CRESST REPORT 792

Julia Phelan
Terry Vendlinski
Kilchan Choi
Yunyun Dai
Joan Herman
Eva L. Baker

THE DEVELOPMENT AND IMPACT OF POWERSOURCE©: YEAR 5

MAY, 2011

The National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Sciences
UCLA | University of California, Los Angeles

The Development and Impact of POWERSOURCE©: Year 5

CRESST Report 792

Julia Phelan, Terry Vendlinski, Kilchan Choi, Yunyun Dai, Joan Herman, and Eva L. Baker
CRESST/University of California, Los Angeles

May, 2011

CRESST/University of California, Los Angeles
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number Award Number R305A050004, as administered by the Institute of Education Sciences, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

To cite from this report, please use the following as your APA reference: Phelan, J., Vendlinski, T., Choi, K., Dai, Y., Herman, J., & Baker, E.L. (2011). *The development and impact of POWERSOURCE©: Year 5.* (CRESST Report 792). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

EXECUTIVE SUMMARY

POWERSOURCE© Background and Rationale

The POWERSOURCE© intervention is intended as a generalizable and powerful formative assessment-based strategy that can be integrated with any ongoing mathematics curriculum to improve teachers' knowledge and practice and, in turn, student learning. Combining theory and research in cognition, assessment and learning with design elements to support the transformation of practice within existing constraints, POWERSOURCE© includes both a system of learning-based assessments and an infrastructure to support teachers' use of those assessments to improve student learning.

The current study focuses on middle school mathematics, starting in Grade 6, and on helping to assure that students possess key understandings they need for success in Algebra I. Our primary research objectives are based on our hypotheses that as a result of POWERSOURCE©, teachers will become more proficient in their subject matter knowledge, more skilled in their formative use of assessment, and better focus their instruction on key ideas; in turn, teachers will be more effective in helping students to improve their understanding, as shown by measures of student learning.

A striking innovation in POWERSOURCE© is its targeting of the big ideas (fundamental concepts and principles) and their interrelationships, which underlie and define a field of knowledge rather than treat specific concepts and topics in isolation—as do traditionally developed tests. The POWERSOURCE© intervention targets big ideas and related skills in four domains that underlie success in Algebra I: a) rational number equivalence (RNE); b) properties of arithmetic (PA); c) principles for solving linear equations (SE); and d) application of core principles in these domains to other critical areas of mathematics, such as geometry and probability (RA). These domains were chosen because of their significant place in mathematics standards across grades 6-8 as well as their importance to later mastery of algebra.

In each domain we have designed a series of short POWERSOURCE© assessments comprised of multiple item types, which are called *Checks for Understanding*, to help teachers assess their students' understanding of basic mathematical principles, to connect their instruction, and to provide feedback to support deeper understanding. A set of instructional resources and targeted professional development activities were also developed for each of these domains. POWERSOURCE© materials are designed to complement existing curricula.

POWERSOURCE© Implementation Study 2008-09

As described in previous reports (Baker, 2007; Baker, 2008), the core undertaking of our work during the 2008-09 school year was to continue with an extended, random assignment implementation study of the POWERSOURCE© program. In the 2008-09 study, we expanded the intervention from only Grade 6 to Grades 6 and 7 (in all participating schools). As in prior years, new teachers were randomly assigned to either POWERSOURCE© or control conditions, with the ultimate goal of determining program impact on both students and teacher learning outcomes. Teachers that continued in the study for another year maintained their prior year's group status. The 2008-09 study was almost identical to the previous year's work, with a few minor changes:

- An interim transfer measure was developed for use in Grade 6.
- We created Grade 7 teacher instructional materials and *Checks for Understanding* assessments.
- We modified the professional development sessions (in Grade 6) to focus more on interpreting student assessment data and less on teaching the big ideas.
- We recruited an additional school district to replace a district that did not continue with the study.

Several supplementary strands of work were also completed as part of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) activities during the 2008-09 school year. The supplementary work included a validation study of teacher math knowledge measures; validation and use of teacher interview and observation protocols; investigation of district contexts for assessment; and lastly, international applications of the POWERSOURCE© work.

Implementation Study 2008-09: Student Outcomes

Six districts participated in the random assignment implementation study in 2008-09. As described earlier, we used two designs (within and between school) based on district needs and configuration. Ultimately, three of the districts used a within-school (W-S) design, where random assignment was accomplished within each school (i.e., a given school had both treatment and control teachers). Two districts used a between-school (B-S) design, where schools within a district were randomly assigned to treatment or control conditions[1].

Taking methodological concerns into account, we used a two-level hierarchical model (HM) to examine the POWERSOURCE© effects on the transfer measure outcome. In order to synthesize two different designs and compromise a unit of analysis issue, we chose teacher

[1] Note that one district adopted both W-S design and B-S design.

as a unit of analysis and individual school effects are also included in a model. School specific fixed effects take care of school blocking factors and intra-class correlation of school in a model. As such, we can examine whether there is a differential treatment effect depending upon two different designs not at the cost of losing statistical power.

Using the Transfer and Interim Measure Total Scores as Student Outcome

Results from the analyses indicated that:

- A short amount of targeted intervention on key mathematical principles has an impact on student performance on transfer measures of related content. The POWERSOURCE© intervention had more impact on the relatively higher-performing students than the lower-performing students. In both grades, on most of the student measures, those students with higher initial pretest (or interim transfer measure) scores tended to benefit more from the treatment when compared to students with lower pretest scores.
- On the Grade 6 transfer measure items related to rational number equivalence concepts, we saw a significant effect of POWERSOURCE©. In both designs, students in the POWERSOURCE© group outperformed control group students on items associated with rational number equivalence and the effect was larger as pretest scores increased.
- Item analyses indicated difficulty ranges on the Grade 6 RNE items between b= -1.54 and b = +1.5, with the range of all the items for all domains b = -1.8 to b = 2.24. Thus, the RNE transfer measure items were spread evenly across the measure in terms of difficulty.
- There were also significant effects of the POWERSOURCE© treatment seen for PA items on the Grade 6 interim transfer measure and transfer measure, when using the pretest as a covariate.
- In two cases we did see a main effect of design. Previously we saw no differences when we compared students in the B-S design treatment with the W-S design treatment. In this year, however, we saw a main effect of design on the Grade 6 interim transfer measure and also on the Grade 7 transfer measure items associated with solving equations. In both cases, scores for the B-S design were higher than for the W-S design.

Future Plans

Currently, we are starting to analyze data collected during the 2009-10 school year—including the student transfer measures, pretest measures, and the multiple teacher outcomes described in this paper. We will also analyze the *Checks for Understanding* that POWERSOURCE© group teachers completed. Moreover, we will examine the statistical quality of the items and track student scores across the school year. Lastly, we will analyze state test data outcomes as they are made available by the districts including (when available) subscale scores of state mathematics items.

The focus for project implementation during the 2009-10 school year continued the experimental (random assignment) study of POWERSOURCE©. In this year of the study, we expanded the intervention from Grades 6 and 7, to Grades 6-8, in all participating schools. As with prior years, new teachers were randomly assigned to either POWERSOURCE© or control conditions with the ultimate goal of determining program impact on both students and teacher learning outcomes. Teachers that continued in the study for another year maintained their prior year's group status. The 2009-10 study was almost identical to the previous year's work, with a few minor changes:

- An interim transfer measure was developed for use in Grade 7 (we had one in Grade 6 only in the previous year).
- We created Grade 8 teacher instructional materials and *Checks for Understanding* assessments.
- We modified the professional development sessions (in Grades 6 and 7) to focus more on interpreting student assessment data and less on teaching the big ideas.

We recruited an additional school district to replace a district that did not continue with the study begun in 2008-09. Specifically, in addition to continuing the study at the Grade 6 and 7 levels, we added Grade 8 teachers in the participating districts to the study (note that depending on district configuration, there may be some overlap in samples; for instance, in cases where the same teachers taught multiple grades of math). The study utilized a similar design and instrumentation to that described in the earlier text regarding the 2008-09 study—with student and teacher outcome instruments adapted to reflect Grade 8 content (as applicable).

TABLE OF CONTENTS

THE DEVELOPMENT AND IMPACT OF POWERSOURCE©: YEAR 5

Julia Phelan, Terry Vendlinski, Kilchan Choi, Yunyun Dai, Joan Herman, and Eva L. Baker
CRESST/ University of California, Los Angeles

Abstract

The POWERSOURCE© intervention is intended as a generalizable and powerful formative assessment strategy that can be integrated with any mathematics curriculum. POWERSOURCE© includes both a system of learning-based assessments and an infrastructure to support teachers' use of those assessments to improve student learning. The core undertaking of our work during the 2008-09 school year was to continue with an extended, random assignment implementation study of the POWERSOURCE© program. Results from our analyses indicated that a short amount of targeted intervention on key mathematical principles had a positive impact on student performance on transfer measures of related content. The POWERSOURCE© intervention had more impact on the relatively higher-performing students than the lower-performing students; that is, those students with higher initial pretest scores tended to benefit more from the treatment when compared to students with lower pretest scores.

Introduction

The POWERSOURCE© intervention is intended as a generalizable and powerful formative assessment strategy that can be integrated with any ongoing mathematics curriculum to improve teachers' knowledge and practice and, in turn, student learning. Combining theory and research in cognition, assessment and learning (for both adults and students) with design elements to support the transformation of practice within existing constraints, POWERSOURCE© includes both a system of learning-based assessments and an infrastructure to support teachers' use of those assessments to improve student learning.

The current study focuses on middle school mathematics, starting in Grade 6, and on helping to assure that students possess key understandings they need for success in Algebra I. Such a focus is motivated by ample research showing the frequency and price of failure for subsequent academic performance—including high school graduation, college preparation and entry (e.g., Brown & Niemi, 2007).

Our primary research objectives are based on our hypotheses that as a result of POWERSOURCE©, teachers will become more proficient in their subject matter knowledge; more skilled in their formative use of assessments; lastly, they will better focus their instruction on key ideas. As a result, teachers will be more effective in helping students improve their understanding—as shown by measures of student learning. Ultimately, we expect that improvements in student understanding will drive better performance on No Child Left Behind (NCLB, 2002) mandated state tests, transfer measures, and future coursework.

Research on Formative Assessment

The intervention builds on recent research showing formative assessment as a powerful strategy for improving learning (Black & Wiliam, 1998a, 1998b; Bloom, 1968; Kluger & DeNisi, 1996). For example, Black and Wiliam's (1998a) landmark meta-analysis, based on a review of 250 studies, found effect sizes that ranged between .4 and .7, and found particularly large effect sizes for low-achieving students, including students with learning disabilities (Black & Wiliam, 1998b). This finding makes intuitive sense—as one of the major functions of formative assessment is to determine where students are relative to learning goals and to use this information to provide feedback and/or make necessary instructional adjustments (such as re-teaching, trying alternative instructional approaches, or offering more opportunities for practice). If students have already mastered the content, there is little need for subsequent adjustment and little room for learning improvement.

Yet, even as research shows the rich potential of formative assessment, it also suggests the limits of current practice. The quality of increasingly popular interim or benchmark testing, marketed as formative assessments to districts and schools, is uneven—assessment tends to be an afterthought rather than a core, quality element of current curriculum materials (Herman & Baker, 2006; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Wolf, Bixby, Glenn, & Gardner, 1991). Moreover, educators often have limited background and capacity to develop or engage in quality assessment practices (Heritage & Yeagley, 2005; Herman & Gribbons, 2001; Plake & Impara, 1997; Shepard, 2001: Stiggins, 2005). For many teachers—current classroom assessment practices are almost exclusively summative (for instance, consisting of end-of-the-week, unit, or semester tests).

Students receive grades or scores on these assessments and their teachers, who have neither the time nor the curriculum resources to remediate deficiencies, move on, disconnecting the assessments from any active function in learning. Yet as Black and Wiliam (1998a, 1998b) note, assessments can only become formative when information from them is used immediately to inform teaching and for the benefit of student learning. Teacher subject matter knowledge offers yet another challenge—as research and our own experiences in assessment development with teacher and districts suggest that many teachers do not have subject area knowledge sufficiently deep to teach or assess mathematics effectively (Ball & Bass, 2001; Ball, Lubienski, & Mewborn, 2001).

Learning to use assessment in a more formative way requires significant changes for many districts, teachers, and students. For districts, it will mean ensuring that teachers have the time and resources to act on the assessment information they receive. For teachers and students, it will involve learning to use assessment information diagnostically to determine the course of instruction and learning. It will also involve dealing with learning difficulties that are revealed by formative assessments. Given the challenges involved in changing assessment practices—a substantial part of our research and development focuses on exploring the types and frequency of assessments and instructional supports that will be feasible to implement and be most beneficial to teachers and students. For instance, helping teachers understand mathematical concepts more deeply, monitoring learning of key ideas and skills, and figuring out the best strategies to improve students' understanding.

Learning-Based POWERSOURCE© Strategy

The POWERSOURCE© intervention thus involves not only the development of formative assessments but also the development of professional development and instructional support resources. The intervention aims to help teachers understand mathematical content, interpret assessment information, provide feedback to students, and adapt instruction as needed. Moreover, a striking innovation in POWERSOURCE© is its targeting of the big ideas—fundamental concepts and principles—and their interrelationships that underlie and define a field of knowledge, rather than treating specific concepts and topics in isolation, as do traditionally developed tests. This innovation is motivated by ample evidence from a range of cognitive psychology perspectives, which suggest that learning (in order to be acquired efficiently and sustained) must enable students to connect to organizing principles that would otherwise would be disconnected knowledge or procedures. Students should be able to integrate and demonstrate their knowledge and skills in many situations, in near and far transfer, and across time (e.g., Atkinson & Shiffrin, 1968; Chi, Feltovich, & Glaser, 1981; Ericson, 2003; Ericson & Simon, 1984; Hiebert & Carpenter, 1992; Mayer, 2003; Brown, Bransford, & Cocking, 2000; Newell, 1990, VanLehn, 1996, Catrambone & Holyoak, 1989).

Similarly, the specific item types used in POWERSOURCE© were developed based on cognitive research demonstrating the value of specific strategies for promoting transfer. Research, for example, suggests that learning and problem solving strategies can be successfully transferred if students are taught to focus on self-evaluation or metacognition (Moreno & Mayer, 2005; Palincsar & Brown, 1984; Pressley & Brainerd, 1985); the conditions for applying strategies (Judd, 1908, 1936; Kilpatrick, 1992); building principled representations of problem situations (Fuchs, Fuchs, Finelli, Courey, & Hamlett, 2004;

Kilpatrick, Swafford, & Findell, 2001); using worked-out examples as a way to build problem schemas that generalize across a range of tasks (Chi & Bassok, 1989; Pawley, Ayres, Cooper, & Sweller, 2005); lastly, explanation and problem solving tasks requiring understanding of core concepts and principles that recur across arithmetic, pre-algebra, and algebra (Carpenter & Franke, 2001; Haverty, 1999; Ready, Edley, & Snow, 2002; Schmidt, McKnight, & Raizen, 1997). POWERSOURCE© not only uses item types that are positioned to uniquely foster learning but it also purposively employs multiple formats to promote transfer, rather than focusing only on those representations adopted by test developers designing for accountability purposes (Richardson-Klavehn & Bjork, 2002).

Targeted Domains Operationalized in *Checks for Understanding*

The POWERSOURCE© intervention targets big ideas and related skills in four domains that underlie success in Algebra I: a) rational number equivalence (RNE); b) properties of arithmetic (PA); c) principles for solving linear equations (SE); and d) application of core principles in these domains to other critical areas of mathematics, such as geometry and probability (RA). These domains were chosen because of their importance to later mastery of algebra and their significant place in state mathematics standards across Grades 6-8.

In each domain we have designed a series of short POWERSOURCE© assessments comprised of multiple item types, which are called *Checks for Understanding*, to help teachers assess their students' understanding of basic mathematical principles and to connect their instruction and provide feedback to support deeper understanding. A set of instructional resources and targeted professional development activities were also developed for each of these domains. Thus, a POWERSOURCE© module around a given domain includes a set of *Checks for Understanding*, targeted instructional resources, and professional development opportunities. POWERSOURCE© materials are designed to complement existing curricula; yet, the time to implement POWERSOURCE© must be found within tight district curriculum frameworks and timelines. It is thus important for POWERSOURCE© to integrate well and easily with existing initiatives and not add an unreasonable burden to the heavy testing requirements already imposed on teachers (e.g., weeks of state and district testing), and not replace large chunks of extant curricula.

More detailed information about the research foundations, content focus, initial development process, and program components of POWERSOURCE© can be found in the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) 2006-2009 progress reports to the Institute of Education Sciences (Baker, 2006, 2007, 2008, 2009).

The present report focuses on providing an update on project activities undertaken since the last progress reporting period (i.e., covering the 2009-10 school year). This update is organized around four general areas:

1. First, we provide updated results from the 2007-08 experimental (randomized) field test of POWERSOURCE© instructional sensitivity for the *Checks for Understanding* and treatment/control differences on student and teacher outcomes.
2. Second, we describe the experimental (randomized) study conducted during the 2008-09 school year, and present findings on both student and teacher outcomes.
3. Third, we describe activities conducted during the 2009-10 school year.
4. Finally, we provide updates on supplemental/synergistic research studies and dissemination activities are also discussed.

Updated Results from 2007-08 POWERSOURCE© Field Test

As described in previous reports (Baker, 2007, 2008, 2009) the core undertaking of our work during the 2007-08 school year was continuing with an extended, random assignment implementation study of POWERSOURCE©. A detailed description of methodology used and preliminary results were presented in these previous progress reports. Following is an updated summary of 2007-08 school year results.

Experimental Comparison Findings

We employed a randomized, controlled design to address the following question: Does using our strategy improve student performance on assessments of key mathematical ideas relative to a comparison group? Eighty-five teachers and 4,091 students were included in the study. Students took a pretest and a transfer measure at the end of the year. Treatment students completed formative assessments; whereas treatment teachers had exposure to professional development and instructional resources.

On average, treatment students did not outperform those in control groups, given that we did not find a statistically significant main effect of the treatment. What we did find, however, was a significant interaction between treatment and pretest score. This indicates that students with higher scores on the pretest tend to benefit more from the intervention compared to students with lower pretest scores. The effect size for those students is as high as a 0.5 pooled standard deviation. In other words, the intervention had more impact on higher-performing students than lower-performing students. In addition treatment students significantly outperformed control students on distributive property items. This effect was larger as pretest scores increased.

POWERSOURCE© Implementation Study 2008-09

As described in previous reports (Baker, 2007, 2008, 2009), the core undertaking of our work during the 2008-09 school year was continuing with an extended, random assignment implementation study of the POWERSOURCE© program. In this year of the study, we expanded the intervention from only Grade 6, to Grades 6 and 7 (in all participating schools). As in prior years, new teachers were randomly assigned to either POWERSOURCE© or control conditions—with the ultimate goal of determining program impact on both student and teacher learning outcomes. Teachers who continued in the study for another year maintained their prior year's group status (for additional details about the 2008-09 study's plan and rationale, see Baker, 2009). The 2008-09 study was almost identical to the previous year's work), with a few minor changes:

1. An interim transfer measure was developed for use in Grade 6.
2. We created Grade 7 teacher instructional materials and *Checks for Understanding* assessments.
3. We modified the professional development sessions (in Grade 6) to focus more on interpreting student assessment data and less on teaching the big ideas.
4. We recruited an additional school district to replace a district that did not continue with the study.

