The Effectiveness of Cognitive Principles in Authentic Education Settings: Research to Practice

Yvonne Kao, WestEd (ykao@wested.org)[1]

Jodi Davenport, WestEd (jdavenp@wested.org)

Bryan Matlen, WestEd (bmatlen@wested.org)

Larry Thomas, WestEd (lthomas@wested.org)

Steve Schneider, WestEd (sschnei@wested.org)

Background

Lab-based research in cognitive and learning sciences provide many recommendations for improving learning and instruction (e.g., Pashler et al., 2007). Tightly controlled experiments demonstrate that learning can be enhanced with strategies such as: 1) facilitating mapping between visual representations, 2) prompting for explanation of worked examples, 3) using quizzing to promote learning, and 4) spacing practice opportunities over time (e.g., Cooper & Sweller (1987); Cepeda, Pashler, Vul, et al. 2006; Clark & Mayer, 2003; Larkin & Simon, 1987; Mayer 2001; Kalyuga, Chandler, and Sweller (2001); Paas & Van Merrienboer, 1994; Rohrer & Taylor, 2006). However, few studies investigate the principles in combination, or in authentic learning environments. We describe a large-scale effort to bridge research and practice by applying four cognitive principles to redesign the 7th grade *Connected Mathematics Project 2* curriculum and testing the efficacy of these revised materials. Figure 1 shows a side-by-side of a page from the original and the revised curricula.

Research Design and Research Questions

In a large-scale, two-year, cluster-randomized trial schools, were randomly assigned to receive either the redesigned 7th grade CMP2 curriculum or original materials. During the first year, treatment teachers became familiar with the redesigned curriculum and accompanying changes to practice, all teachers familiarized themselves with study requirements, and researcher-created unit tests were field tested. Impact on students' math achievement was measured in the second year.

Students' math achievement was measured using the Mathematics Diagnostic Testing Project (MDTP) pre-algebra readiness assessment (pre-test and post-test) and post-tests for each of the eight units in the curriculum[2]. The study addressed the following research questions:

1. Do 7th grade students receiving the redesigned curriculum (treatment) show greater learning than students receiving the original curriculum (control)?
2. Does the effect of the redesigned curriculum differ for traditionally lower-performing students?

[1] Presenting author.

[2] One unit is not reported due to substantially high attrition.

Participants

Study participants were recruited from public and charter schools in the United States. All participating teachers had at least one year of experience teaching the original CMP2 curriculum and planned to use CMP2 as their core 7th grade math curriculum for the two years of the study. There was high study attrition in the first study year, so a second cohort was recruited. In all, 114 schools (N_{Treat} = 59, $N_{Control}$ = 55) in 22 states were enrolled in the study, with 181 participating teachers at these schools (N_{Treat} = 93, $N_{Control}$ = 88). Data were collected from 2,465 students (N_{Treat} = 1,222, $N_{Control}$ = 1,243) in the impact year.

Data Analysis

The results reported in this paper use multidimensional item response theory (MIRT) to estimate student math achievement on the outcome measures. Specifically, the study used a multilevel extension of the Rasch model (Rasch, 1960) for dichotomous items and the partial credit model (Masters, 1982) for open-ended items. These MIRT models were applied to each assessment to estimate expected *a posteriori* (EAP) student pre- and post-test scores. These resulting EAP scores were then used in additional impact analyses using a two-level hierarchical linear model (HLM).

The level 1 (student and teacher) and level 2 (school) HLM models are presented below:

Level 1:

$$OUTCOME_{ij} = \pi_{0j} + \pi_{1j}(PRE)_{ij} + \pi_{2j}(GEND)_{ij} + \pi_{3j}(ELL)_{ij} + \pi_{4j}(SPED)_{ij} + \pi_{5j}(STEM)_{ij} + \pi_{6j}(COHORT)_{ij} + \pi_{7j}(DEGREE)_{ij} + \pi_{8j}(MKT)_{ij} + \pi_{9j}(EXP)_{ij} + \epsilon_{ij}$$

