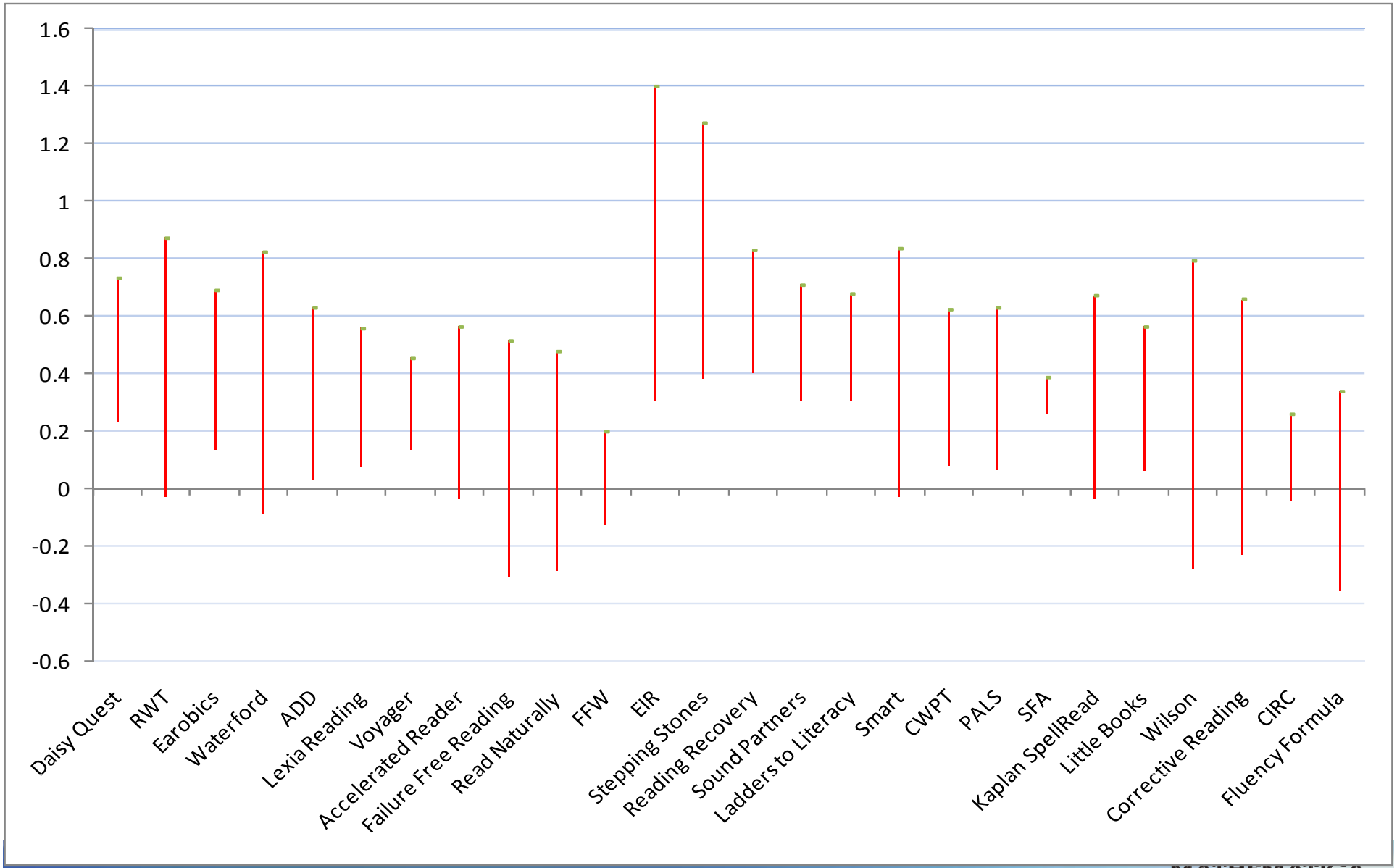
	Beginning Reading Topic Area						
Type of Program	n	M	SE	95% L	95% U	Z-value	P-value
Computer-assisted programs	33	0.28	0.06	0.16	0.40	4.71	0.000
Othe BR programs	47	0.39	0.04	0.32	0.47	9.84	0.000
Beginning Reading Total	80	0.35	0.03	0.29	0.42	10.65	0.000