Measure Quality and Item Analysis

This section documents the technical characteristics of the reliability, validity, and item analysis for some of the 2008-09 POWERSOURCE© *Checks for Understanding* assessments as well as the pretest and posttest (transfer measure) administered to all students. To examine the measure quality, alpha was used to calculate reliability; furthermore, exploratory factor analysis was applied to check the construct validity. To investigate the quality of test items, two different angles—according to different theories—could be applied: One is Classical Test Theory (CTT) and the other is Item Response Theory (IRT). Because both CTT and IRT can provide valuable information about a test, we used them to evaluate the items in the POWERSOURCE© *Checks for Understanding* assessments.

In keeping with findings in Phelan, Kang, Niemi, Vendlinski, and Choi (2009), which showed the appropriateness in using unidimensional Rasch models for POWERSOURCE© test items, the one parameter logistic model (1PLM) for dichotomous items and a partial credit model (PCM; Masters, 1982) for polytomous items were employed. Additionally, with the data sets for the pretest and transfer measure having items representing all domains—

factor analysis was conducted to estimate the amount of variance explained by the main construct.

Reliability

Table 1 shows the number of items, the actual number of examinees, and reliability for each form considered in the report. All of the data sets are not yet complete and so included in this analysis are the pretest, transfer measure, and the RA assessments. The reliability was computed with coefficient alpha as shown in Table 1.

Table 1

Sample Size and Reliability of the 2008-09 POWERSOURCE© Grade 6 and 7 Assessments

POWERSOURCE© assessments	Number of items	Sample size	Reliability (Cronbach's alpha)
Grade 6 assessments			
Pretest	28 (=28+0)	3,805	.81
Transfer measure (posttest)	30 (=26+4)	3,676	.83
Interim transfer measure	20 (=16+4)	3,419	.83
Grade 7 assessments			
Pretest	27 (=27+0)	3,570	.73
Transfer measure (posttest)	30 (=29+1)	3,104	.85

Note. The symbol = represents the number of dichotomous items; the symbol + represents the number of polytomous items.

Confirmatory Factor Analysis

Each of the student measures used in the POWERSOURCE© study included items from different conceptual domains. Thus, we conducted confirmatory factor analysis (CFA) to confirm these underlying structures. In our CFA analyses, the latent factors are the domains in each measure (pretest, interim measure, transfer measure). The items (indicator) are only linked to their corresponding domain (factor). The CFA models are slightly modified by allowing some residuals to be correlated. Only the residuals for items belonging to the same domain are correlated. We use the software SAS 9.1 to conduct the CFA analyses and use weighted least squares (WLS) estimation method for categorical indicators recommended by Albright and Park (2009).

Table 2 presents the model fit index for each measure in Grades 6 and 7 POWERSOURCE©. Hu and Bentler (1999) suggested that a Root Mean Squared Error of Approximation (RMSEA) of <.06 and a Comparative Fit Index (CFI) that is > .95 are the cut-off point for acceptable model fit. For CFA using categorical data, Yu (2002) suggested that the cut-off recommended by Hu and Bentler (1999) are acceptable. As shown in Table 2, the CFA of Grade 6 interim measure achieved acceptable model fit for both CFI and RMSEA. For other measures, the corresponding RMSEA values were within acceptable range. In our future research, we will conduct exploratory factor analysis to further analyze the underlying structures among items for each measure.

Table 2

Model Fit Index for Each Measure in 2008-09 POWERSOURCE©

Measure in POWERSOURCE©	Chi-square	Degree of freedom	Sample size	CFI	RMSEA	NFI
Grade 6:Pretest	3231	337	3805	0.65	0.05	0.62
Grade 6:Interim measure	993	204	3419	0.95	0.03	0.94
Grade 6:Transfer measure	1331	385	3676	0.90	0.03	0.86
Grade 7:Pretest	1037	304	3570	0.79	0.03	0.73
Grade 7:Transfer measure	1551	392	3104	0.81	0.03	0.76

The Grade 6 pretest includes items from four domains, thus, we fitted a four-factor CFA on this measure. Each latent factor corresponds to each domain. The unstandardized and standardized factor loadings for the Grade 6 pretest are reported in Appendix A: CFA Result of PS Grade 6 Pretest. We expected that items belonging to the same domain would obtain positive factor loadings on their corresponding latent factors; yet, one item, PRE23, loaded negatively on factor PA (properties of arithmetic). This indicates that this item (see Figure 1 for the item) did not fit very well with other items belonging to the PA domain. As reported in the item analysis section, PRE23 is the most difficult item. This could be the reason why the factor loading for this item is negative while factor loadings for all other items are positive. Statistically, all the factor loadings for the Grade 6 pretest are statistically significant from zero.

23. $(4 \div 6) \div 2$ has the same value as:

a) $4 \cdot 6 \cdot 2$

b) $4 \div 6 \cdot \frac{1}{2}$

c) $4 \div (6 \div 2)$

d) $4 \div 6 \cdot 2$

Figure 1. Item PRE23 on the Grade 6 pretest.

We present the factor loadings for the Grade 6 interim measure in Appendix B: CFA Result of PS Grade 6 Interim Measure. In the Grade 6 interim measure, test items belong to two domains, PA (properties of arithmetic) and RNE (rational number equivalence). As shown in Appendix B, all factor loadings are positive. In addition, our analysis results show that the estimated coefficients are statistically different from zero. The results indicate positive relations of the items to their corresponding domains.

The grade 6 transfer measure (see Appendix C: CFA Result of PS Grade 6 Transfer Measure), presents the standardized and unstandardized factor loadings. For this measure, based on the domain structure in the grade 6 transfer measure, we imposed a three-factor CFA model on the response data. Two items, POST03 and POST09, had negative factor loadings on factor SE, which represents the domain of principles for solving linear equations. The negative loadings indicate that these two items did not fit very well with other items belonging to the scale creating the SE domain. It is important to note that POST09 is the most difficult item in the grade 6 transfer measure. Based on our item analysis results, we found out that POST09 was a poor quality item; hence, this could be the reason that this item did not have a positive relation with the SE latent factor.

In Appendix D (CFA Result of PS Grade 7 Pretest), we present the factor loading for the Grade 7 pretest. The results show that the loadings for item PRE02 (on factor PA) and PRE 24 (on factor SE) are negative. In addition, the factor loading from PRE24 on factor SE is not statistically significant from zero. This means that the PRE02 and PRE 24 did not fit very well with the scale created for their corresponding domains, PA and SE. Figure 2 shows the test item PRE24.

24. A student is trying to solve the equation $3x + 6 = -9$. Which of the following shows a correct first step?

a) $\frac{3x}{3} + 6 = \frac{-9}{3}$

b) $3x + 6 + -9 = -9 - -9$

c) $\frac{3x}{3} + \frac{6}{3} = \frac{-9}{3}$

d) $9x = -9$

Figure 2. Item PRE24 on the Grade 7 pretest.

We present the factor loadings for the Grade 7 transfer measure in Appendix E (CFA Result of PS Grade 7 Transfer Measure). For the Grade 7 transfer measure, the CFA model is a four-factor model. Factor loadings from items and their corresponding factors (i.e., domain) are positive. In addition, these estimated coefficients are statistically significant from zero. These indicate the positive relation of the items towards their latent factors.

Item Analysis

Classical test theory. A descriptive analysis was used initially, which contained mean and standard deviation of the test score. Each item was examined using the proportion which answered the item correctly, *p*-values, and point-biserial correlation, r_{pbis}. The former and latter provide the information of item related to difficulty and discrimination, respectively. The point-biserial correlation is the correlation between the test-takers' performance on one item compared to the test-takers' performances on the total test score.

Item response theory. Because the 1PLM having every item discrimination to be 1 is nested within the PCM, the IRT model used in this report can be written:

$$P(z \mid \theta_j, \beta_i, \tau_{ci}) = \frac{\exp \sum_{c=0}^{z} \left[\theta_j - (\beta_i - \tau_{ci}) \right]}{\sum_{y=0}^{Z_i} \exp \sum_{c=0}^{y} \left[\theta_j - (\beta_i - \tau_{ci}) \right]} \qquad (1)$$

Under the PCM, the probability that an examinee *j* scores *z* with $z = 0, \ldots, Z_i$ on item *i* with $Z_i + 1$ response categories. β_i denotes the difficulty of item *i*, and τ_{ci} represents the location parameter for a category on item *i*. Equation 7 needs to set $\tau_{0i} = 0$, $\sum_{c=1}^{Z_i} \tau_{ci} = 0$ and

$\exp\sum_{c=0}^{0}\left[\theta_j-(\beta_i-\tau_{ci})\right]=1$ for model identification. For a dichotomous item with $Z_i=1$, there are two response categories (i.e., 0 and 1) and only β_i exists as the related item parameter.

2008-09 POWERSOURCE© Pretest

Grade 6

In the Grade 6 pretest, there were 28 multiple-choice items. Among these items, item PRE08 was the easiest item (b=-1.54, *p*-value=0.94) and item PRE23 was the most difficult item (b=3.60, *p*-value=0.13). PRE03 and PRE23 are shown in Figure 3. Regarding the polyserial correlation coefficient between item and test scores, the correlation indexes for pretest items were larger than 0.30, except for item PRE23 and PRE24. Especially for PRE23, the polyserial correlation is equal to 0.06 and the total scores for students who answered this item correctly or incorrectly were very similar. This means that PRE23 was a poor quality item.

8. Terry has 4 hats and Dan has 5 more hats than Terry. How many hats does Dan have?

RA-PT-5

a) 4
b) 5
c) 1
d) 9

Item PRE08

23. (4 ÷ 6) ÷ 2 has the same value as:

PA-PT-4

a) 4 • 6 • 2
b) 4 ÷ 6 • $\frac{1}{2}$
c) 4 ÷ (6 ÷ 2)
d) 4 ÷ 6 • 2

Item PRE23

Figure 3. Easiest item and most difficult item on Grade 6 pretest.

We calculated the item reliability and test reliability based on IRT. The method we used in our calculation was the one suggested by Dimitrov (2003). The overall test reliability for the Grade 6 pretest was 0.92 (Cronbach's alpha=0.81). The most difficult item, PRE23, had the lowest item reliability (item reliability=0.05). The item that had the highest reliability was PRE15 (item reliability=0.34). Overall, the item reliability estimates for the Grade 6 pretest items were around 0.30. The higher an item's reliability, the more an item contributes to the test reliability. Compared to other test items, PRE23 makes a relatively small contribution to the overall test reliability. Appendix F (Item Analysis Results of PS Grade 6 Pretest) includes the detailed results from the item analysis of Grade 6 data.

We show item information curves in Figure 4. The easiest item (PRE08) provided the highest amount of information for examinees with low ability and the most difficult item (PRE23) gave a relatively large amount of information for examinees with high ability.

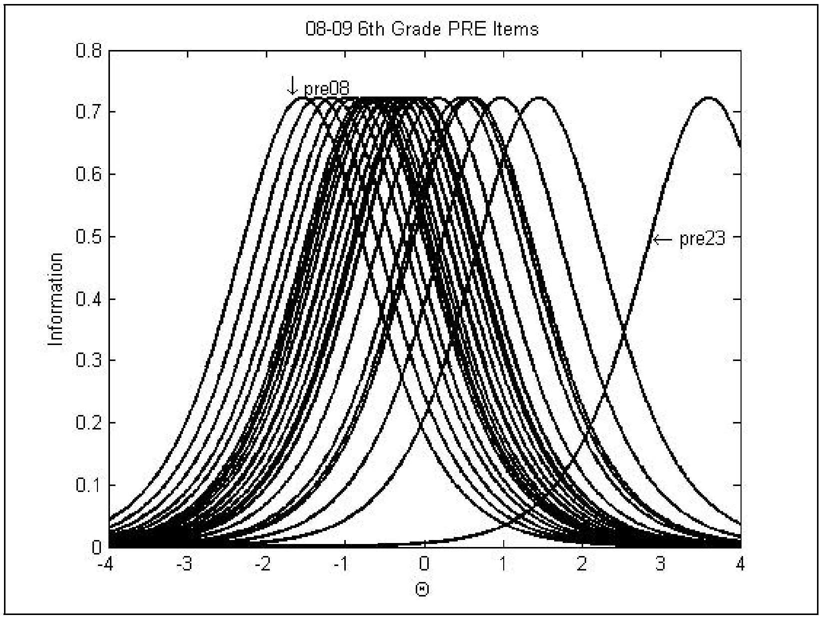


Figure 4. The item information curves of POWERSOURCE© Grade 6 pretest items.

Grade 7

In the 2008-09 Grade 7 POWERSOURCE© pretest, there were 27 multiple-choice items. These items were treated as dichotomous items in the item analyses. The Rasch item difficulty ranges from -2.26 to 1.23. Item PRE05 is the easiest item (b=-2.26, p=0.97) and item PRE18 is the most difficult (b=1.23, p=0.13) among the pretest items.

We also calculated item and test reliability based on IRT. For the Grade 7 pretest, the test reliability was 0.92 (Cronbach's alpha=0.73). The item with the lowest item reliability was also the easiest item (PRE05). Whereas, the item with the highest item reliability was PRE06, for which the item difficulty was around the middle range of the ability continuum. The detailed item analysis results are shown in Appendix G (Item Analysis Results of PS Grade 7 Pretest).

Figure 5 shows the item information curves for the Grade 7 pretest items. The easiest item (PRE05) provides more information for examinees with low ability and the most difficult item (PRE18) yields more information for examinees with high ability.

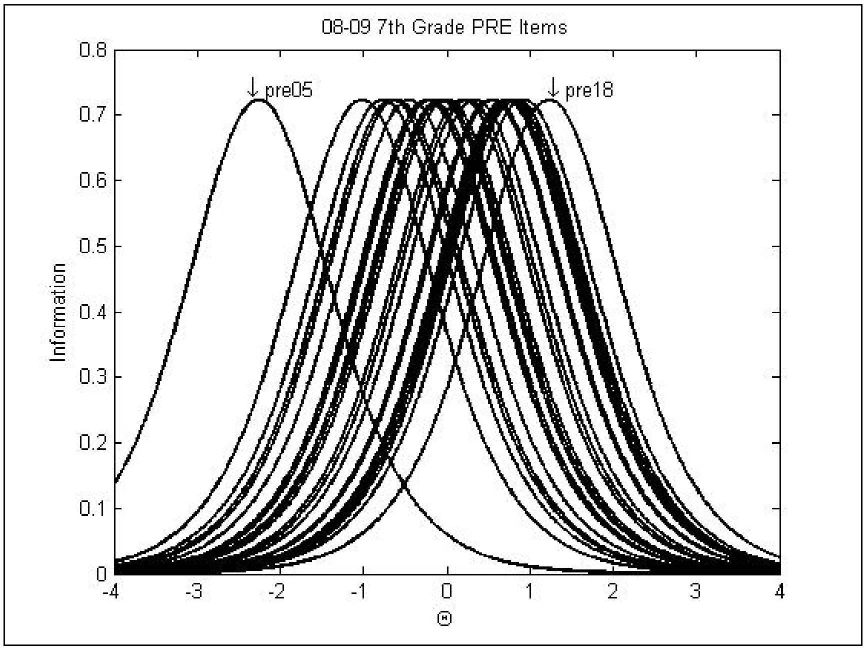


Figure 5. The item information curves of POWERSOURCE© Grade 7 pretest items.

Implementation Study 2008-09: Transfer Measure

Grade 6

The Grade 6 transfer measure was first used in 2007-08. This transfer measure was developed using items from several sources including the Trends in International Mathematics and Science Study (TIMSS); the National Assessment of Educational Progress (NAEP); the Qualifications and Curriculum Authority (QCA) Key Stage 3 exam; the Programme for International Student Assessment (PISA); and benchmark tests used in one of our pilot districts (see Appendix H: Sources of Transfer Measure Items). An initial set of 44 items were selected from the various sources. Items were selected based on their relevance to the POWERSOURCE© domains and their appropriateness for a transfer task (related to POWERSOURCE© content, but not exact replicas of item types used in the *Checks for Understanding*). A final set of items (29) were selected from the initial 44 items. Of these items, 19 were multiple choice; nine short answer; and one was an explanation task. Items were selected based on their representation in the California (CA) state standards and relevance to POWERSOURCE© items (see Appendix I: Alignment of CA Standards and

NCTM). Some of the initially selected items were deemed more appropriate for Grade 7 and were used for the Grade 7 transfer measure.

In the Grade 6 transfer measure there were 27 transfer measure items. Among them, there were 20 multiple-choice items. The rest of the test items were six extended-response items—one item (POST10) with two parts (POST10A and POST10B) and one item (POST27) with three parts (POST27A, POST27B and POST27C). For item POST10, POST10A was a short-answer item and POST10B was an extended response item. For POST27, all three parts were short-answer. Among the Grade 6 transfer measure items, POST10 included POST10A and POST10B. They were separated and treated as one dichotomous item (POST10A) and one polytomous item (POST10B). Additionally, item POST27 was separated and treated as three dichotomous items since students' answers to short-answer items were scored as 0 for an incorrect response and as 1 for a correct response. Similarly, the multiple choice items were also scored as either 0 or 1. Thus, along with the 20 multiple-choice items, there were 24 dichotomous items. The 5 extended-response items and 1 extended response item part (10B) were treated as polytomous items in the item analysis. In total, there were 6 polytomous items.

Our preliminary analysis showed that for some polytomous items, very few students received the highest scores. For example, for item POST04, the highest score was 4 but less than 0.5% of students received a score equal to 4 for this item. To obtain stable estimation of item parameters, we determined to collapse some score categories and combined them with the adjacent score category. In the item analyses, score categories for three polytomous items were collapsed. Originally, two items had three score categories. After collapsing the score categories, these two items became dichotomous items. Thus in total, there were 26 dichotomous (24+2=26) items in our final item analyses.

In the item analyses of Grade 6 transfer measure items, we found out that POST01 was the easiest item (b=-1.80 and *p*-value=0.94); whereas, POST03 (b=2.12, *p*-value=0.04) and POST09 (b=2.24, *p*-value=0.03) were very difficult items. There were less than 5% of students who were able to answer them correctly. The polyserial correlation coefficients for the transfer measure items were all positive and larger than 0.35 except for POST03 and POST09. For POST03, the polyserial correlation coefficient was close to zero and for POST09 the polyserial correlation coefficient was negative. The total scores for students who answered these two items correctly were lower than the total scores for students who answered these two items incorrectly. Thus these two items had poor discrimination and were bad quality items.

The IRT test reliability for Grade 6 transfer measure was 0.93 (Cronbach's alpha=0.83). Two polytomous items, POST04 (item reliability=0.59) and POST10b (item reliability=0.57), had relatively high item reliability compared to other items. This means that these two items contributed a lot more to the overall test reliability than other items. For dichotomous items, the item reliability estimates were around 0.3. We present the detailed item analysis results in Appendix J (Item Analysis Results of PS Grade 6 Posttest).

Figure 6 shows the item information curves for Grade 6 transfer measure. To differentiate the polytomous items from dichotomous items, we highlight the item information curves for these items with different colors. As seen in Figure 6, polytomous items yielded a relatively larger amount of information than the dichotomous items. The easiest item (POST01) mainly gave information for examinees with low ability while the most difficult item (POST09) mainly provided information for examinees with high ability. Interestingly, Figure 6 shows that the item information curve for polytomous item POST25 was bimodal; furthermore, we can see that this item provided information for both examinees with relatively low ability and examinees with high ability.