Level 2:

$$\pi_{0j} = \beta_{00} + \beta_{01}(TREAT)_j + \beta_{02}(LOCALE)_j + \beta_{03}(MATH)_j + \beta_{04}(FRL)_j + u_{0j}$$

The $OUTCOME_{ij}$ variable in the level-1 model represents students' outcome (i.e., MDTP score or unit test scores) for the i-th student in the j-th school. PRE corresponds to students' MDTP pre-test score. Other level-1 covariates represent the students' gender (GEND), English language learner status (ELL), special education status (SPED), and underrepresented ethnicity in STEM status (STEM). Teacher variables consisted of each student's teacher cohort (COHORT), advanced degree status (DEGREE), score on a baseline Mathematics Knowledge for Teaching (MKT) assessment (Schilling & Hill, 2007; Schilling, Blunk, & Hill, 2007), and years of teaching experience (EXP).

Level-2 covariates included school treatment status (TREAT), school locale (LOCALE), percentage of students who were proficient on the state standardized math test (MATH), and percentage of students who qualified for free- or reduced-price lunch (FRL). ϵ_{ij} and u_{0j} are student and school residuals, respectively. All covariates were grand-mean centered.

Results

Results of the overall average treatment impact are summarized in Table 1. Across the eight outcomes evaluated, treatment effects were positive in favor of students who received the redesigned CMP2 units. However, only Unit 2 and 3 showed a statistically significant positive effect. The effect size estimate for the summative MDTP assessment was 0.12, which is considered small, but within-expectation for studies of this type (Cheung & Slavin, 2015). The average effect size for the more specialized unit assessments was .26 (range = .08 - .49), which is considered substantively important positive effect, regardless of statistical significance (WWC, 2014).

To evaluate research question 2, cross-level treatment and student covariate interactions were included in the above model. Following procedures by Bauer and Curran (2005) and Tate (2004), Figure 2 shows the treatment-by-pre-test interaction effects. The treatment was more effective for lower-performing students on four of the eight outcomes. Only Unit 7 showed a statistically significant interaction.

The treatment tended to be more effective for students who are members of traditionally underperforming subgroups in mathematics: females, English language learners, special education students and underrepresented ethnicities in mathematics (see Table 2). However, the interactions are not consistently statistically significant.

Conclusions

Overall, the redesigned curriculum exhibited trends for positive impact. The lack of statistical significance may be due in part to low power as a result of high study attrition. Moderator analyses showed variable effects, but when interaction effects were at least marginally significant, they suggested that the treatment consistently favored the traditionally underperforming subgroup.

These findings suggest that purposefully engineering curricula based on learning sciences research is productive and may increase equity in educational outcomes. At the same time, the observed effects from this study were small and variable. Future analyses will explore how the content of individual units and teacher practice may have mediated learning outcomes.

References

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40*, 373–400.

Cheung, A., & Slavin, R.E. (2015, September). How methodological features affect effect sizes in education. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. Source: http://www.bestevidence.org/methods/methods.html

Clark, R.C., and Mayer, R.E. (2003). e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia Learning. San Francisco: Jossey-Bass.

Cooper, G., and Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. Journal of Educational Psychology, 79, 347–362.

Kalyuga, S., Chandler, P., and Sweller, J. (2001). Learner experience and efficiency of instructional guidance. Educational Psychology, 21, 5–23.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, *11*(1), 65-100.

Masters, G. N. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47, 149-174.

Mayer, R.E. (2001). Multimedia learning. New York: Cambridge University Press.

Paas, F., and van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problemsolving skills: A cognitive-load approach. Journal of Educational Psychology, 86, 122–133.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., and Metcalfe, J. (2007) *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen and Lydicke.

Rohrer, D., and Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. Applied Cognitive Psychology, 20, 1209-1224.

Schilling, S.G. & Hill, H.C. (2007). Assessing Measures of Mathematical Knowledge for Teaching: A Validity Argument Approach. *Measurement: Interdisciplinary Research and Perspectives* (5), 2-3, 70-80.