Beginning Reading programs, random and fixed effects

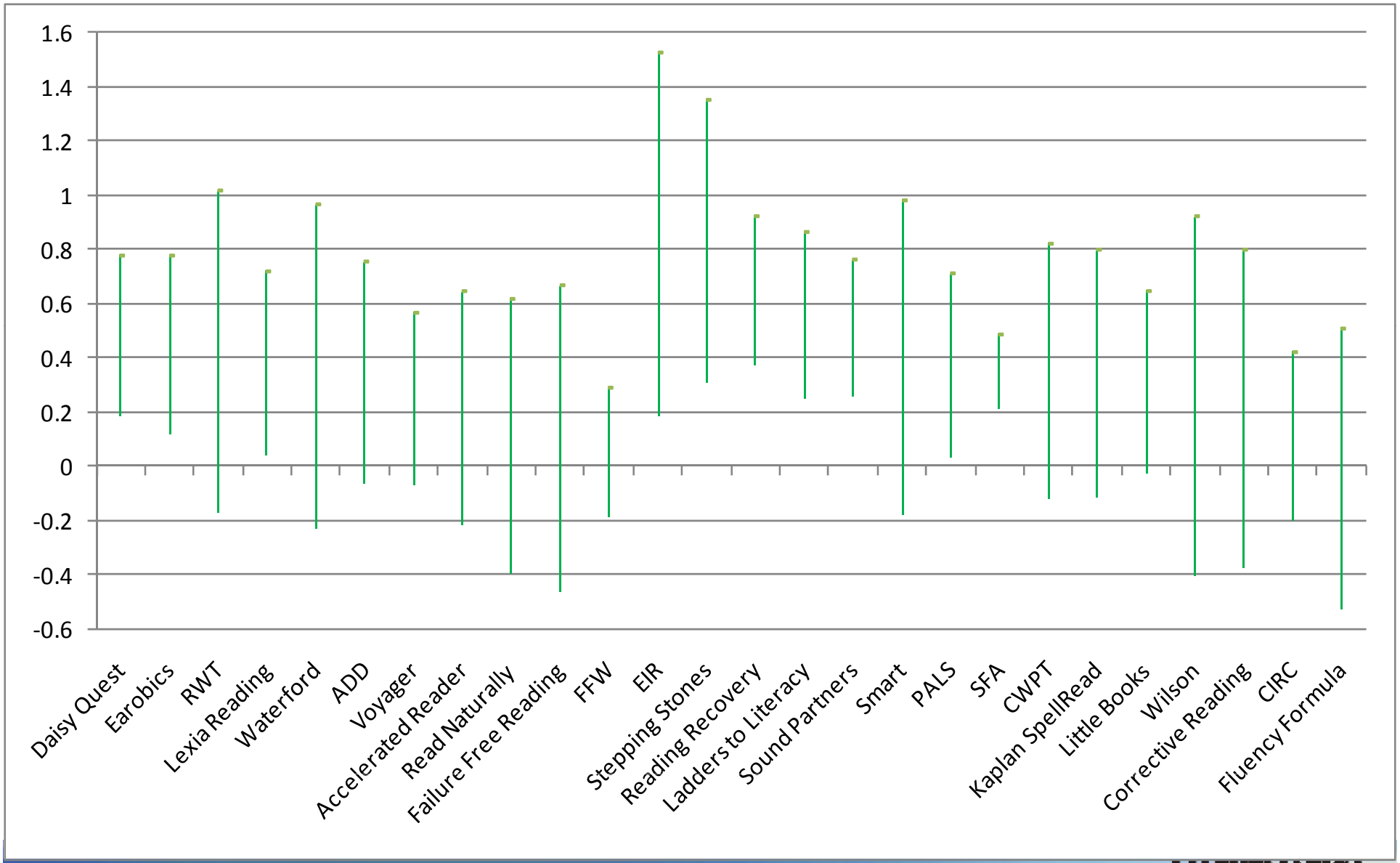
	Beginning Reading Topic Area						
Type of Program	n	M	SE	95% L	95% U	Z-value	P-value
Computer-assisted programs	33	0.28	0.06	0.16	0.40	4.71	0.000
Othe BR programs	47	0.39	0.04	0.32	0.47	9.84	0.000
Beginning Reading Total	80	0.35	0.03	0.29	0.42	10.65	0.000