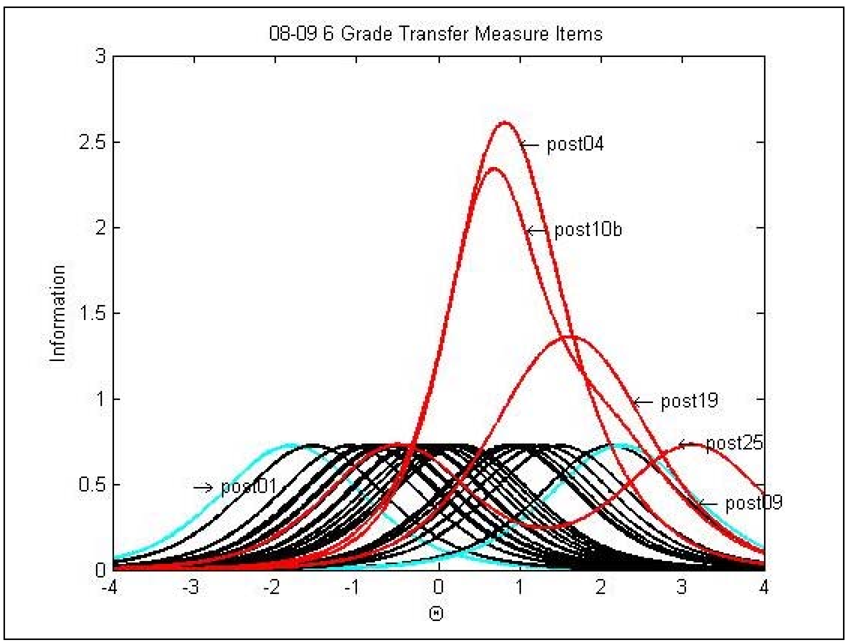


Figure 6. The item information curves of POWERSOURCE© Grade 6 transfer measure items.

Grade 7

The Grade 7 transfer measure was developed using similar procedures as the Grade 6 transfer measure. Items were selected from TIMSS, NAEP, the QCA Key Stage 3 exam, PISA and benchmark tests used in one of our pilot districts (see Appendix H for the sources of all items). Items were selected based on their relevance to the POWERSOURCE© domains and their appropriateness for a transfer task (related to POWERSOURCE© content but not exact replicas of item types used in the *Checks for Understanding*). An initial set of 51 items were selected and narrowed down to a final pool of 26 items. Of these items, 17 were multiple choice and the rest were either short answer or explanation tasks (or a combination of both types). Items were selected based on their representation in the CA state standards and relevance to POWERSOURCE© items (see Appendix K: Grade 7 Transfer Measure 2008/2009).

The Grade 7 transfer measure includes 26 test items. Several items had multiple parts and one item was the explanation task item. The multiple parts for a test item were treated as separate and different items in the item analyses. The extended response item was treated as a polytomous item. In total, there were 29 dichotomous items and 1 polytomous item in the item analyses. Item difficulty was calculated for each item. Among these items, POST01 was the easiest item (b=-1.647, *p*-value=0.91); whereas, the two difficult items were POST25 (b=2.500, p=0.03) and POST26A (b=2.770, *p*-value=0.02). The polyserial correlation coefficients for all items were larger than 0.30—except for one item with polyserial correlation coefficient equal to 0.30.

For Grade 7 transfer measure item analysis, we calculated the item and test reliability. The test reliability was 0.92 (Cronbach's alpha=0.85). The two difficulty items, POST25 and POST26A obtained relatively low item reliability. As expected, the polytomous item (POST20) obtained the highest item reliability (item reliability=0.51). Figure 7 shows the item information curves. The item information curve for POST20 also showed that this item provided the largest amount of information. The easy item (POST01) gave more information for examinees with low ability and the difficult item (POST26A) mainly provided information for examinees with high ability. The related item analysis results are presented in Appendix L (Item Analysis Results of PS Grade 7 Posttest).

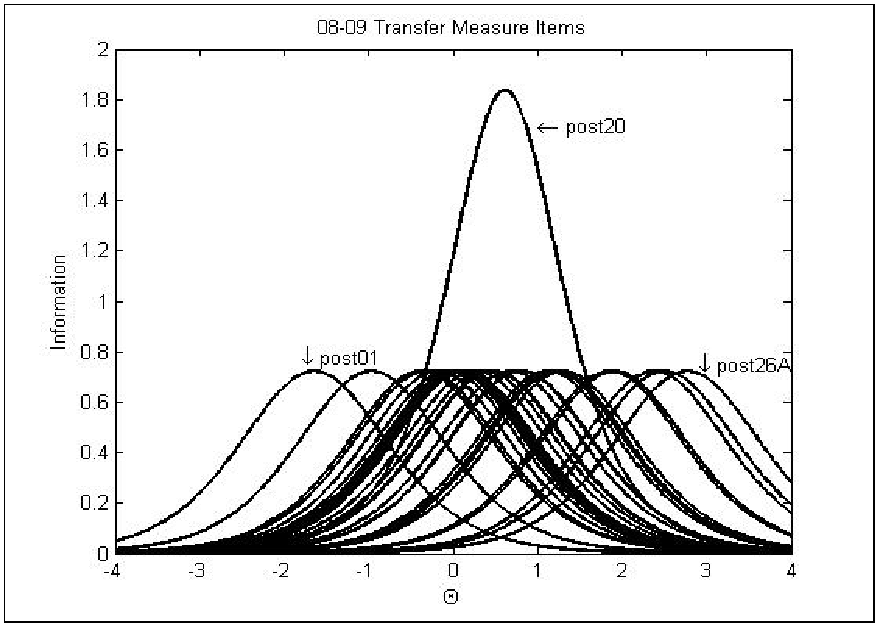


Figure 7. The item information curves of POWERSOURCE© Grade 7 transfer measure items.

Implementation Study 2008-09: Interim Transfer Measure

Grade 6

In an effort to gather more student outcome data, we designed an interim transfer measure to be given to students after completion of the first two POWERSOURCE© domains (PA and RNE). We created a 20 item test form with 20% of the items requiring students to explain a concept in their answer. We selected two items per domain from the pretest (of medium difficulty) and changed the numbers in the items. The remaining items were taken from the transfer measure and again were modified to include different numbers and/or situations. Items selected for the interim transfer measure had a range of difficulty from b=1.101, *p*-value = .17, to b= -1.441, *p*-value = 0.88 (see Appendix M: Grade 6 Transfer Measure 2008/2009).

The Grade 6 interim measure included 20 test items; several items had multiple parts. In our item analysis, items with multiple parts were treated as separate items. In this way, we had 18 dichotomous items and 4 polytomous items. Item difficulty was calculated for each item. Among these items, INTER01 was the easiest (b=-1.43, *p*-value=0.88) and the most difficult was INTER09 (b=2.68, *p*-value=0.14). The polyserial correlation coefficients for all items were larger than 0.40 except for one item with polyserial correlation coefficient lower than to 0.40.

In the Grade 6 Interim measure item analysis, we calculated the item and test reliability. The test reliability was 0.92 (Cronbach's alpha=0.83). The most difficult item and the easiest item, INTER09 and INTER01, obtained relatively low item reliability. As expected, the polytomous items (INTER18B, INTER19B) obtained higher item reliability The item reliability values for these two items were equal to 0.58 and 0.61 (respectively). Figure 8 shows the item information curves. The item information curve for INTER19B reveals that this item provided the largest amount of information. The easy item (INTER01) gives more information for examinees with low ability; while, the difficult item (INTER09) mainly provides information for examinees with high ability. The related item analysis results are presented in Appendix N (Item Analysis Results of PS Grade 6 Intertest).

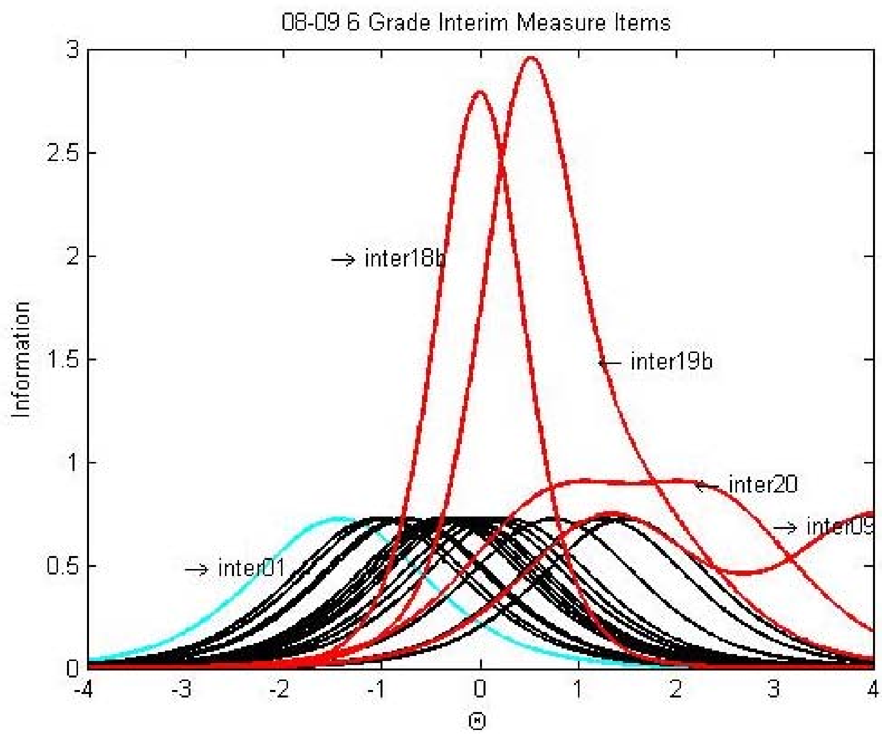


Figure 8. The item information curves of POWERSOURCE© Grade 6 interim measure items.

Implementation Study 2008-09: Student Outcomes

Six districts participated in the random assignment implementation study in 2008-09. As described earlier, we used two designs (within and between school) based on district needs and configuration. Ultimately, three of the districts used a within-school (W-S) design—where random assignment was accomplished within each school (i.e., a given school had both treatment and control teachers). Two districts used a between-school (B-S) design—where schools within a district were randomly assigned to treatment or control conditions. Note that one district adopted both W-S and B-S design.

For Grade 7, fifty three teachers from 20 schools in six school districts participated in the study. Table 3 shows the number of students and teachers in each study design; while Table 4 shows the distribution of teachers in each district.

Table 3

Grade 7 Sample Distribution for 2008-09 School Year

Control/treatment	*N* of students	*N* of teachers	*N* of schools
Between			
Control	689	16	5
Treatment	567	13	7
Subtotal	1,256	29	12
Within			
Control	527	9	8
Treatment	810	15	8
Subtotal	1,337	24	8

Table 4

Grade 7 Sample Distribution by School District for the 2008-09 School Year

District	*N* of students	*N* of teachers	*N* of schools	Design
AZ-1	220	4	3	B/S
CA-1	640	9	2	W/S
CA-2	367	7	2	W/S
CA-3	90	2	1	W/S
CA-6	862	19	5	B/S
CA-7	414	12	7	B/S, W/S

In Grade 6, there were 46 teachers from 19 schools in the B-S design and 23 teachers from 5 schools in the W-S design (see Table 5). Table 6 shows the distribution of teachers in each district.

Table 5

Grade 6 Sample Distribution for the 2008-09 School Year

Control/treatment	*N* of students	*N* of teachers	*N* of schools
Between			
Control	806	18	9
Treatment	1,050	28	10
Subtotal	1,856	46	19
Within			
Control	579	10	5
Treatment	745	13	5
Subtotal	1,324	23	5

Table 6

Grade 6 Sample Distribution by School District for the 2008-09 School Year

District	*N* of students	*N* of teachers	*N* of schools	Design
AZ-1	590	9	3	BS
CA-1	1225	16	3	WS
CA-2	805	7	2	WS
CA-3	245	7	4	BS
CA-6	1727	33	9	BS
CA-7	170	3	3	WS

Descriptive Statistics

We will now present the descriptive statistics for both the pretest score and transfer measure. Tables 7 and 8 depict the total mean score on both measures for the Grade 7 sample. We must note that this result is based on students who have both pretest and transfer measure scores. Table 6 shows that the observed mean difference in pretest score between the POWERSOURCE© and control group is very small (0.14) for the B-S design; this also true for the W-S design. In sum, these results suggest that the two groups are equivalent in terms of pretest scores.

Mean scores on the transfer measure were similar between the control students (*M* = 10.32) and POWERSOURCE© (*M* = 9.95) students in the B-S design. The observed mean difference is approximately 0.37, which is a 0.07 pooled standard deviation. In the W-S design, POWERSOURCE© students had a higher mean score on the transfer measure (M = 11.68) than the control group students (M =11.31). The observed mean difference was also 0.37, which is the size of a 0.07 pooled standard deviation.

Table 7

Descriptive Statistics of Grade 7 Pretest Scores

		Pretest total			
Design	*N*	Mean	*SD*	Min	Max
Between					
Control	689	11.61	4.04	0	24
Treatment	567	11.47	3.67	1	23
Within					
Control	527	13.61	4.30	0	24
Treatment	810	13.22	4.72	0	27

Table 8

Descriptive Statistics of Grade 7 Transfer Measure Scores

		Pretest total			
Design	*N*	Mean	*SD*	Min	Max
Between					
Control	689	10.32	5.71	1	28
Treatment	567	9.95	5.12	1	26
Within					
Control	527	11.31	5.50	0	27
Treatment	810	11.68	5.80	0	29

Tables 9-11 present the total mean score on the pretest, transfer measure, as well as on the interim measure (only used in Grade 6 in the 2008-09 study year). The mean pretest score for the control group in the B-S design (M = 18.81) was higher than the mean score for the POWERSOURCE© group (M = 17.63); in fact, the difference between the mean pretest scores was 1.2, which is a little larger than 1/5 of the pooled standard deviation. In the W-S design, however, mean pretest scores for the POWERSOURCE© students (M = 20.29) were 1.4 points (1/3 of the pooled standard deviation) higher than the control group students (M =

18.88). This difference is larger than the Grade 7 students described earlier; however, the difference will be taken into account when the pretest score is included as a covariate in the subsequent analyses.

For the interim transfer measure, in the B-S design there was a 0.26 point difference between the two groups, which is negligibly small (a 0.05 pooled standard deviation difference). The observed difference (1.74) in W-S design (a 0.35 pooled standard deviation difference) is larger.

For students in the B-S design, mean scores on the transfer measure were similar between the POWERSOURCE© students (M = 14.21) and the control group students (M = 14.59). The observed mean difference is approximately 0.38, which is a 0.07 pooled standard deviation. In contrast the POWERSOURCE© students in the W-S design had a higher mean transfer measure score (M = 16.32) than the control group students (M = 14.32). This difference of two points is the size of a 0.36 pooled standard deviation.

Table 9

Descriptive Statistics of Grade 6 Pretest Scores

		Pretest total			
Design	*N*	Mean	*SD*	Min	Max
Between					
Control	818	18.81	4.39	0	28
Treatment	1,058	17.63	4.52	0	27
Within					
Control	579	18.88	4.06	0	27
Treatment	745	20.29	4.11	0	28

Table 10

Descriptive Statistics of Grade 6 Interim Test Scores

		Pretest total			
Design	*N*	Mean	*SD*	Min	Max
Between					
Control	691	11.53	5.13	0	25
Treatment	1,055	11.79	5.27	0	25
Within					
Control	569	11.84	4.17	2	26
Treatment	768	13.57	5.67	0	25

Table 11

Descriptive Statistics of Grade 6Transfer Measure Scores

		Pretest total			
Design	*N*	Mean	*SD*	Min	Max
Between					
Control	818	14.59	5.51	0	29
Treatment	1,058	14.21	5.54	0	33
Within					
Control	579	14.32	4.71	2	27
Treatment	745	16.32	6.47	1	32

We also calculated descriptive statistics for each of the content domains, in each of the grades. Results of these more detailed analyses can be found in Appendix O (Descriptive Statistics by Content Domain). Descriptive statistics were also calculated for each district, school and teacher (see Appendix P: Additional Descriptive Statistics for Grades 6 and 7).

Methodological Concerns

As described earlier, two different designs (B-S and W-S) were implemented due to districts' needs and configuration. Given the relatively small sample size in both the B-S design (30 teachers, 12 schools for Grade 7; 46 teachers, 19 schools for Grade 6) and the W-S design (25 teachers, 8 schools for grade 7; 23 teachers, 5 schools for Grade 6), the statistical power of the key parameter of interest (capturing treatment effect) is not as high as we would like. Furthermore, in situations where two designs are implemented, it is common to analyze those results separately. This raises two issues: 1) synthesis of two different results; and 2) losing statistical power.

Lastly, there is a concern on the choice of unit of analysis. At first glance, the data has three-level hierarchical structures (i.e., students are nested within teachers, who in turn are nested within schools) in both designs. However, given the small number of teachers in some schools, especially in the B-S design where most of schools only have two or three teachers, it does not seem to be appropriate to use "teacher" as another level in the hierarchical model. One possible solution is that in both designs a 2-level hierarchical model could be used (i.e., students in level-1 and schools in level-2). The problem with this approach, however, is that within-school individual teacher variability is ignored. Furthermore, all the valuable teacher information (e.g., three different individual teacher pre-and post-surveys) can be used only as a school aggregate.

HLM Results

Taking methodological concerns into account, we used a two-level hierarchical model (HM) to examine the POWERSOURCE© effects on the transfer measure outcome. In order to synthesize two different designs and compromise unit of analysis issue, we chose teacher as a unit of analysis; individual school effects are also included in a model. School specific fixed effects take care of school blocking factors and intra-class correlation of school in a model. As such, we can examine whether there is a differential treatment effect depending upon two different designs not at the cost of losing statistical power (see Appendix Q for the complete statistical model).

Grade 7 Transfer Measure

In this section, we will present HM results (where the Grade 7 posttest score is used as an outcome). Appendix R (Estimates for Fixed Effects and the Variance) presents all the estimates for fixed effects and the variance components in the model. The results show that there are no statistically significant main effects for treatment and design on the total posttest score. The interaction term between treatment and design is also not significant. However, the pretest main effect is significant. The estimate of pretest mean is 1.16 and its *p*-value is smaller than 0.001. This means that teachers whose students have a higher pretest mean score tend to have higher posttest mean scores.

It is important to note that the interaction effect between treatment and students' pretest scores is significant. The estimate of the interaction effect is 0.17 with the *p*-value being very close to 0.05. We show this interaction effect in Figure 9, which presents the fitted relationships between pretest and posttest scores. As shown in Figure 9, the slopes of the lines for both B-S and W-S are steeper for treatment groups than for the control groups. It indicates that students with higher pretest scores tend to benefit more from the treatment as compared to those with lower pretest scores.

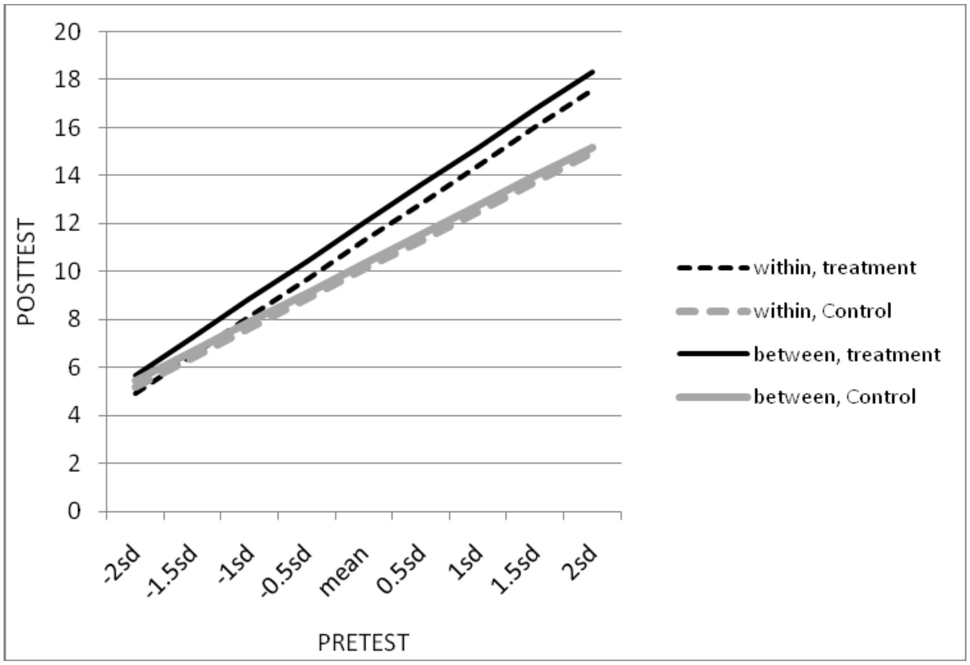


Figure 9. HM result (Grade 7 transfer measure total score): Fitted relationships between pretest score and posttest by design and treatment condition.