Schilling, S.G., Blunk, M. & Hill, H.C. (2007). Test Validation and the MKT Measures: Generalizations and Conclusions. *Measurement: Interdisciplinary Research and Perspectives* (5), 2-3, 118-127.

Tate, R. L. (2004). Interpreting hierarchical linear and hierarchical generalized linear models with slopes as outcomes. *The Journal of Experimental Education, 73*, 71–95.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, March). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0).* Retrieved from http://whatworks.ed.gov

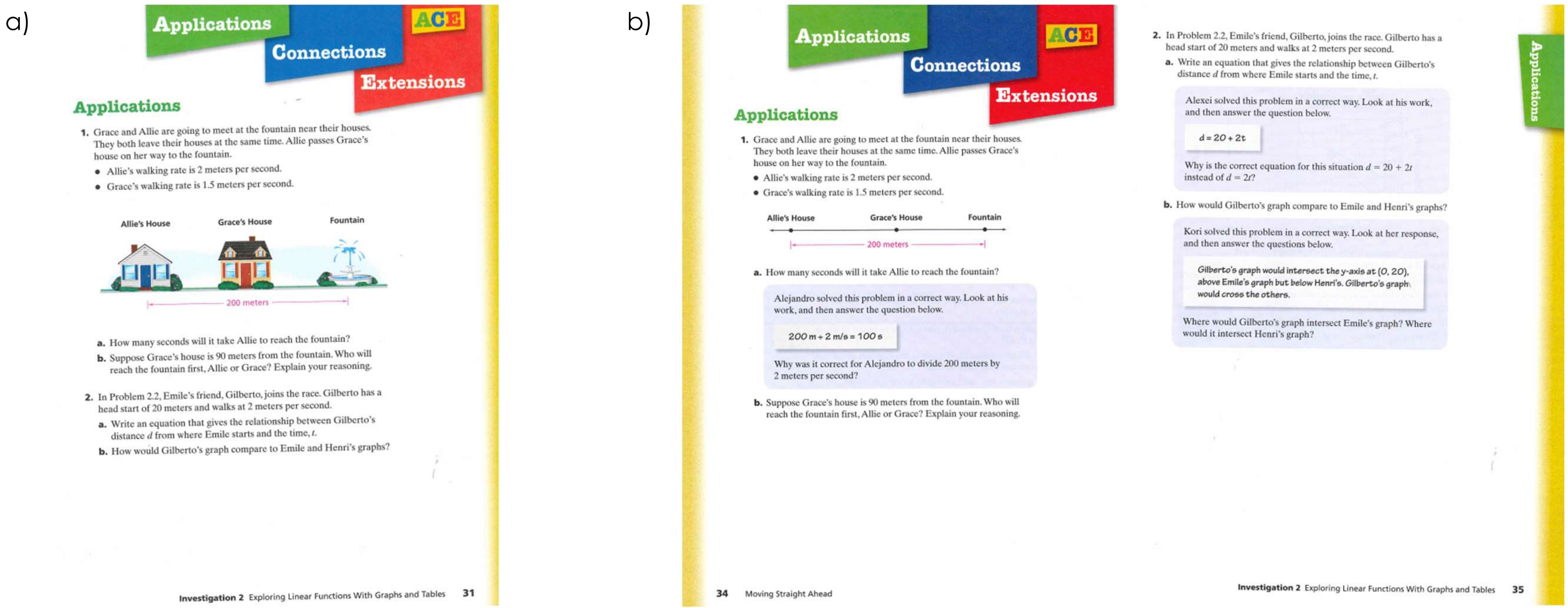

Figure 1. A page from the original CMP2 curriculum (a) and the redesigned version (b). The redesigned version contains a revised figure and introduces worked examples into the problems.

Table 1. Four of eight outcome measures showed substantively important effect sizes in favor of the redesigned curriculum (shaded rows). This effect was statistically significant for Unit 3: Comparing and Scaling.