Type of Program	n	M	Standard Error	95% Lower	95% Upper	Z-value	P-value
Computer-assisted programs	33	0.26	0.04	0.18	0.34	6.50	0.000
Othe BR programs	47	0.34	0.02	0.29	0.39	14.35	0.000
Beginning Reading Total	80	0.32	0.02	0.28	0.36	15.65	0.000

Beginning Reading Interventions, Fixed Effects, 95% Confidence Intervals



Beginning Reading Interventions, Random Effects, 95% Confidence Intervals



Moderator Analysis, random effects

Modeling between study variability:

- **Categorical models (analogous to a one-way ANOVA)**
- **Regression models (continuous variables and/or multiple variables with weighted multiple regression)**

Categorical analysis: moderators of program effectiveness

- Population
- Design
- Sample size
- Control group
- Reading domain

Weighted mean Effect Sizes for moderators: 80 studies, Beginning Reading, random effects

Study Characteristics	Overall			Computer-assisted			Other		
	n	M	SE	n	M	SE	n	M	SE
Type of Population[a]									
Universal	30	0.30	0.05	8	0.22	0.12	22	0.32	0.05
At Risk (struggling readers)	54	0.39	0.04	25	0.30	0.07	29	0.47	0.05
Evaluation Design									
Random	46	0.35	0.05	24	0.34	0.07	22	0.36	0.06
Non-Random	34	0.36	0.05	9	0.15	0.11	25	0.42	0.05
Sample Size									
Small	46	**0.48**	0.05	24	**0.39**	0.07	22	**0.56**	0.06
Large	34	**0.27**	0.04	9	**0.13**	0.09	25	**0.31**	0.04

[a]Sum of programs is greater than 80 because some programs collected data for multiple subgroups

Weighted mean Effect Sizes for moderators: 80 studies, Beginning Reading, random effects

Study Characteristics	Overall			Computer-assisted			Other		
	n	M	SE	n	M	SE	n	M	SE
Type of Control Group									
Business as usual	68	**0.39**	0.04	25	0.31	0.07	43	**0.42**	0.04
Other program/intervention	12	**0.17**	0.08	8	0.19	0.12	4	**0.14**	0.12
Domain[b]									
Alphabetics	57	0.44	0.04	25	**0.38**	0.07	32	0.48	0.05
Fluency	25	0.36	0.07	6	**0.16**	0.15	19	0.42	0.08
Comprehension	41	0.16	0.05	13	**0.02**	0.09	28	0.22	0.05
General Reading	22	0.41	0.06	2	**0.30**	0.19	20	0.42	0.06

[b]Sum is greater than 80 because programs collected data for multiple domains

Dummy Variables for Regressions

Variables		
	Random	Non-random
Design	1	0
	Buisness-as-usual	Other program
Control group	1	0
	Computer-assisted	Other BR programs
Computer-Assisted Programs	1	0

Regression Statistics for BR Programs, Random effects

$$ES_i = \beta_0 + \beta_1 C_i + \varepsilon_i$$

Variable	Coefficient	Standard Error	- 95% CI	+95% CI	Z-statistic	P-value
Constant	0.40	0.04	0.32	0.48	9.61	0.000
Computer-assisted programs	-0.12	0.07	-0.26	0.20	-1.72	0.084
Note: Q (model)=2.97, df=1, p=0.084						
Test for homogeneity: Q(error)=90.60, df=78, p=0.156						
v=0.037						