Figures 10 and 11 present model-fitted lines, respectively, for the W-S and for B-W design. For the W-S design, the two fitted lines are crossed at about 2.0 *SD* below the pretest score mean. In other words, despite no main effect of the treatment, the effect of treatment may have a small magnitude but become positive for most of students (i.e., students whose pretest scores are higher than -2 *SD* of the pretest score mean). Statistically, POWERSOURCE© students who had a mean pretest score of 2 *SD* above the mean, significantly outperformed the W-S design control group students on the transfer measure.

In the B-S design (see Figure 11), the lines for treatment group and control group are also crossed around 2.0 *SD* below the pretest score mean. That is, the fitted line for treatment group (dark solid line) is located above the fitted line for control group (gray solid line) from the point of 2 *SD* above the pretest mean. This shows that POWERSOURCE© students have higher posttest scores than control students through almost the entire range of pretest scores. The difference in posttest score between POWERSOURCE© students and control group students increases as the pretest score increases. Specifically, the estimated difference between the two groups at the pretest mean is approximately 2 points (approximately 1/3 of the pooled standard deviation). In addition, statistical tests show that the difference between the two groups of students becomes statistically significant from mean pretest score and above.

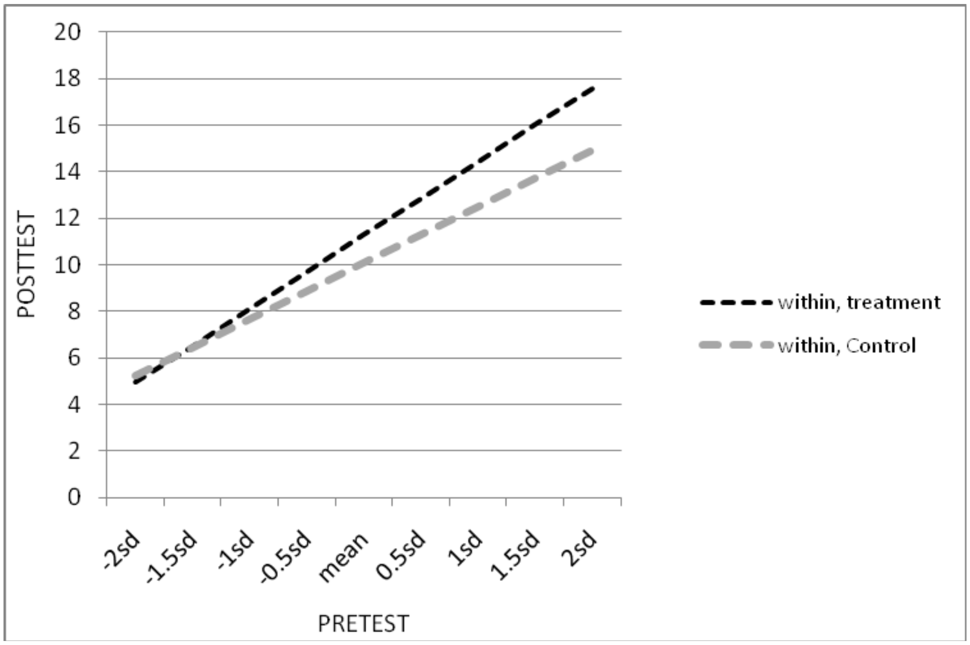


Figure 10. HM result (Grade 7 total score): Fitted relationship between pretest and posttest for treatment conditions in within-school design.

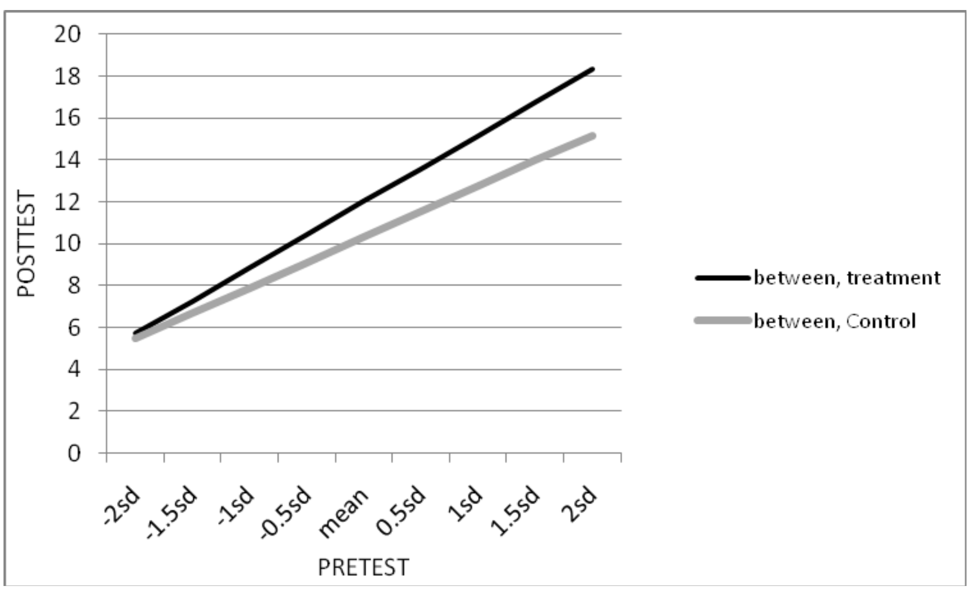


Figure 11. HM result (Grade 7 total score): Fitted relationship between pretest and posttest for treatment conditions in between-school design.

Transfer measure subscore (Grade 7 RNE). In the transfer measure, there were five items related to RNE. The result of the HM analysis results (using subscore for domain RNE as the outcome variable) showed no main effect of treatment of design. In addition, the interaction effect of treatment and design is not statistically significant. However, the interaction effect of treatment and student pretest score is just marginally insignificant (*p*-value = 0.06). Taking the small sample size into account, it seems that POWERSOURCE© students with higher RNE pretest scores benefit more than those with lower RNE pretest scores (see Appendix S Subdomain HLM Analysis Results for Grade 7).

HM result: transfer measure subscore (Grade 7 PA). Appendix S illustrates the results from items assesssing knowledge of properties of arithmetic (specifically the distributive property) of which there were three on the transfer measure. On these items we found that there is no main effect of treatment that is statistically signficant (estimate= 0.14, *p*-value = 0.190); moreover, the interaction effect between pretest and treatment is not significant (estiamte = 0.02, *p*-value = 0.467). Likewise, the main effect of design is not statistically significant nor is it statistically significant in treatment or design interaction.

HM result: transfer measure subdomain (Grade 7 SE). We had 13 solving equations-related items on the transfer measure. One noticeable finding from the HLM analysis was that the estimate of design was statistically significant (estimate = 2.35, *p*-value=0.031). Figure 12 shows the average SE transfer measure scores are different between the two groups. Thus, the average score for the B-S design is higher than for the W-S design (see Appendix S). However, there were no statistically significant main effects of treatment, nor an interaction effect of treatment and design.

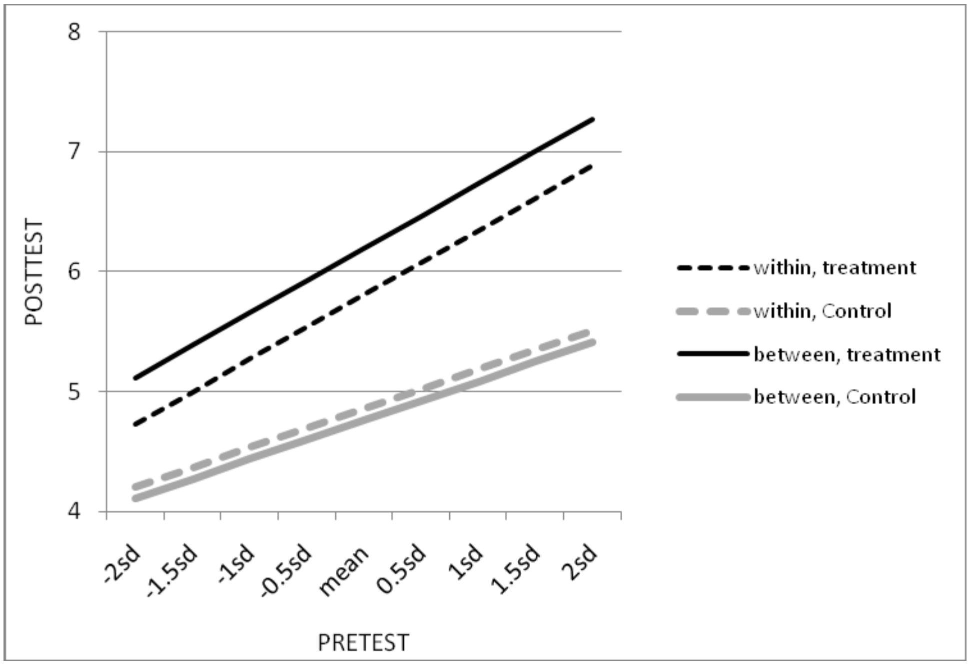


Figure 12. HM result for Grade 7 (SE transfer measure subscore): Fitted relationships between pretest SE subscore and posttest SE subscore by design and treatment condition.

Grade 6 Measures

For Grade 6 students there were three student measures: pretest, interim transfer measure, and transfer measure. To identify the POWERSOURCE© effects, we conducted three HM analyses. In the first HM analysis, we used the transfer measure score as the outcome and the interim measure as a covariate. In the second HM analysis, the interim measure was the outcome variable and pretest score the covariate in the 2-level HM model. In the third analysis, we treated the transfer measure as the outcome and pretest score as covariate.

Grade 6 Transfer Measure

For the Grade 6 HM analysis, in this section we will present the results where the Grade 6 posttest score was used as an outcome and the interim measure as a covariate. Appendix T presents all the estimates for fixed effects and the variance components in the model (see Appendix T: Fixed Effects and the Variance Components in the Model_Grade 6). The results show that there are no statistically significant main effects of treatment and design on the total transfer measure score. The interaction term between treatment and design is also not significant; however, the interim measure main effect is significant. The estimate of the interim measure mean is 1.14 and its *p*-value is smaller than 0.001. This means that teachers with students who have higher interim measure mean scores tended to have higher posttest mean scores.

The interaction effect between treatment and students' interim measure score is significant. The estimate of the interaction effect is 0.14 with the *p*-value less than 0.05. We show this interaction effect in Figure 13, which presents the fitted relationships between the interim and transfer measure scores. As shown in Figure 13, the slopes of the lines for both B-S and W-S are steeper for treatment groups than for the control groups. This indicates that students with higher interim measure scores tend to benefit more from the treatment as compared to those with lower interim measure scores.

Figures 14 and 15 present separate lines in Figure 13 based on different design in the analysis. Figures 14 and 15 present model-fitted lines, respectively, for the W-S and the B-S design. For the W-S design, the two fitted lines are crossed at about 2.0 *SD* below the interim measure score mean. This means that despite no main effect of the treatment, the effect of treatment may have a small magnitude but become positive for most of students (i.e., students whose pretest scores are higher than -2 *SD* of pretest score mean).

In the B-S design (see Figure 15), the lines for the treatment and control groups are also crossed at 2.0 *SD* below the interim measure score mean. The fitted lines for treatment group and control group in Figure 15 are similar to those in Figure 14. The difference between these two figures is that for the same interim measure score, the average posttest scores in the B-S design are slightly higher than those in the W-S design. In addition, statistical tests show that the difference between the two groups of students (in both designs) becomes statistically significant from the point where the pretest score is 2 *SD* above the mean.

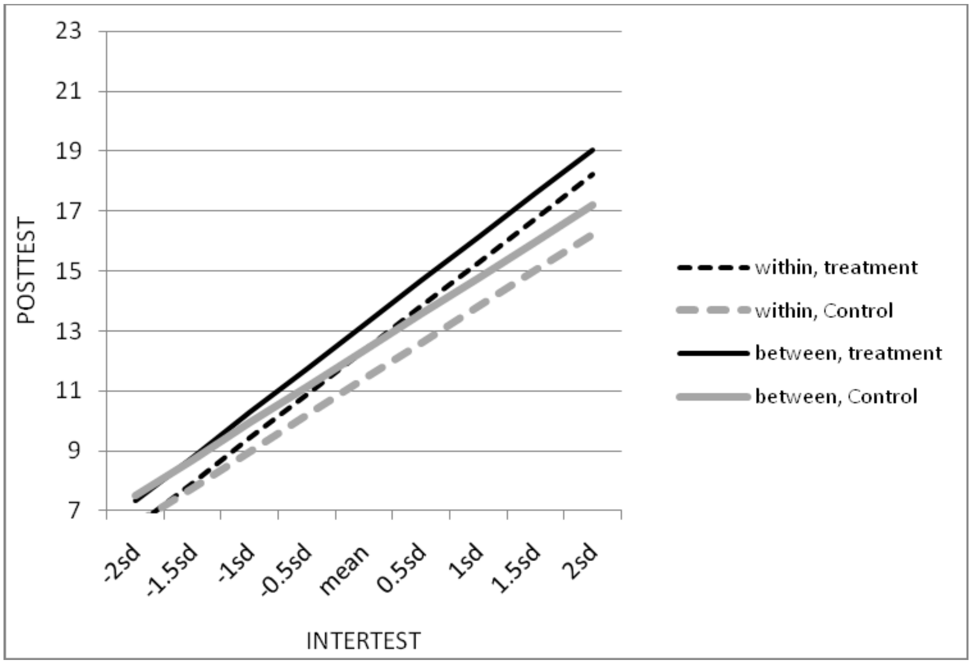


Figure 13. HM result (Grade 6 transfer measure total score): Fitted relationships between interim measure and posttest by design and treatment condition.

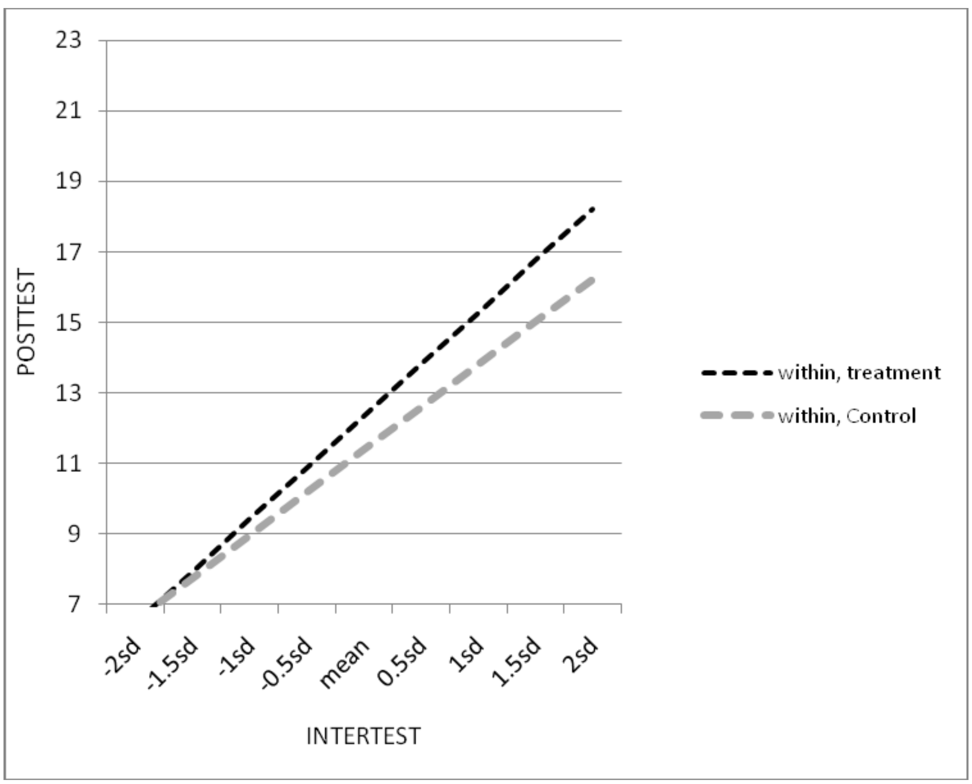


Figure 14. HM result (Grade 6 total score): Fitted relationship between interim measure and posttest for treatment conditions in within-school design.

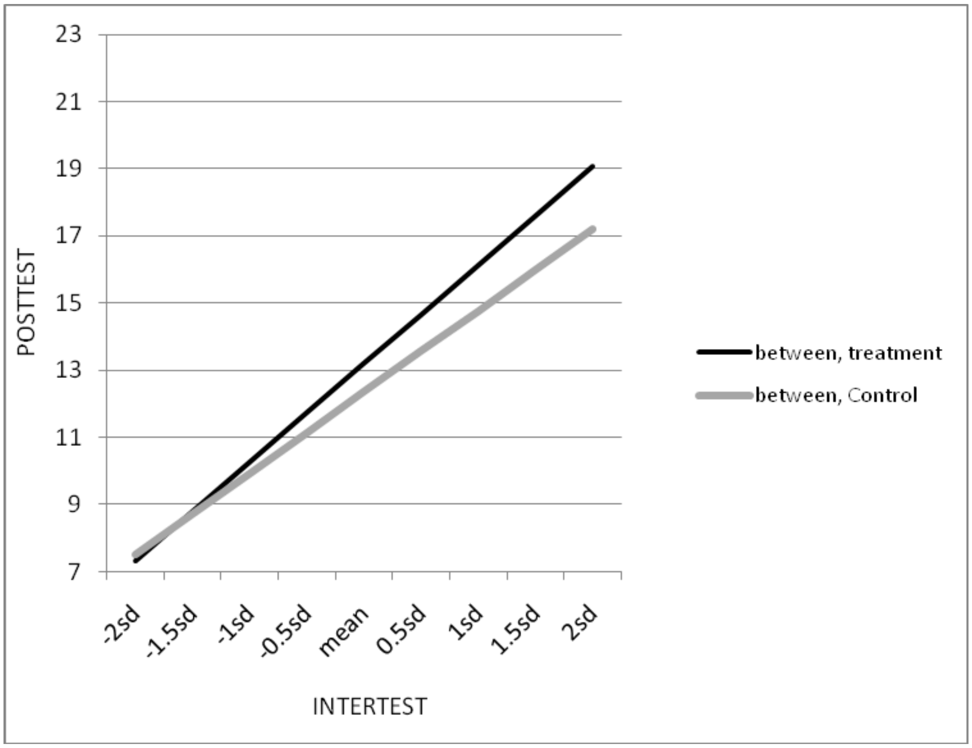


Figure 15. HM result (Grade 6 total score): Fitted relationship between interim measure and posttest for treatment conditions in between-school design.