				Effect Size	
Outcome Measure	**Coefficient**	**Standard Error**	***p***	**Estimate**	**95% CI**
MDTP Post-test	0.11	0.12	0.36	0.12	0.01 - 0.23
Unit 1: Variables and Patterns	0.06	0.14	0.66	0.08	-0.07 - 0.22
Unit 2: Stretching and Shrinking	0.41	0.23	0.08⁺	0.36˚	0.24 - 0.48
Unit 3: Comparing and Scaling	0.35	0.14	0.02*	0.38˚	0.26 – 0.49
Unit 4: Accentuate the Negative	0.12	0.11	0.31	0.17	0.05 – 0.28
Unit 5: Moving Straight Ahead	0.07	0.15	0.63	0.09	-0.03 - 0.21
Unit 6: Filling and Wrapping	0.47	0.29	0.12	0.49˚	0.33 – 0.65
Unit 7: What Do You Expect?	0.15	0.15	0.32	0.26˚	0.12 – 0.40

Note. ⁺ indicates $p < 0.1$, * indicates $p < 0.05$, and ˚ indicates a substantively important effect size

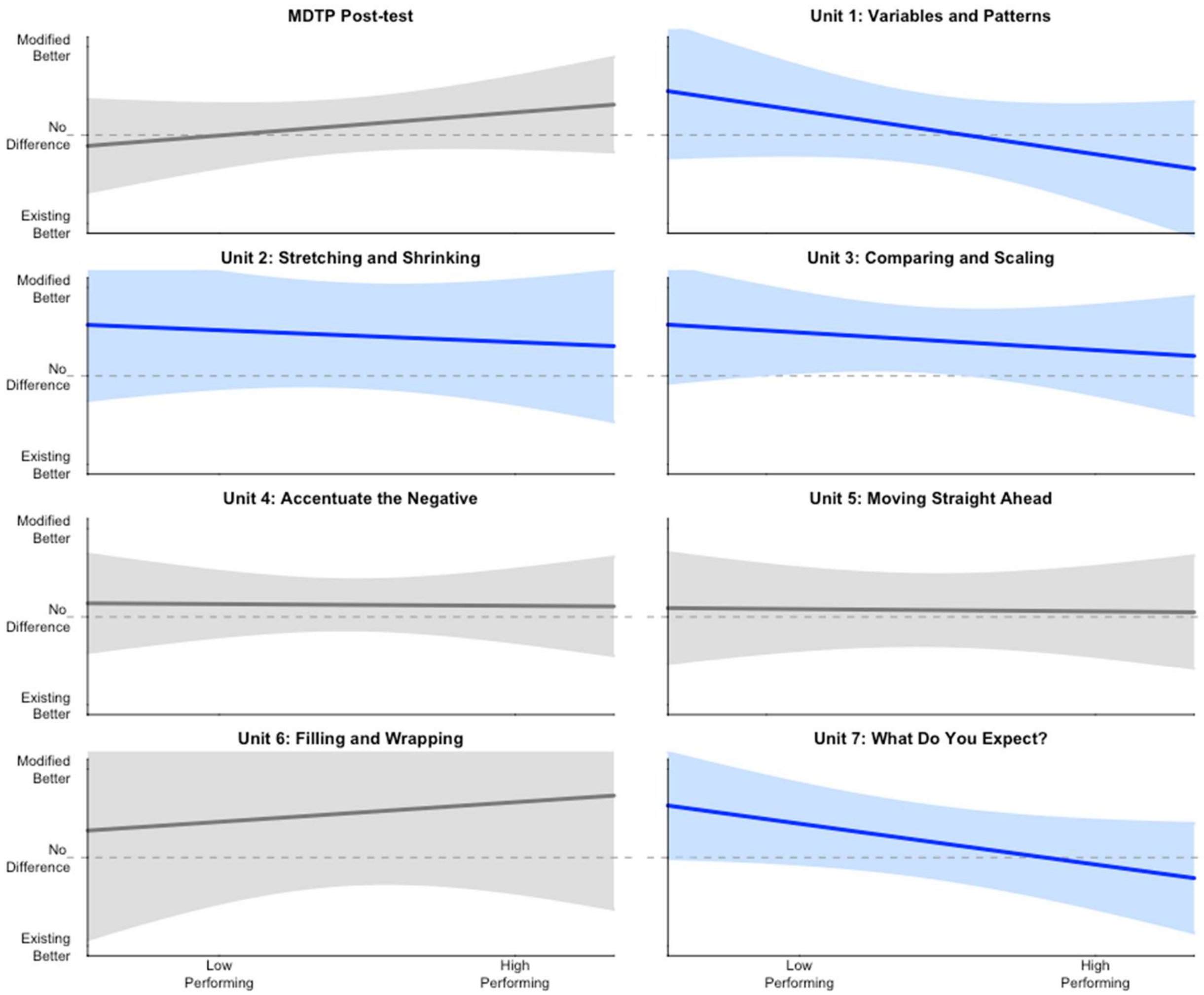

Figure 2. Summary of treatment-by-pre-test interaction effects by outcome measure showing that low performing students benefited more from the modified curriculum in four of the eight outcomes. Bolded lines show the effect of the treatment for students at different performance levels on the pre-test. Shaded region represent 95% confidence bands.

Table 2. The treatment tended to be more effective for students in traditionally underperforming subgroups (shaded rows). This interaction was statistically significant for special education students studying Unit 3: Comparing and Scaling.

Moderator	**Outcome Measure**	**Coefficient**	**Standard Error**	***p***
Gender	MDTP Post-test	0.04	0.05	0.43
	Unit 1: Variables and Patterns	-0.03	0.07	0.68