Regression Statistics for BR Programs, Random effects

Variable	Coefficient	Standard Error	- 95% CI	+95% CI	Z-statistic	P-value
Constant	0.40	0.04	0.32	0.48	9.61	0.000
Computer-assisted programs	-0.12	0.07	-0.26	0.20	-1.72	0.084
Note: Q (model)=2.97, df=1, p=0.084						
Test for homogeneity: Q(error)=90.60, df=78, p=0.156						
v=0.037						

	Beginning Reading Topic Area						
Type of Program	n	M	SE	95% L	95% U	Z-value	P-value
Computer-assisted programs	33	0.28	0.06	0.16	0.40	4.71	0.000
Othe BR programs	47	0.39	0.04	0.32	0.47	9.84	0.000
Beginning Reading Total	80	0.35	0.03	0.29	0.42	10.65	0.000

Regression Statistics for BR Programs, Random Effects

$$ES_i = \beta_0 + \beta_1 C_i + \beta_2 LnW_i + \beta_3 D_i + \beta_4 CG_i + \varepsilon_i$$

Variable	Coefficient	Standard Error	- 95% CI	+95% CI	Z-statistic	P-value
Constant	0.70	0.17	0.38	1.03	4.26	0.000
Computer-Assisted Programs	-0.14	0.07	-0.28	-0.001	-1.97	0.049
Program Size (Ln Weight)	-0.13	0.04	-0.20	-0.06	-3.59	0.000
Design	-0.06	0.07	-0.19	0.08	-0.86	0.393
Control group	0.20	0.09	0.03	0.38	2.24	0.025
Note: Q (model)=20.86, df=4, p=0.000						
Test for homogeneity: Q(error)=79.64, df=75, p=0.335						

Meta-Analytic Multiple Regression Results From the Wilson/Lipsey SPSS Macro

```
 *****  Inverse Variance Weighted Regression  *****

 *****  Random Intercept, Fixed Slopes Model  *****

------- Descriptives -------
      Mean ES     R-Square            k
        .3510        .2076      80.0000

------- Homogeneity Analysis -------
                     Q           df            p
Model          20.8631       4.0000        .0003
Residual       79.6431      75.0000        .3351
Total         100.5062      79.0000        .0517

------- Regression Coefficients -------
                 B       SE    -95% CI  +95% CI       Z         P     Beta
Constant      .7038    .1651     .3802   1.0273   4.2630    .0000    .0000
Program size -.1324    .0368    -.2046   -.0601  -3.5920    .0003   -.3852
Computer     -.1418    .0720    -.2829   -.0006  -1.9686    .0490   -.2119
Design       -.0585    .0685    -.1927    .0758   -.8537    .3933   -.0920
Cntrl group   .2036    .0909     .0253    .3818   2.2386    .0252    .2284

------- Method of Moments Random Effects Variance Component -------
v      =    .03056
```

Conclusions

- **The present work appears to lend some support to the proposition that computer-assisted interventions in reading are effective. For example, the average effect for beginning reading computer-based programs is positive and substantively important (that is >0.25).**
- **For the Beginning Reading topic area, the effect appears smaller than the effect achieved by non-computer reading programs.**

AMERICAN INSTITUTES FOR RESEARCH®

MATHEMATICA Policy Research, Inc.

References

- Borenstein, M., Hedges, L.V., Higgins, J.P., and Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons.

- Hedges, L. V. and Olkin I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.

- Lipsey, M.W., & Wilson, D.B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.

- Tobler, N.S., Roona, M.R., Ochshorn, P., Marshall, D.G., Streke, A.V., & Stackpole, K.M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *Journal of Primary Prevention*, *20*(4), 275-336.

For More Information

- **Please contact:**
  - **Andrei Streke**
    - AStreke@mathematica-mpr.com
  - **Tsze Chan**
    - TChan@air.org

AMERICAN INSTITUTES FOR RESEARCH®

MATHEMATICA Policy Research, Inc.