HM result: Transfer measure subdomain. In the Grade 6 transfer measure, there were eight items related to RNE. Appendix U (Subdomains as an Outcome Variable_Grade 6) presents the HM analysis results using the subdomains as an outcome variable. The results show that there was no main effect of design. However, the main effect of treatment was statistically significant (estimate=0.67, *p*-value=0.03) and the interaction effect between interim measure and treatment was also statistically significant (estimate=0.08, *p*-value=0.401). As can be seen in Figure 16, the average RNE posttest scores are higher for the treatment group than those for the control group for both W-S design and B-S design. In other words, for RNE, the POWERSOURCE© students achieved significantly higher transfer measure scores and the difference in the posttest for POWERSOURCE© students and non-POWERSOURCE© students became larger for those whose interim measure scores were higher.

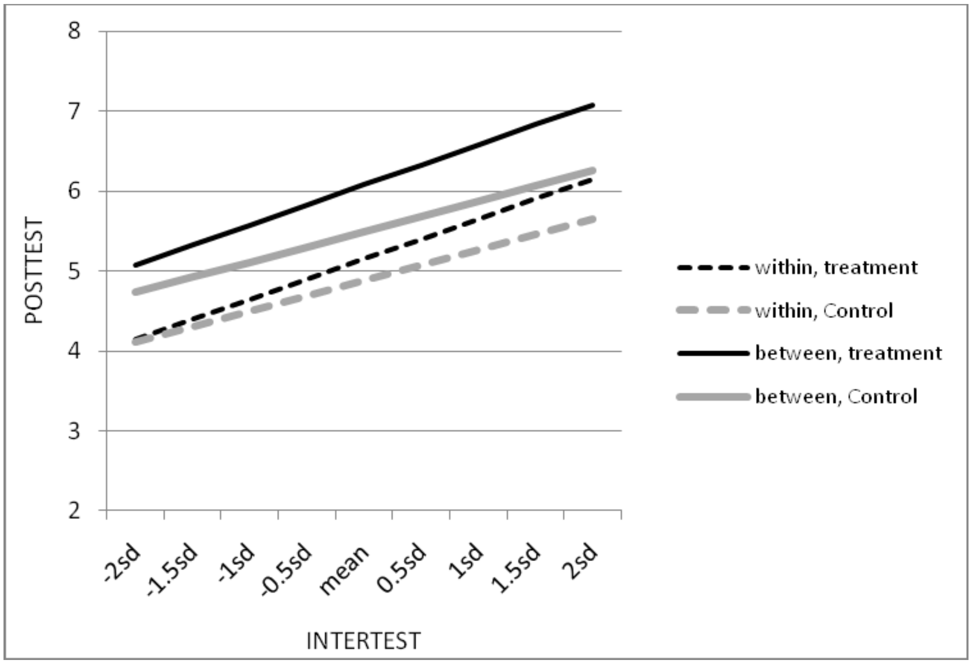


Figure 16. HM result for Grade 6 (RNE transfer measure subscore): Fitted relationships between interim measure RNE subscore and posttest RNE subscore by design and treatment condition.

HM result: Transfer measure subdomain (PA). There were five items assessing PA on the Grade 6 transfer measure. On these items, we found no statistically significant main effect of treatment or of design (see Appendix U: Subdomains as an Outcome Variable_Grade 6). The interaction term between treatment and design was also not significant; however, the interim measure test main effect was significant. The estimate of the interim measure mean was 0.44 with a *p*-value smaller than 0.001. This means that teachers with students who have a higher interim measure PA subscore tend to have higher mean scores on the transfer measure. In addition, the interaction effect between students' interim measure scores and treatment was also significant (estimate = 0.09, *p*-value = 0.01). As shown in Figure 17, for both W-S design and B-S design, only students in the POWERSOURCE© groups with higher pretest scores had higher transfer measure PA subscores than control group students.

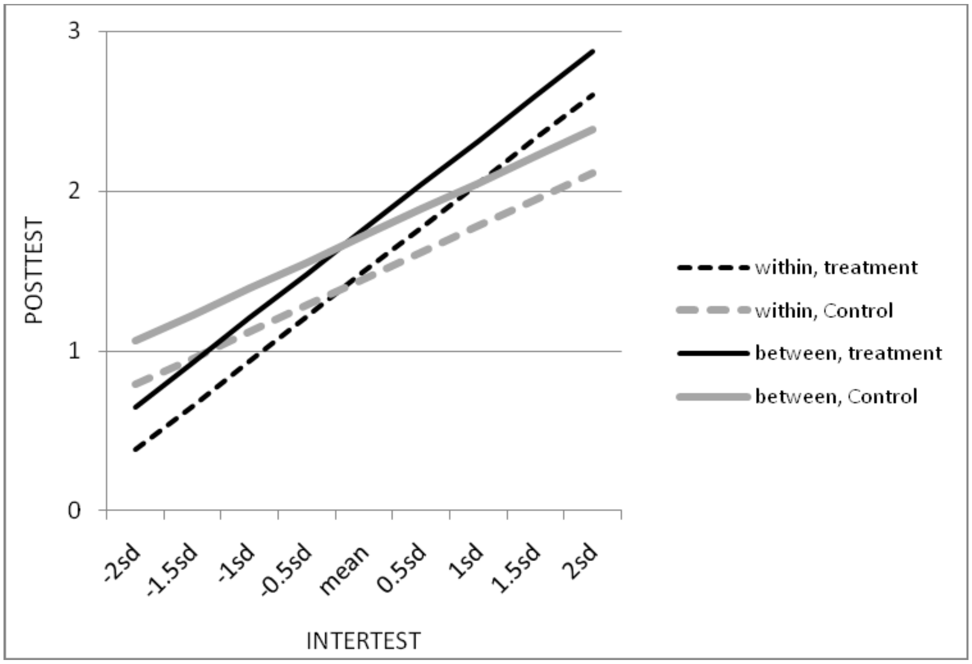


Figure 17. HM result for Grade 6 (PA transfer measure subscore): Fitted relationships between interim measure PA subscore and posttest PA subscore by design and treatment condition.

Grade 6 Interim Transfer Measure

In our second Grade 6 HM analysis, we used the interim measure as the outcome variable and pretest score as a covariate in our model. Appendix V (Interim Measure Estimates of Fixed Effect and the Variance Components) presents all the estimates of fixed effect and the variance components. The results show that the treatment main effect and the interaction between treatment and design are not statistically significant. However, we found a significant main effect of design with an estimate equal to 4.26 and *p*-value equal to 0.043. Figure 18 illustrates the fitted lines for the B-S design are higher than those for the W-S design. In addition, the pretest mean main effect is also significant (estimate=0.81, *p*-value<0.0001). This means that teachers who have higher average pretest scores tend to have higher average interim scores. Figure 18 presents the fitted relationships between pretest score and interim measure score. Figures 19 and 20 present separate lines in Figure 18 based on different design in the analysis. Figures 19 and 20 present model-fitted lines, respectively, for W-S design and for B-S design.

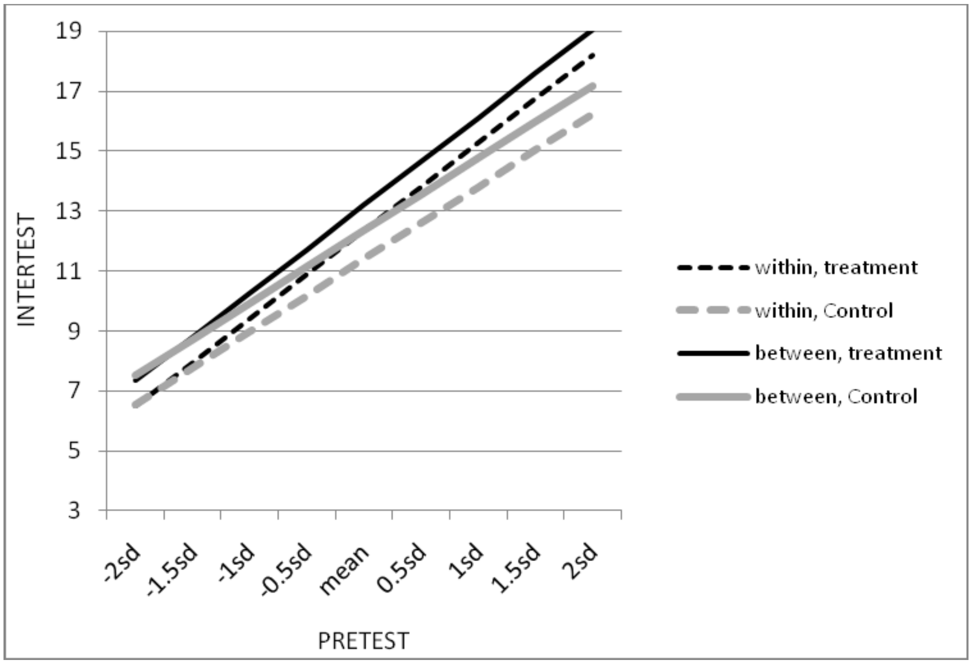


Figure 18. HM result (Grade 6 interim measure total score): Fitted relationships between pretest and interim measure by design and treatment condition.

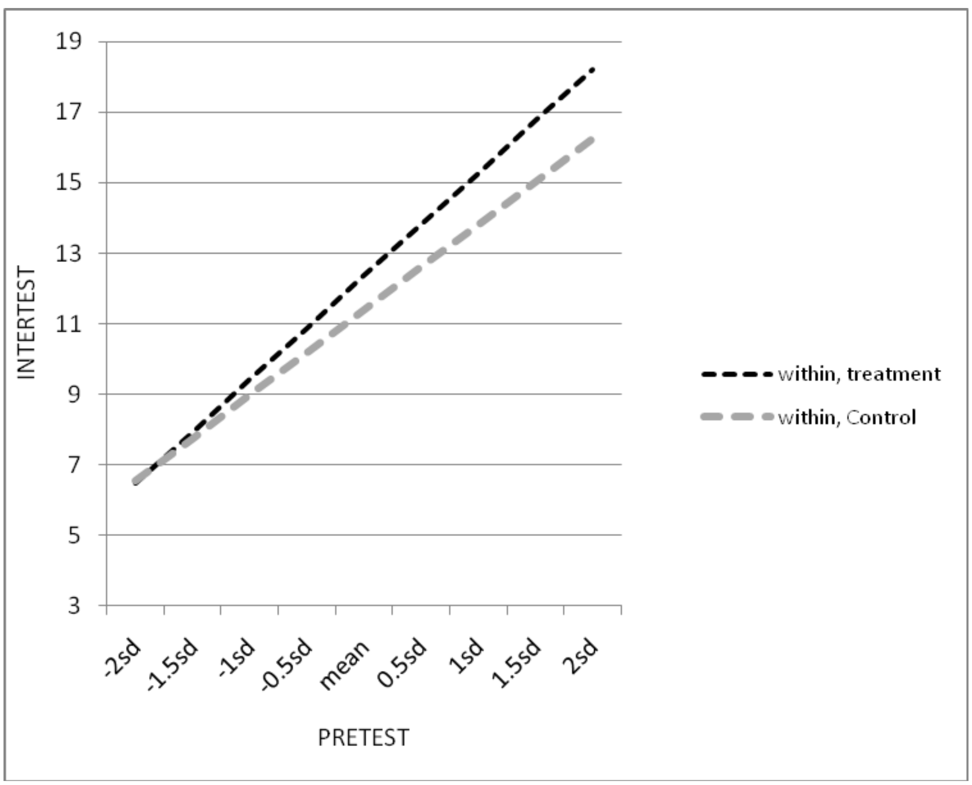


Figure 19. HM result (Grade 6 interim measure total score): Fitted relationship between pretest and interim measure for treatment conditions in within-school design.

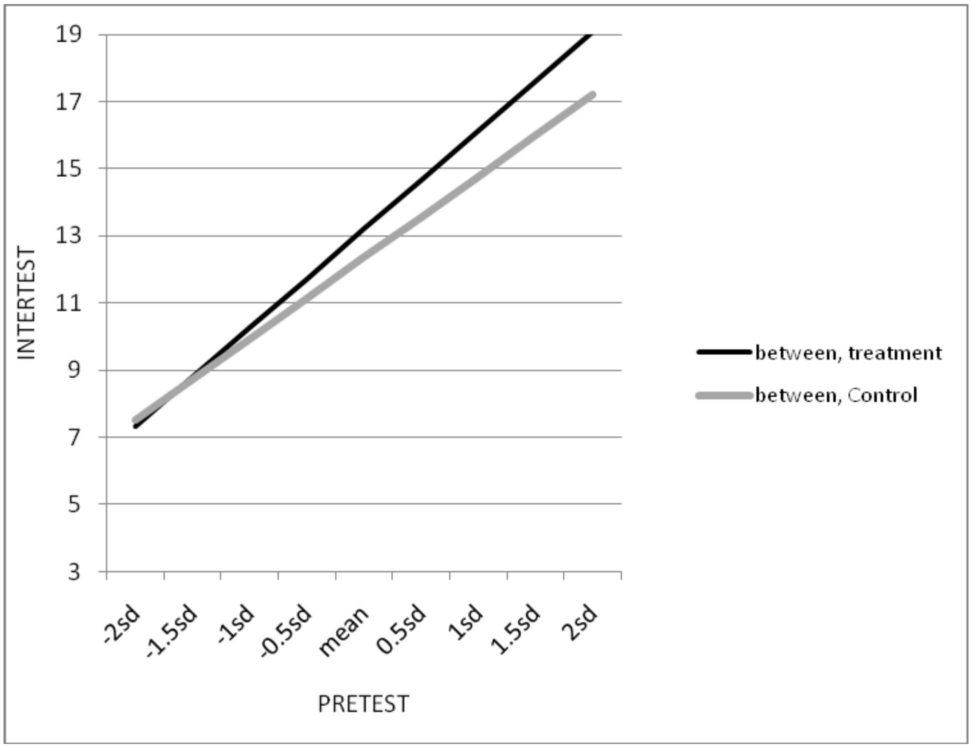


Figure 20. HM result (Grade 6 interim measure total score): Fitted relationship between pretest and interim measure for treatment conditions in between-school design.

Interim measure subscore (Grade 6 RNE). In the interim measure, there were 12 items related to RNE (see Appendix V for HM results). The result presents that the main effect of treatment was not statistically significant. In addition, the interaction effect of treatment and design was not statistically significant. However, the design main effect was statistically significant (estimate=2.40, *p*-value=0.037). This is clearly shown in Figure 21 in which the fitted lines for B-S design are significantly higher than those for W-S design. Additionally, the estimate of pretest mean is 2.21 and its *p*-value is smaller than 0.0001. This suggests that teachers who have higher pretest RNE mean scores tend to have higher interim RNE mean scores.

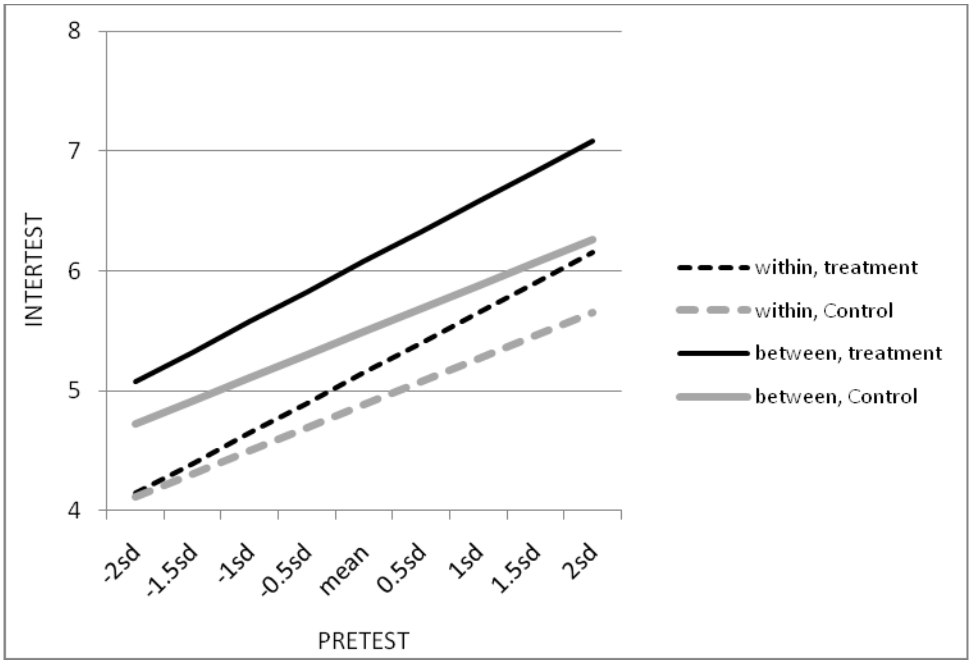


Figure 21. HM result for Grade 6 (RNE interim measure subscore): Fitted relationships between pretest RNE subscore and interim measure RNE subscore by design and treatment condition.

Interim measure subscore (Grade 6 PA). HLM results from 10 items assessing knowledge of properties of arithmetic are presented in Appendix V. We found that the main effect from treatment (estimate=0.99, *p*-value=0.024) was statistically significant. In addition, the interaction between pretest and treatment was also statistically significantly. Figure 22 clearly shows these effects. First, the fitted line for treatment groups in both W-S design and B-S design are significantly higher than those for control group. Second, the fitted lines for B-S design are higher than those for W-S design—especially for the treatment group. Third, the fitted lines for the POWERSOURCE© students are steeper than for control group students.

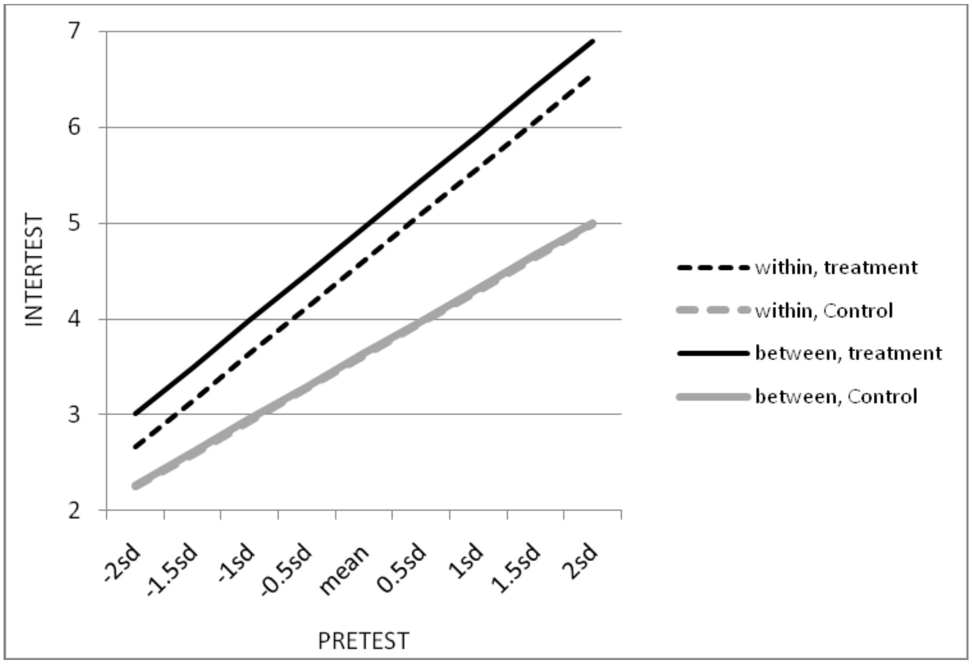


Figure 22. HM result for Grade 6 (PA interim measure subscore): Fitted relationships between pretest PA subscore and interim measure PA subscore by design and treatment condition.