	Unit 2: Stretching and Shrinking	0.04	0.08	0.57
	Unit 3: Comparing and Scaling	-0.08	0.07	0.22
	Unit 4: Accentuate the Negative	0.02	0.06	0.71
	Unit 5: Moving Straight Ahead	0.10	0.06	0.07+
	Unit 6: Filling and Wrapping	0.00	0.10	1.00
	Unit 7: What Do You Expect?	0.01	0.05	0.91
English language learner status	MDTP Post-test	0.11	0.17	0.54
	Unit 1: Variables and Patterns	0.06	0.27	0.83
	Unit 2: Stretching and Shrinking	0.03	0.24	0.90
	Unit 3: Comparing and Scaling	-0.24	0.22	0.28
	Unit 4: Accentuate the Negative	0.30	0.19	0.11
	Unit 5: Moving Straight Ahead	0.15	0.18	0.39
	Unit 6: Filling and Wrapping	0.12	0.39	0.75
	Unit 7: What Do You Expect?	-0.13	0.15	0.40
Special education status	MDTP Post-test	0.03	0.12	0.83
	Unit 1: Variables and Patterns	-0.08	0.14	0.58
	Unit 2: Stretching and Shrinking	0.12	0.16	0.44
	Unit 3: Comparing and Scaling	0.39	0.15	0.01*
	Unit 4: Accentuate the Negative	0.04	0.12	0.74
	Unit 5: Moving Straight Ahead	0.20	0.13	0.12
	Unit 6: Filling and Wrapping	-0.05	0.22	0.80
	Unit 7: What Do You Expect?	-0.1	0.11	0.34
Underrepresented ethnicity in STEM (i.e., not White or Asian)	MDTP Post-test	0.07	0.07	0.34
	Unit 1: Variables and Patterns	0.02	0.11	0.89
	Unit 2: Stretching and Shrinking	0.08	0.10	0.47
	Unit 3: Comparing and Scaling	-0.05	0.09	0.59
	Unit 4: Accentuate the Negative	-0.06	0.08	0.46
	Unit 5: Moving Straight Ahead	0.14	0.08	0.07+
	Unit 6: Filling and Wrapping	0.13	0.13	0.30
	Unit 7: What Do You Expect?	0.09	0.07	0.20

Note. + indicates $p < 0.1$, * indicates $p < 0.05$.