HM results—Grade 6: Transfer measure (posttest) outcome (total score). As previously mentioned, we conducted three HM analyses. In the third HM analysis, we used the transfer measure as our outcome variable and the pretest score as a covariate. Appendix W presents all the estimates for fixed effects and the variance components in the model (see Appendix W: Estimates for Fixed Effects and the Variance Components in the Model_Grade 6 Transfer Measure). The results indicate that neither the main effect of treatment or of design were statistically significant. In addition, the interaction effect of treatment and design was not statistically significant.

Figure 23 presents the fitted relationship between pretest score and posttest score. Figures 24 and 25 separate the fitted lines in Figure 23 based on different school design. Though there was no statistically significant interaction effect found, statistical tests show that the POWERSOURCE© students who scored above 1.5 *SD* on the pretest, significantly outperformed the control group students on the posttest (in the B-S design).

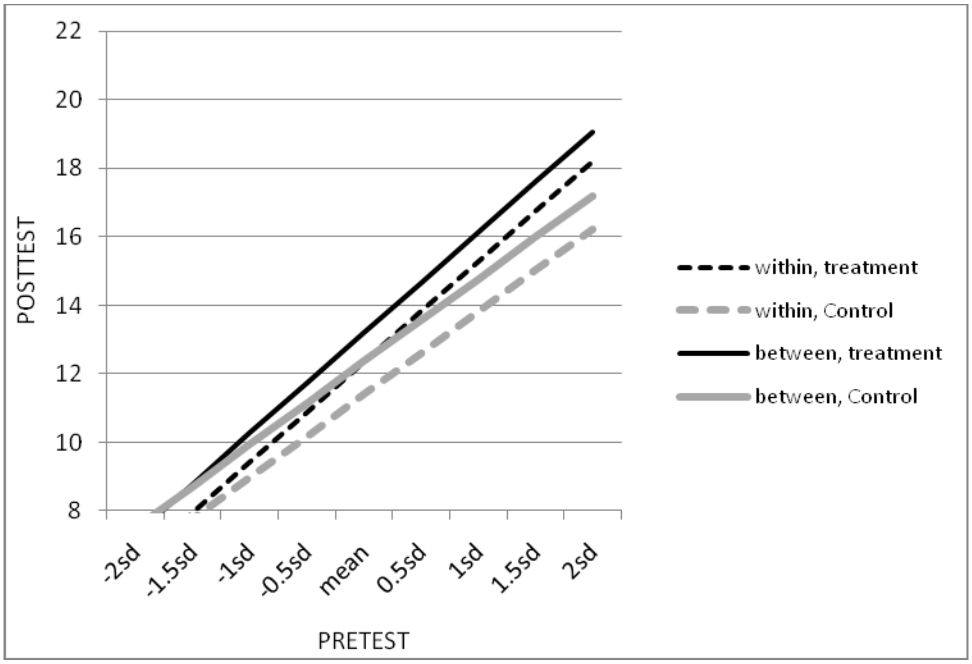


Figure 23. HM result (Grade 6 transfer measure total score): Fitted relationships between pretest and posttest by design and treatment condition.

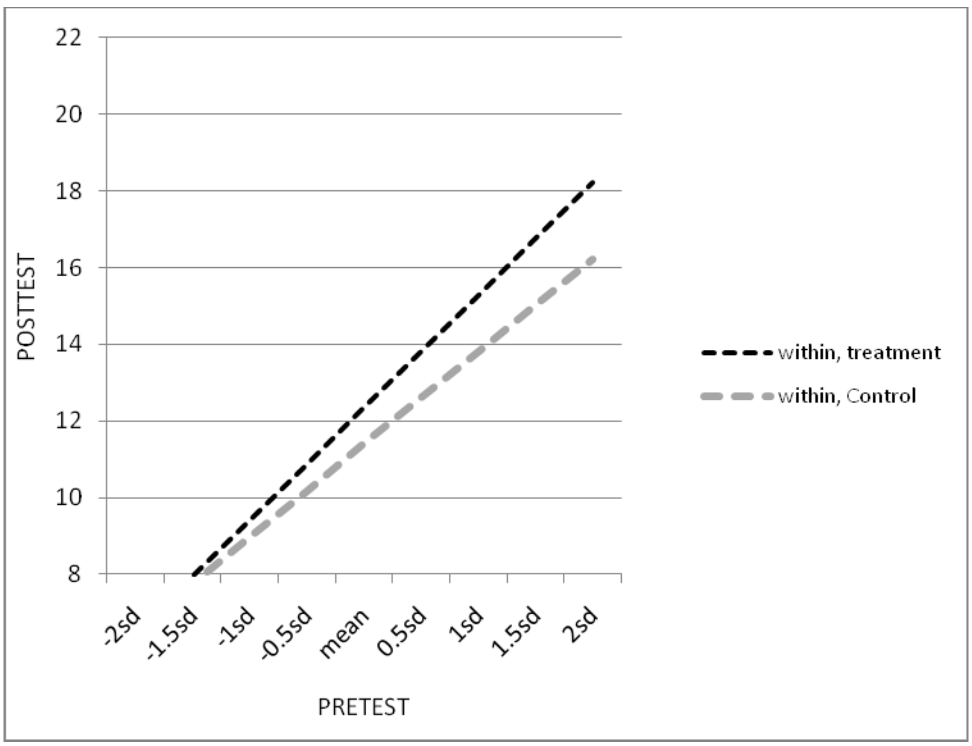


Figure 24. HM result (Grade 6 transfer measure total score): Fitted relationship between pretest and posttest for treatment conditions in within-school design.

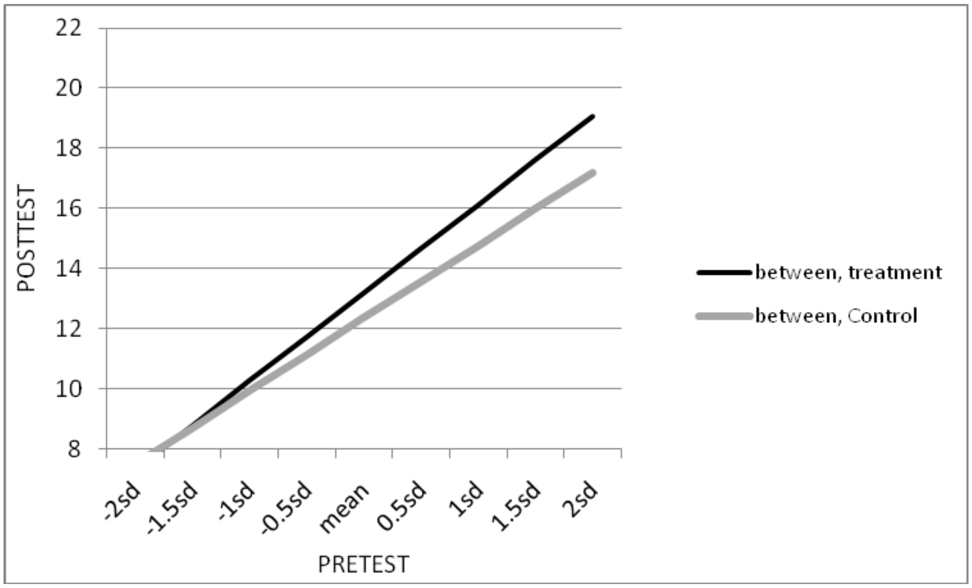


Figure 25. HM result (Grade 6 transfer measure total score): Fitted relationship between pretest and posttest for treatment conditions in between-school design.

HM result: Transfer measure subdomain (Grade 6 PA). There were five items related to PA domain on the Grade 6 transfer measure. We found that both the treatment main effect (estimate=0.59, *p*-value=0.012) and design main effect (estimate=1.77, *p*-value=0.001) were both statistically significant. In addition, the interaction between pretest and treatment was also statistically significant. Figure 26 shows these effects clearly. We can see that in Figure 26, the fitted lines for the treatment group are higher than those for the control group in both W-S design and B-S design. Additionally, the fitted lines for the treatment group are steeper than those for the control group. This indicates the interaction between pretest score and treatment effect. At last, the fitted line for B-S design (solid line) is higher than the fitted line for W-S design (dotted line) for both treatment group and control group.

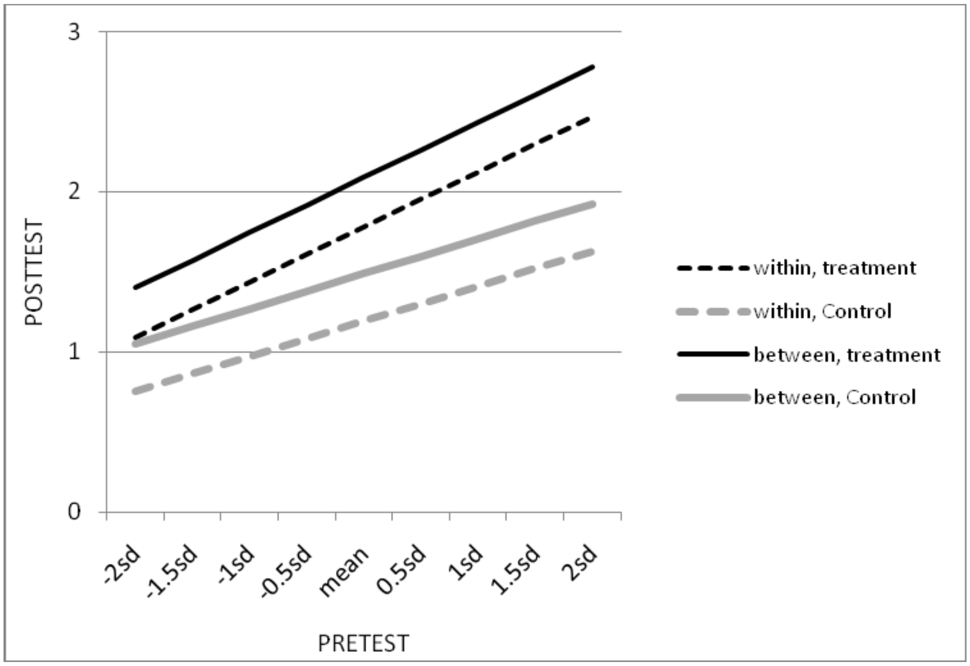


Figure 26. HM result for Grade 6 (PA transfer measure subscore): Fitted relationships between pretest PA subscore and posttest PA subscore by design and treatment condition.

Results from the analyses indicated that:

- A short amount of targeted intervention on key mathematical principles has an impact on student performance on transfer measures of related content. The POWERSOURCE© intervention had more impact on the relatively higher-performing students than the lower-performing students. In both grades, on most of the student measures, those students with higher initial pretest (or interim transfer measure) scores tended to benefit more from the treatment when compared to students with lower pretest scores.
- We saw a significant effect of POWERSOURCE© on the Grade 6 transfer measure items related to rational number equivalence concepts. In both designs, students in the POWERSOURCE© group outperformed the control group students on items associated with rational number equivalence; the effect was larger as pretest scores increased.
- Item analyses indicated difficulty ranges on the Grade 6 RNE items between b = -1.54 and b = +1.5, with the range of all the items for all domains b = -1.8 to b = 2.24. Thus, the RNE transfer measure items were spread evenly across the measure in terms of difficulty.
- There were also significant effects of the POWERSOURCE© treatment seen for PA items on the Grade 6 interim transfer measure and transfer measure, when using the pretest as a covariate.
- In two cases we did see a main effect of design. Previously, we saw no differences when we compared students in the B-S design treatment with the W-S design treatment. In this year, however, we saw a main effect of design on the Grade 6

interim transfer measure and also on the Grade 7 transfer measure items associated with solving equations. In both cases scores for the B-S design were higher than for the W-S design.

Implementation Study 2008-09: Professional Development and Teacher Measures

In the 2008-09 academic year, POWERSOURCE© teacher professional development expanded to include both Grade 6 and 7 teachers. As many of the 2007-08 Grade 6 teachers returned, this part of our professional development program focused almost exclusively on how to modify instruction based on the results of student performance on our formative assessments. On the other hand, the professional development program for the Grade 7 teachers blended content knowledge on key, foundational math concepts as they apply to the Grade 7 curriculum; on student misconceptions; and on instructional modifications likely to dispel those misconceptions. In three of the seven participating districts, we also provided a program of alternative professional development for teachers not assigned to the POWERSOURCE© treatment. We randomized teachers to treatment or control groups during their first year of participation, whether in Grades 6 or 7, and these teachers remained in their respective groups until the end of the study. A brief description of the POWERSOURCE© professional development at each grade is provided immediately below and is followed by a description of the Alternative Professional Development program.

2008-09 Grade 6 POWERSOURCE© Professional Development

The Grade 6 professional development meetings were designed with two points of focus: 1) reviewing student response data from the previous POWERSOURCE© unit (completed before the meeting), and 2) utilizing student response data from the previous year of the study to prepare for the upcoming POWERSOURCE© unit. These points of focus were targeted through activities designed to deepen teachers' content knowledge through the analysis of students' responses and response patterns, and through discussions of instructional implications, including teaching strategies. At the beginning of the meeting, the teachers were presented with the most recently completed student response data from the POWERSOURCE© unit. This activity had two phases: 1) comparing percentage of correct and incorrect answers overall, and 2) then looking more in-depth at student response patterns. We analyzed the frequency of correct and incorrect responses; solicited hypotheses from teachers about what these results might mean in terms of student learning; examined response/error patterns to further confirm (or challenge) these hypotheses; and then identified possible instructional responses based on these analyses.

After the student responses were discussed, the teachers were given a POWERSOURCE© assessment, corresponding with the upcoming unit that had been completed using the most frequent incorrect responses from their district (from the previous year). Teachers were also given a worksheet that asked them to identify student errors and possible misconceptions by analyzing the POWERSOURCE© assessment at item level and as a whole assessment. The worksheet also asked teachers to identify various components related to the assessment items, student responses, and potential student feedback, and had them develop a lesson plan that would effectively teach the unit content and instructional strategies that addressed potential student misconceptions identified.

The meeting concluded with a review of the materials related to the upcoming unit they would be receiving and providing time to answer any logistical questions. Professional development meetings were primarily attended by teachers who had been in the study for more than a year. The meetings were focused on allowing teachers to use their experience and available student data to think about how to modify instruction and to avoid or mitigate misconceptions and student errors. We also wanted to create the possibility for continued peer professional development among teachers without facilitation from CRESST members by increased teacher involvement and a more collaborative environment.

At the end of each professional development meeting, Grade 6 treatment teachers completed a confidential evaluation form that provided feedback on their opinion of the meeting. Aggregating across all POWERSOURCE© professional development meetings in 2008-09 academic year, 123 respondents provided feedback for *organization of the session*, *session presenter*, *ideas/activities presented*, *benefit to them (treatment teachers) as an educator*, *overall rating of the session* and were given response options of *Poor*, *Fair*, *Good*, and *Excellent*. For *organization of the session*, 74.59% marked *Excellent* and 25.41% marked *Good*. In another category, 72.36% of respondents thought the *session presenter* was *Excellent* while *27.64%* thought the *session presenter* was *Good*. The *ideas and activities* presented were rated as 66.67% *Excellent* and 32.52% *Good*. 70.49% of the respondents marked *Excellent* and 29.51% marked *Good* in rating the professional developments as a *benefit to them as an educator*. As for the *overall rating of session*, 67.20% of the respondents felt that the professional development meetings were *Excellent* and 32.52% *Good*.

2008-09 Grade 7 POWERSOURCE© Professional Development

As was the case for the Grade 6 teachers, each Grade 7 teacher in the POWERSOURCE© (treatment) group received slightly more than nine hours of professional development in small clusters (usually between 5 and 20 teachers) by district. These sessions were conducted largely outside of school hours at the district office or at one of the school sites within each district. The initial four hours of professional development was almost always done prior to the beginning of the academic year. During this four-hour block, teachers were introduced to the importance of key, foundational topics—referred to as "Big Ideas." The "Big Ideas" are foundational in that most of the content in Grades 6, 7, and 8 mathematics can be explained and developed from these concepts. In POWERSOURCE©, we focused on three "Big Ideas"—the Multiplicative Identity (as applied to RNE), the meaning of multiplication and other Properties of Arithmetic (as applied to Distribution), and the meaning of the equal sign (as applied to Solving Equations). The last half of the first professional development session, focused on Rational Number (Expression) Equivalence, including proportion with variables. The first session was followed up with three 90 minute sessions with the teachers (during after school hours at approximately two-and-a half month intervals). During the first 45 minutes of each of these follow-up sessions, teachers and researchers discussed student work (from the teachers' students) on the formative assessments associated with a particular foundational concept, possible misconceptions identified by those assessments, and possible instructional misconceptions. The last 45 minutes of each session focused on another single "Big Idea" (the meaning of multiplication or the meaning of equality) and its application, how that big idea would be developed from its nascent form into abstract concepts in algebra, and how the Big Idea could be appropriately taught and applied to Grade 7 subject matter. To aid teachers with their upcoming instruction on each foundational concept, teachers were given an instructional handbook on that concept during this portion of each session. The professional development integrated this instructional handbook (pedagogical content) with the conceptual development of each of the "Big Ideas" (content knowledge).

The treatment teachers then returned to their classrooms to develop their actual instructional plan and to provide dedicated instruction to their students on the applicable "Big Idea" for two class periods of approximately 40 minutes each. They also administered the *Checks for Understanding* associated with each unit during this time. After the initial presentation of a "Big Idea" to their students, teachers were encouraged to continue to use each "Big Idea" in other instructional units they developed during the year to teach other concepts.

Like Grade 6 treatment teachers, the Grade 7 treatment teachers also completed evaluation forms after each professional development meeting. Aggregating across all POWERSOURCE© professional development meetings in 2008-09 academic year, 104 respondents provided feedback for *organization of the session*, *session presenter*, *ideas/activities presented*, *benefit to them (treatment teachers) as an educator*, *overall rating* *of the session* and were given response options of *Poor*, *Fair*, *Good*, *Excellent*. For the organization of the session, 67.31% marked *Excellent* and 31.73% marked *Good.* In the category of *session presenter*, 78.64% of respondents marked *Excellent* while 21.36% marked *Good*. The *ideas and activities* category received 58.10% *Excellent* and 39.05% *Good*. 60.19% of the respondents marked *Excellent* and 35.92% marked *Good* in rating the professional developments as a *benefit to them as an educator*. The category, *overall rating* of *the session*, yielded 64.72% *Excellent* and 34.62% *Good*.

Teacher Measures

Teacher Knowledge Maps

The concept map of Leinhardt (see Chung and colleagues, 2006; Leinhardt & Smith, 1985), was used as a pre/post measure to track changes in the way teachers cognitively organized mathematical concepts. The knowledge map task required teachers to organize a list of mathematical concepts and math problems to create a web-like representation of how the concepts and the problems related to one another.

Teachers received a manila envelope containing the materials from which to construct the knowledge map. The envelope included a large piece of paper (approximately 2' X 3'), written instructions, and four sheets of peel off stickers. One sheet of stickers was a list of mathematical concepts (e.g., multiplicative inverse, additive property of equality, fractions). A second sheet of stickers consisted of math problems. The remaining two sheets of stickers were arrows—one set of black arrows with text labels that described relationships (e.g., "is a property of," "applies to") and the other were blue arrows with a square box in the middle.

Teachers were first asked to spend 15 minutes applying the concept stickers to the paper and then using the black text arrows to link the concepts together appropriately. Teachers were then asked to spend 15 minutes adding the math problem stickers to the map and linking the math problems to the concepts with the blue arrows. Teachers were instructed to fill in the box in the blue arrow with a 1 if understanding the concept was necessary but not sufficient to solve the problem. Teachers were asked to fill in the box with a 2 if understanding the concept was sufficient to solve the problem (for a full description of the knowledge map measure, see Chung et al., 2006).

Of the 131 teachers participating in the POWERSOURCE© program for 2008-09, 91 (68%) completed the knowledge map measure.

Scoring Knowledge Maps

Knowledge maps created by teachers were first saved electronically by a researcher duplicating the map onto CRESST Concept Mapper software designed for this purpose. A researcher recreated the map with this software and portion of maps entered were double-checked by a second researcher. The two researchers transcribing the maps met to resolve any discrepancies. Second, the Concept Mapper software assigned each knowledge map a score by comparing the teacher-submitted knowledge map to an “expert map” created by the researchers who conducted the professional development. This “expert map” was created by combining the individual maps researchers created in isolation from one another and from the teachers. Although these maps were identical on 98% of the relationships and concepts, researchers met to resolve the remaining discrepancies to form one “expert” standard map.

Agreement between each teacher map and the expert map was analyzed for similarity for 1) the degree to which both maps linked concepts together in the same way and 2) the degree to which both maps linked problems to concepts in the same way. Although several strategies exist for scoring concept maps, strict scoring was used in this case. That is, a match between teacher and expert maps required not only that two identical concepts (or two identical problem-concept pairs) be connected using an identical link, but that those concepts be connected with the identical link in the same direction. For example, {Additive Inverse} (a concept) is linked with {Property of Addition} (a concept) using the black text arrow link, {“is a”} going from {Additive Inverse} to {Property of Arithmetic}. This connection would then “read,” “Additive inverse is a Property of Addition.” If both expert and teacher maps displayed the linkage described earlier, this proposition would be scored as a match. However, if the same two concepts were linked with the same text arrow, but linking the concepts in the other direction (i.e., read, “Property of Addition is a additive inverse,”) the proposition would not be scored as a match. For problem-concept pairs, propositions were scored as exact matches if both the teacher and the expert map created a problem-concept link with the blue arrow and filled in the identical number, (i.e., either 1 or 2), depending on if the concept was necessary and sufficient, or necessary, but not sufficient, for solving the problem.

Analyses & Results

The extent to which teachers’ knowledge maps resembled the expert map was compared using paired t-tests within treatment and control conditions. These analyses were conducted for both concept links and problem-concept (i.e., problem) links. For concept linkages, although the mean match between expert and teacher maps within the treatment group was greater for the post measure (*M* = 1.12, *SD* = 1.68) than the pre measure (*M* = 0.81, *SD* = 1.39), this difference was not significant, $t(57) = 1.27$, ns. Within the control group, however, the mean match between the expert map and teacher map was significantly greater at the post measure (*M* = 1.16, *SD* = 1.66) than at the pre measure (*M* = .38, *SD* = .89), $t(36) = 2.40$, $p < .05$. For problem linkages, the mean match between the expert and teacher maps were significantly greater for the post measure (*M* = 5.63, *SD* = 4.21) than for the pre measure (*M* = 3.96, *SD* = 2.69), $t(54) = 2.63$, $p < .01$. The control group showed no gain in matches between expert and teacher maps from the pre (*M* = 4.32, *SD* = 4.37) to the post (*M* = 4.32, *SD* = 3.73) measure, $t(37) = 0$, $p = 1.00$.

Teacher Evaluation of Student Work

For the teacher evaluation of student work task, each teacher was asked to examine and respond to actual student work on an assessment of the Properties of Arithmetic unit of the POWERSOURCE© curricula. Specifically, teachers were instructed to respond to three sets of questions:

1. What is (are) the key principle(s) that these assessments address? Why do students need to understand this principle for Algebra I?
2. What inferences would you draw from this student's responses? What does this student know? What does this student not know?
3. If this student were in your class, based on your responses to questions 1 and 2, what would you do next in your instruction?

Teachers were asked to repeat this process three times, once for each instructional unit in the POWERSOURCE© curricula: (Properties of Arithmetic [Task 1], Solving Equations [Task 2], and RNE [Task 3]).

Analyses & Results

All teacher evaluations of student work (pre- and post- PD) were scored by four raters based on rubrics developed by Heritage and Vendlinski (2006). Given that previous analyses (i.e., Heritage, et al., 2006) found relatively little variation in Tasks 1 and 2, teacher Task 3 was examined for change in pre-to-post scores. Because the Teacher Evaluation of Student Work data were analyzed with a rubric resulting in ordinal data, differences in pre and post scores were evaluated with the Kruskal-Wallis test. This test does not allow for a comparison of means, but does allow researchers to compare pre and post scores when the dependent variable is ordinal. Please see Appendix X (Teacher Evaluation of Student Work-1) for descriptive results for each sub-question (i.e., a, b, and c) of Task 3.

According to MacDonald (2009), rather than testing the null hypothesis that the means or the median scores within any two data sets are the same, the Kruskal-Wallis test first substitutes the rank in the overall data set for each measurement value (the smallest with 1, etc.). The sum of ranks is calculated for each group before the test statistic H is calculated. H represents the variance within the ranks of each group, adjusting for the number of ties and is approximately chi-square distributed. Results from the Kruskal-Wallis test for none of the three sub-questions of Task 3 were significant (i.e., $p > .05$, in each case), suggesting that neither the POWERSOURCE© intervention (i.e., in the case of the treatment group) or an additional year of teaching middle school math (i.e., in the case of the control group) affected how teachers responded to this measure.

Results for the teacher knowledge measures are modestly encouraging. On one hand, teachers became more similar to POWERSOURCE© "experts" in how they think about the mathematical concepts students need to draw upon to solve math problems. That is, for the knowledge map measure, teachers exposed to POWERSOURCE© professional development in the treatment condition created maps more similar to the expert map at the end of the school year. This finding is particularly encouraging given POWERSOURCE©'s emphasis on moving away from procedural knowledge and rote memorization and toward more conceptual strategies to solve math problems. On the other hand, teachers in both the treatment and control groups related mathematical concepts to one another in a manner more similar to the POWERSOURCE© experts at the end of the school year. An overall recency effect may have been at work wherein all teachers, regardless of exposure to the professional development, were more adept at linking concepts to one another after a school year of teaching math than at the pre meeting after the summer vacation. The Teacher Evaluation of Student Work task also proved to be inconclusive. There was no item on which teachers' responses differed significantly between treatment or control either before or after the professional development intervention for the treatment group. It should be noted, however, that some teachers were missing data for the pre-assessment, which may have affected the results.

In conclusion, of the two teacher knowledge measures used, the teacher knowledge maps provided more clear information about how the POWERSOURCE© intervention affected teacher knowledge. This measure suggests that the way teachers think about math problems in relation to the mathematical concepts needed to solve them becomes more expert-like as a result of participation in POWERSOURCE© professional development.

POWERSOURCE© Implementation Study 2009-10

The core undertaking of our work during the 2009-10 school year was continuing with an extended, random assignment implementation study of the POWERSOURCE© program. In this year of the study we expanded the intervention from Grades 6 and 7 to Grades 6-8 in all participating schools. As with prior years, new teachers were randomly assigned to either POWERSOURCE© or control conditions with the ultimate goal of determining program impact on both students and teacher learning outcomes. Teachers continuing in the study for another year maintained their prior year's group status. The 2009-10 study was almost identical to the previous year's work, with a few minor changes:

1. An interim transfer measure was developed for use in Grade 7 (we had one in Grade 6 only in the previous year).
2. We created Grade 8 teacher instructional materials and *Checks for Understanding* assessments.
3. We modified the professional development sessions (in Grades 6 and 7) to focus more on interpreting student assessment data and less on teaching the big ideas.
4. We recruited an additional school district to replace a district not continuing with the study.

In the following section, we summarize changes made for the treatment and comparison conditions for the 2009-10 implementation study (including the alternative professional development offered to the control teachers); this is followed by brief descriptions of the design, measures and the analysis plan for the study. Additional details about the plan and its rationale can be found in the supplemental design report submitted to IES in August, 2007. The data collection for these activities is in its final stages.

Development of Grade 8 Materials

Pilot Testing of Grade 8 Items

Around 75 Grade 8 items have been developed and 40 items pilot-tested on 11 teachers in three schools. Using the same assessment model as the Grades 6 and 7 items, we have developed different types of assessment: basic computation tasks, partially worked problems, explanation tasks, word problems and problems involving graphics. Items were grouped together (within domains) to create the *Checks for Understanding* assessment forms. We used an overlapping design to allow us to compile item data and conduct IRT analyses on all items. The items we have pilot tested to date were compiled into 14 forms.

Pilot Testing Process

For pilot testing, the tasks described were assembled into forms that students should have been able to complete in about 15 minutes. This time frame was imposed by the districts we were working with for the study. They felt that any assessment longer than 15 minutes would be viewed by teachers as a test and would evoke complaints about the large amount of district testing. However, as it has turned out, the 15 minute time frame actually has a number of advantages in focusing teachers and students' attention on students' understanding of a single concept. Moreover, the shorter time frame encouraged deep assessment without being too intrusive or engendering teacher hostility because of intrusion into instructional time.

Every teacher participating in pilot tests received at least two different test forms—each focusing on the same big idea, with each form containing between 3-5 tasks. The forms were randomly assigned to students within classrooms; each teacher administered the assessments to all of their Grade 8 students. In all cases the first 2-3 items on the test forms were basic computation items. The subsequent items were open-ended explanation tasks, partially worked problems, word problems, or problems with a graphic prompt. Forms containing explanation tasks did not contain any other tasks besides the basic computational items.

All pilot data from the closed-ended responses were entered by a group of undergraduate and graduate student workers as well as by other CRESST staff. Three-point scoring rubrics were developed for the open-ended items.

From the set of Grade 8 items piloted in the 2008-09 year, we chose items to include on our *Checks for Understanding* forms and instructional materials for the extension of the POWERSOURCE© study in Grade 8. Items were analyzed using the same procedures outlined in previous reports (Baker, 2008, 2009). Several criteria (including confirmatory factor analyses, reliability analyses, and IRT analyses) were used to evaluate the items used in the pilot-testing phase.

Grade 8 Instructional Materials Development

Concurrent to the development of the *Checks for Understanding* items in Grade 8, we developed instructional materials to be used by teachers. We designed these materials for teachers to use as support when teaching each of the domains addressed in the study. Working with the expert teachers from one of our participating districts, we developed four Teacher Handbooks—each one closely aligned with the *Checks for Understanding* items in each domain (RNE, principles for solving equations, the distributive property, review and applications). Knowledge from teaching experience, research on teaching in these areas, and information gathered during the pilot testing year all played a role in developing these instructional materials.

Professional Development 2009-10

In the 2009-10 academic year, POWERSOURCE© teacher professional development expanded to include Grade 6, 7, and 8 teachers. Many of the 2008-09 Grade 6 and 7 teachers were returning participants to the study; thus, part of our program focused almost exclusively on how to modify instruction based on the results of student performance on our formative assessments. On the other hand, the professional development program for the Grade 8 teachers was a composite of content knowledge on key, foundational algebra concepts as they apply to the Grade 8 curriculum, on student misconceptions, and on instructional modifications likely to dispel those misconceptions. In one of the six participating districts, we also provided a program of alternative professional development for teachers not assigned to the POWERSOURCE© treatment. We randomized teachers to treatment or control groups during their first year of participation, whether in Grades 6, 7, or 8, and these teachers remained in their respective groups until the end of the study.

Each Grade 6, 7, and 8 teacher in the POWERSOURCE© treatment group received approximately nine hours of professional development in small clusters, usually between 1 and 20 teachers, by district. These meetings were conducted during after school hours at the district office or at one of the school sites within each district. The initial four hours of professional development was almost always done prior to the beginning of the academic year. During the four-hour block, teachers were introduced to the importance of key, foundational topics, referred to as “Big Ideas.” A brief description of the POWERSOURCE© professional development at each grade level is provided next. A description of the Alternative Professional Development program follows.

In the 2009-10 academic year Grade 6 and 7 teacher professional development meetings were structured very similarly—with the one exception that the Grade 7 meetings were facilitated by a CRESST research member while the Grade 6 meetings were facilitated by a school-affiliated personnel member. Prior to the 2009-10 academic year, the Grade 6 teacher professional development meetings were facilitated by CRESST members. In the 2008-09 academic year, CRESST had a vision to create the opportunity for continued peer professional development among Grade 6 teachers without the facilitation from CRESST research members. This vision was realized in the 2009-10 academic year when the Grade 6 teacher professional development meetings were facilitated by POWERSOURCE© coordinator participants in four of six participating districts; all of whom had been participants in the Grade 6 POWERSOURCE© treatment group in the previous years. One of the two remaining districts had only one Grade 6 participant who attended the Grade 7 teacher professional development meetings instead. The other remaining district facilitator was a math coach selected by the district's administration.

For both the Grade 6 and 7 teacher professional development meetings, the design was composed of two parts. The first point of focus was on reviewing the "Big Idea" of the current unit and the student response data from the 2008-09 academic year. The second focus of the meeting was on utilizing the student response data to modify instruction and develop teaching strategies to help students avoid and alleviate misconceptions and student errors prior to teaching of the unit. At the beginning of each meeting the teachers were presented with the 2008-09 student response data from the current POWERSOURCE© unit of focus. The teachers were asked to compare percentage of correct and incorrect answers overall, and then look more in depth at student response patterns. After analyzing the response patterns, the facilitators solicited hypotheses from teachers about what these results might mean in terms of student learning, examined response/error patterns to further confirm (or challenge) these hypotheses, and then identified possible instructional responses based on those analyses. Teachers were then given a worksheet that asked them to identify student errors and possible misconceptions by analyzing the POWERSOURCE© assessment at the item level and as a whole assessment. The worksheet also asked teachers to identify various components related to the assessment items, student responses, and potential student feedback. Furthermore, the worksheet asked teachers to develop a lesson plan that would effectively teach the unit content and instructional strategies that addressed potential student misconceptions identified. Each meeting concluded with the review of materials related to the upcoming unit teachers would be receiving and providing time to answer any logistical questions. At the end of each meeting, teachers were also asked to provide feedback on their opinion of the meeting by completing a confidential evaluation form.

Teacher Measures

All Grade 8 teachers and Grade 6 and 7 teachers new to the POWERSOURCE© program (in both treatment and control groups) completed two measures of teacher knowledge prior to any professional development. After completion of treatment and control professional development activities, all participating teachers—new and returning—completed teacher knowledge measures again. Teachers in the treatment group completed the measures of teacher knowledge at the final professional development meeting of the school year. Teachers in the control group completed the measures on their own and returned them to their school or district coordinator.

Teacher Implementation of Formative Assessment

As in previous years, both Group 1 (POWERSOURCE© treatment group) teachers and Group 2 (control group) teachers were asked to fill out the Teacher Implementation Surveys. Surveys for each of the four domains were sent to all teachers after Group 2 teachers taught each domain. The Group 1 survey questions address a variety of issues—including how the teacher handbooks lessons were used, how the *Checks for Understanding* were utilized as formative assessment tools, and any difficulties or concerns that arose when using POWERSOURCE© materials. The Group 2 survey was shorter and asked teachers if and how they used any assessments for teaching each domain's material. Upon return of the survey forms, data was analyzed to identify any trends.

Website Resources

The website that was created to provide participating teachers with resources to assist and enhance their experience while participating in the POWERSOURCE© study has been updated. Website content for Grades 6 and 7 have been updated and website content for Grade 8 has been added. Users of the website are able to access information and materials that range from logistical information concerning the organization and use of materials to the research behind the study content.

The website address is www.cresstpowersource.com. Members can access the site by entering a Member ID number. Upon entering the site, users are presented with a brief overview of the POWERSOURCE© study and links to download study background and implementation surveys, as well as a content map of the three Big Ideas. Users are also given the option to view 1 of 4 portals representing the domain units of the study. Once a unit is selected, options for viewing information regarding the Big Ideas, Teaching Resources, and Teacher Handbook for that unit are offered.

Having direct access to materials and resources on demand provides more flexibility to POWERSOURCE© users and decreases a participant's level of dependence on us for materials. This, along with a more collaborative professional development setting, creates the possibility of a sustainable professional development program within participating districts.

Sample and Design

Six districts participated in the random assignment implementation study in 2009-10. As described earlier, we used two designs (within and between school) based on district needs and configuration. Ultimately, three schools used a between-school design and three other schools used a within-school design. The total number of participants in the study in 2009-10 is shown in Tables 12, 13, and 14.

Table 12

Sample Distribution by School District ('09-'10 school year) Grade 6

District	*N* of students	*N* of teachers	*N* of schools	Design
AZ-1	477	7	3	BS
CA-1	1048	17	3	WS
CA-2	731	10	3	WS
CA-3	182	6	4	BS
CA-6	1343	32	11	BS
CA-7	35	1	1	WS

Table 13

Grade 7 Sample Distribution by School District for the 2009-10 School Year

District	*N* of students	*N* of teachers	*N* of schools	Design
AZ-1	228	3	2	BS
CA-1	591	7	2	WS
CA-2	822	7	3	WS
CA-3	307	3	2	BS
CA-6	764	13	5	BS
CA-7	212	6	4	WS

Table 14
Grade 8 Sample Distribution by School District for the 2009-10 School Year

District	*N* of students	*N* of teachers	*N* of schools	Design
AZ-1	419	6	3	BS
CA-1	580	6	2	WS
CA-2	1062	6	3	WS
CA-3	-	-	-	-
CA-6	1251	11*	5	BS
CA-7	453	6	6	WS

Note. *One teacher had a substitute for a while. Both teachers were counted together as one teacher.

Transfer Measure

Over the course of the 2009-10 school year, the treatment group students in our POWERSOURCE© study received instruction and formative assessments (*Checks for Understanding*) on the four POWERSOURCE© domains. Also included in the study were a control group of students who received their regular instruction.

We hypothesized that students in the POWERSOURCE© group would possess a better understanding of the basic mathematical principles contained within each domain. We also hypothesized that students would be able to apply concepts they had learned, solve complex problems, and transfer the principles covered by the POWERSOURCE© domains. For example, having received instruction and formative assessment on RNE, students should understand the multiplicative identity principle and be able to use it to: a) demonstrate that a set of rational numbers are equivalent, b) find equivalent fractions, c) find missing numbers in proportions, and d) solve proportional reasoning problems. In order to answer these questions we used a transfer measure (posttest) to compare the POWERSOURCE© and control groups on novel items related to our four POWERSOURCE© domains.

Grade 7 Interim Transfer Measure

In an effort to gather more student outcome data, we designed an interim transfer measure to be given to students after completion of the first two POWERSOURCE© domains (i.e., PA and RNE). In 2008-09 we created an interim transfer measure for Grade 6; in 2009-10 we created one for Grade 7. We created a 16 item test form with ~ 20% of the items requiring students to explain a concept in their answer. We selected two items per domain from the pretest (of medium difficulty) and changed the numbers in the items. The remaining items were taken from the transfer measure and again were modified to include different numbers, and/or situations. Items selected for the interim transfer measure had a range of difficulty from b=1.26, *p*-value = .15, to b= -1.63, *p*-value = 0.91 (see Appendix Y: Grade 7 Interim Transfer Measure 2009/2010).

Grade 7 Transfer Measure Revision

Based on our item analyses we modified the Grade 7 transfer measure. Since the amount of information we tend to get from an extended response item is greater than for a multiple choice item—the more extended response items on the test, the fewer multiple choice items are required. We modified some of the existing item formats from multiple choice to extended response (either short answer or explanation). We removed one question (which was answered correctly by 91% of the students); changed three item formats; and deleted three items all with $p > 0.97$ (meaning that at least 97% of students were answering incorrectly). We removed these three items and instead added them to the Grade 8 transfer measure. Based on the data collected in 2008-09, we re-organized the items to reflect item difficulty with items ordered from easiest to most difficult (see Appendix Z: Grade 7 Transfer Measure, Revised Version).

Grade 8 Pretest

The Grade 8 pretest was developed using similar procedures as the Grades 6 and 7 pretests. The pretest consisted of items used previously on other CRESST projects, items adapted from the Grade 7 California Standards Test released items, and items created by us specifically for this project. Items reflected precursor math content for the three Big Ideas being covered by the *Checks for Understanding* assessments. Each POWERSOURCE© domain contained 7-8 items on the pretest associated with relevant precursor knowledge, which yielded a total of 29 items (see Appendix AA: Grade 8 Pretest).

Grade 8 Transfer Measure

The Grade 8 transfer measure was developed using similar procedures as the Grade 6 and Grade 7 transfer measures. Items were selected from TIMSS, NAEP, the QCA Key Stage 3 exam, PISA, and benchmark tests used in one of our pilot districts (see Appendix H for sources of all items). Items were selected based on their relevance to the POWERSOURCE© domains and their appropriateness for a transfer task (related to POWERSOURCE© content but not exact replicas of item types used in the *Checks for Understanding*). An initial set of items were selected and narrowed down to a final pool of 21 items. Of these items 12 were multiple choice and the rest were either short answer or explanation tasks, or a combination of both types. Items were selected based on their representation in the CA state standards and relevance to POWERSOURCE© items (see Appendix BB: Grade 8 Transfer Measure 2009/2010).

Observation and Interview Study

As part of the 2009-10 POWERSOURCE© implementation research, we conducted classroom observations and teacher interviews. This followed pilot studies in 2007-08 and 2008-09 of the interview and observation measures. These observations/interviews had several inter-related purposes: First, they provided first-hand data, to supplement the self-report surveys about how teachers were using POWERSOURCE© materials in the classroom (including assessments, instructional activities, and learning supports). Second, they provided a more open-ended opportunity for teachers to provide feedback about their POWERSOURCE© implementation and professional development experiences. Finally, it allowed us to pilot instruments and methodology for scaled up qualitative data collection in the remaining years of the study.

Five trained CRESST observers/interviewers visited Grade 6 and 7 teachers at five schools within three local school districts to observe them as they taught PA and SE lessons from POWERSOURCE© teacher handbooks. Six lesson 1 PA classes, six lesson 2 PA classes, 10 lesson 1 SE classes, and nine lesson 2 SE classes were observed. The purpose of these visits was to monitor how teachers were implementing the POWERSOURCE© program. These classroom observations helped us ensure that teachers' self-reports of their classroom activities were accurate. A total of 15 interviews were conducted after teachers taught lesson 1 or lesson 2. These one-on-one interviews were conducted to gain more insight into how the POWERSOURCE© program impacted teachers' use of formative assessment when teaching mathematics.

Student Interviews

As part of the 2009-10 POWERSOURCE© study, we conducted a series of student interviews with ten treatment group students. The objectives of this study were to 1) interview students while they complete mathematics assessments and discuss their rationale for answering questions the way they did; and 2) to observe how students explained their reasoning on problem-solving. We had already looked at student responses patterns and made inferences as to why students may have answered a question in a particular way; yet, this discussion with students gave us a truer sense of how they solved problems and justified their answers. One limitation of using paper and pencil explanation tasks with students was that given their lack of familiarity and experience with such tasks, the type of information put forth in student responses was sometimes sparse. Thus we were forced to infer what a student was thinking or trying to communicate in a response to a more complex explanation task. By sitting down with a subset of students and having them explain how they would solve a problem (or why they have selected a particular response), we hoped to gain a deeper and richer insight into how students solve problems as well as the depth of their conceptual understanding of the topics presented within the *Checks for Understanding* assessments.

Ten student interviews were conducted, each lasting between 20-30 minutes. The think-aloud protocols added additional insight into how students thought about the problems and helped us gain insight into how students solved math problems.

Supplementary Research Activities

Following is a brief update of a supplementary strand of work undertaken as part of CRESST's activities during the 2009-10 school year. This work includes an investigation of district contexts for assessment.

Use of Interim Assessment Data/District Contexts

This research activity takes a broader contextual approach to interim assessment use by examining the ways in which middle school mathematics teachers use the data provided by POWERSOURCE© (and other types of interim assessments) and how the features of the assessments are related to data use. The project was conducted simultaneously in three sites—Central Colorado (coordinated by Lorrie Shepard, CU Bolder), Southern California (coordinated by Brian Stecher, RAND), and Northern California (coordinated by Hilda Borko, Stanford).

We are now in the process of writing up the final results of the "Use of Interim Assessment Data" component of the POWERSOURCE© study. This component examined the ways in which middle school mathematics teachers used the data provided by POWERSOURCE© (as well as other types of interim assessments) and how the features of the assessments were related to data use. For the past year, the project has been analyzing data drawn from eight districts located in three geographic regions—Central Colorado, Southern California, and Northern California, and we are currently drafting our final deliverable.

The deliverable will be structured as three interrelated reports. The first report will focus on the districts that were using assessment systems that we characterized as "interim assessments." These districts generally adopted externally developed assessments that measure students' skills a few times each year and provide feedback on mastery or progress toward annual standards. Using detailed examples from our interviews and collected classroom artifacts, the report describes the range of information teachers learned from the assessments and how the information is used in their instruction. It also examines the relationship between the features of the assessment and the patterns of use. The second report examines those districts that were engaged with more “formative” assessment systems. These efforts, which took a variety of forms, focused more on providing information for ongoing instructional planning. The formative assessment report has a similar structure, describing the formative system and the assessment context in each district and then focusing on the kinds of information teachers obtained from the assessments, how they use this information instructionally, and whether the features of the assessment were related to use. The final report, which will cover all sites, focuses on “district intent,” that is, the reasons districts’ administrators espoused for adopting each system, the manner in which they expected the assessments to be used, and whether those expectations were fulfilled.

Analysis of Student Understanding of Mathematical Equality

Ubiquitous in mathematics—understanding of the equals sign is crucial for understanding many mathematical topics (including algebra). Many studies (mostly conducted with elementary students) have shown that there is shaky understanding (at best) when it comes to the concept of the equals sign (see Baroody & Ginsburg, 1983; Kieran, 1981; Carpenter, Franke, & Levi, 2003; McNeil & Alibali, 2005). In order to further investigate student understanding of equality, we conducted a small study with five Grade 6 teachers and eight Grade 7 teachers. Students from ten Grade 6 and twenty two Grade 7 classes participated.

One goal of this pilot study was to determine whether the way students define an operator (in this case the equal sign) relates to student performance on problems that involve use of this operator. Specifically, what is the relationship between how students define the meaning of the equal sign and their ability to write an equation from a word problem, solve a simple equation, and identify the equal sign? Also of interest is how scaffolding relates to students’ ability to translate a word problem into an equation. Finally, we wanted to investigate how the complexity of an equation relates to how students define the equal sign. Data for this small study are still being analyzed; hence, results are not finalized.

Leadership

A core, planned set of supplemental activities is the leadership strand of work. Our leadership activities intend to support states and districts in their desire to develop coherent instructional programs to engage in standards-based reform; this work focuses on two areas. First, it looks at the collaborative development of methodology and annotated examples that practitioners and contractors can use to align instruction and assessment developmentally—with key priorities for student capability in mathematics as well as with standards. The methodology seeks deeper understanding and communication of the learning demands, inherent standards, and the developmental progressions that are essential to accomplishing key standards. The methodology lays out a systematic framework describing these learning demands and progression, rather than simply working backward from one existing test. Products from the proposed effort will include software with embedded tutorials for conducting alignment analyses, paper and poster illustrations, and the results of workshops and webinars held with experts in math, math education, test developers, and other researchers—as well as with the practitioner and policy communities.

Formative Assessment Group

Recently, several CRESST researchers have formed a working group to define assessment quality as it applies in its broadest sense to formative assessment. While there is a growing body of empirical research which examines the benefits of formative assessment to student learning (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black &Wiliam, 1998a: Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002), this literature has mainly addressed the process of formative assessment. The formative assessment process is characterized as continuous—carried out during the course of teaching and learning to provide feedback to teachers and students to improve teaching and learning. Discussions of assessment quality are less prominent in the formative assessment literature. The goal of our CRESST working group is to establish a framework for considering formative assessment quality.

Prior work (Phelan, et al., 2009) has shown us that we can establish technical quality of formative assessments; moreover, data suggest that relatively brief formative assessments that focus on key conceptual domains can provide reliable and useful information on students' levels of understanding and possible misunderstandings in the domain. These results, however, are just part of the evidence needed to validate the tasks as formative assessments. Other evidence includes information on the sensitivity of the tasks to instruction (so that they are not just measuring, for example, general intelligence or mathematics achievement) and the utility of the tasks in a formative assessment system, which means that teachers are able to use the assessments to make more informed and effective instructional decisions.

Formative assessment can include questioning, discussions, tasks, representations, and explanations. Whatever the assessment strategy, formative assessment is not "formative" unless action is taken on the basis of the evidence the assessment provides. The action is intended to lead to further learning and thus to have positive consequences (e.g., Moss, 2003; Stobart, 2006). However, positive consequences hinge directly on teachers' abilities to interpret the evidence and to know what action to take as a result. Effectively interpreting and using evidence is dependent on teachers' domain and pedagogical content knowledge. As a step toward developing an assessment quality framework, our working group is currently engaged in analyzing the range of teacher knowledge needed for different types of formative assessment. A structure for our analysis is shown in Table 15.

Table 15

Structure for Analyzing Teacher Knowledge

Assessment cycle	Cognitive demand	Formative assessment	Type of evidence	Teacher knowledge	Teacher action
Length of the assessment cycle – e.g., 5 minutes, 1 lesson, 1week	Cognitive demand of the assessment task	Example of a formative assessment linked to cycle and cognitive demand	Evidence provided from the formative assessment	Knowledge needed to interpret the evidence i.e. what does this tell me about current learning status?	Desirable action to move learning forward

Although we are in the early stages of this analysis, we anticipate it will yield insights into some key considerations of assessment quality, which will inform the next step of our work toward establishing a framework for assessment quality related to formative assessment.

Future Plans

Analysis Plan

Plans for analyzing data from the 2009-10 year include:

- POWERSOURCE© was implemented in Grades 6, 7, and 8 during the 2009-10 school year. Since the study has implemented a similar design and instrumentation described earlier, we will basically utilize a similar statistical model and analyses plan to the one employed in the 2008-09 study. Note, however, that the analyses will be conducted separately by grade level.
- One of the key distinctions for the 2009-10 data analyses is that we will explore some possibilities of examining the student growth trajectory during a year with three time-series measures: pretest, interim transfer measure, and post transfer

measure. We will address the following interesting questions: What does the growth trajectory look like? How much variability in the student growth trajectory is observed? Does the rate of growth differ between the control group and the POWERSOURCE© group?

- Given that this is the third year of the POWERSOURCE© large-scale implementation, we are keenly interested in differential/cumulative effects of the POWERSOURCE© experience--both in students and teachers. For example, we expect there to be a significant impact on number of years a teacher has been involved in POWERSOURCE©. We hope to see that teachers will become more proficient in their subject matter knowledge, more skilled in their formative use of assessment, and better equipped to focus their instruction on key ideas; as a result, teachers will be more effective in helping students to improve their understanding of key algebra principles.

Alongside analyzing our remaining data we are currently exploring avenues through which we might extend our POWERSOURCE© work into more content areas and age levels. Our results, thus far, have been promising and we would like to explore the possibility of expanding our intervention to perhaps reach some of the students we have not had significant impact on thus far. One such project involves the combination of the power of research in cognitive science, mathematics teaching and learning, measurement theory and formative assessment with technology to develop a stand-alone intervention to prepare Grade 7 and 8 students for success in whatever algebra course they are to begin. This intervention will feature computer-based "smart assessments" and technology to deliver tailored instruction for use as a daily, one hour supplement to a six-week summer school remedial math program (for example). Leveraging our prior work in formative assessment and technology-based instruction in middle school mathematics (see Choi, 2008; Phelan, Niemi, & Vendlinski, 2008), the program focuses on efficiently developing student understanding of critical foundational ideas that are key precursors to readiness for an understanding of algebra.

References

Albright, J. J., & Park, H.M. (2009). *Confirmatory factor analysis using Amos, LISREL, Mplus, and SAS/STAT CALIS.* Working paper, The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.), *The psychology of learning and motivation, vol. 8.* London, UK: Academic Press.

Baker, E. L. (2006, June). *National Center for Research on Evaluation, Standards and Student Testing. The development and impact of POWERSOURCE©* (Deliverable to IES). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L. (2007, June). *National Center for Research on Evaluation, Standards and Student Testing. The development and impact of POWERSOURCE©* (Deliverable to IES). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L. (2008, June). *National Center for Research on Evaluation, Standards and Student Testing. The development and impact of POWERSOURCE©* (Deliverable to IES). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L. (2009, June). *National Center for Research on Evaluation, Standards and Student Testing. The development and impact of POWERSOURCE©* (Deliverable to IES). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Ball, D. L., & Bass, H. (2001). What mathematical knowledge is entailed in teaching children to reason mathematically? In National Research Council, *Knowing and learning mathematics for teaching: Proceedings of a workshop* (pp. 26-34). Washington, DC: National Academy Press.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching, (4th ed.).* New York, NY: Macmillan.

Baroody, A. J., & Ginsburg, H. P (1983). The effects of instruction on children's understanding of the "equals" sign. *The Elementary School Journal*, *84*(2), 198-212.

Black, P., Harrison C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*, 9-21.

Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, *5*(1), 7-74.

Black, P. J., & Wiliam, D. (1998b*). Inside the black box: Raising standards through classroom assessment.* London, UK: School of Education, King's College.

Black, P.J., & Wiliam, D. (1998c). Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan*, *80*(2), 139-148.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1-12.

Brown, A., Bransford, J., & Cocking, R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school, (Expanded ed.).* Washington, DC: National Academy Press.

Brown, R. S., & Niemi, D. N. (2007). *Investigating alignment of high school and community college assessments in California.* San Jose, CA: National Center for Public Policy in Higher Education.

Carpenter, T., & Franke, M. (2001). Developing algebraic reasoning in the elementary school. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), *Proceedings of the 12th ICMI Study Conference, 1* (pp. 155-162). Melbourne, Australia: The University of Melbourne.

Carpenter, T. P., Franke, M.L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Portsmouth, NH: Heinemann.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem solving transfer. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(6), 1147-56.

Chi, M.T. H., & Bassok, M. (1989). Learning from examples via self-explanations. In L.B. Resnick (Ed.), *Knowledge, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5, 121-152.*

Choi, K. (2008). *Multi-sites multiple-cohorts growth model with gap parameter: 4-level latent variable regression hierarchical model*. Paper presented at the annual Institute of Education Sciences conference, Washington, DC.

Chung, K. W. K., Baker, E. L., Brill, D.G., Sinha, R., Saadat, F., & Bewley, W. L. (2006). *Automated assessment of domain knowledge with online knowledge mapping.* (CSE Tech Report 692). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Delacruz, G. C., Chung, G.K.W. K., Heritage, M., Vendlinski, T., Bailey, A., & Kim, J. 0. (2007, April). *Validating knowledge elicitation techniques: Examining the relation between measures of content knowledge and knowledge of teaching algebra.* Paper presented at the annual National Council on Measurement in Education meeting, Chicago, IL.

Dimitrov, D. M. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement*, *27*, 440–458.

Ericsson, K. A. (2003). The search for general abilities and basic capacities: Theoretical implications from the modifiability and complexity of mechanisms mediating expert performance. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Perspectives on the psychology of abilities, competencies, and expertise* (pp. 93-125). Cambridge, UK: Cambridge University Press.

Ericsson, A. K., & Simon, H. A. (1984). *Protocol analysis. Verbal reports as data.* Cambridge, MA: MIT Press.

Fuchs, L.S., Fuchs, D., Finelli, R., Courey, S. J., & Hamlett, C. L. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41*(2), 419-445.

Haverty, L. (1999). *The importance of basic number knowledge to advanced mathematical problem solving.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Heritage, M., & Vendlinski, T. (2006). *Measuring teachers' mathematical knowledge* (CSE Tech. Report 696). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Heritage, M., & Yeagley, R., (2005). Data use and school improvement: Challenges and prospects. In Herman, Joan L. and Haertel, Edward H. (Eds.). *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education, part 2.* Oxford, UK. Blackwell Publishing.

Herman, J. L., & Baker, E. L. (2006). Making benchmark testing work for accountability and improvement: Quality matters. *Educational Leadership, 63*(3), 48-55.

Herman, J. L., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation* (CSE Tech. Report 535). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, J., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2005, April). *The nature and impact of teachers' formative assessment practices.* Paper prepared for the annual American Educational Research Association conference, Montreal, Canada.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York, NY: Macmillan.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review, 36*, 28-42.

Judd, C. H. (1936).*Education as the cultivation of higher mental processes*, New York, NY: Macmillan.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*, 317-326.

Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3-38). New York, NY: Macmillan.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics—Report of the Mathematics Learning Study Committee*. Washington, DC: National Research Council, National Academy Press.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Leinhardt, G. & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, *77*(3), 247-271.

McDonald, J. H. (2009). *Handbook of Biological Statistics* (*2nd ed.*, pp. 165-172). Baltimore, MD: Sparky House Publishing.

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

Mayer, R. E. (2003). *Learning and instruction.* Upper Saddle River, NJ: Merrill Prentice Hall.

McNeil, N. M., & Alibali, M. W. (2005). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition and Development, 6*(2), 285-306.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97* (1), 117-128.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, *22*, 13–25.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 C.F.R. (2002).

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, *1*, 117-175.

Pawley, D., Ayres, P., Cooper, M., & Sweller, J. (2005). Translating words into equations: A cognitive load theory approach. *Educational Psychology, 25,* 75-97.

Phelan, J., Kang, T., Niemi, D. N., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessments in middle school mathematics* (CRESST Report 750). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Phelan, J., Niemi, D., Choi, K., Vendlinski, T., & Kang, T. (2008). *Some aspects of the technical quality of formative assessments in middle school mathematics* (CSE Tech. Report 750). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment (pp. 53–68).* San Diego, CA: Academic Press.

Pressley, M., & Brainerd, C. J. (Eds.). (1985). *Cognitive learning and memory in children; Progress in cognitive development research*, New York, NY: Springer-Verlag.

Ready, T., Edley, Jr., C., & Snow, C. E. (Eds.). (2002*). Achieving high educational standards for all: Conference summary.* Washington, DC: National Academy Press.

Richardson-Klavehn, A., & Bjork, R. A. (2002). *Memory: Long term* .In L. Nadel (Ed.), *Encyclopedia of cognitive science, 2 (*pp. 1096-1105). London, UK: Nature Publishing Group.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39(5*), 369-393.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Boston, MA: Kluwer Academic Publishers.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching, (4th ed.*, pp. 1066–1101). Washington, DC: American Educational Research Association.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324-328.

Stobart, G. (2006). Influencing classroom assessment. *Assessment in Education: Principals, Policy & Practice, 12*(3), 235-238.

VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly & D. J. Foss (Eds.), *Annual Review of Psychology, 42* (pp. 513-539). Palo Alto, CA: Annual Reviews.

Wolf, D., Bixby, J., Glenn III, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education, 17* (pp. 31-74). Washington, DC: American Educational Research Association.

Yu, C.Y. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles, CA. Retrieved from Mplus website http://www.statmodel.com/download/Yudissertation.pdf

Appendix A:

CFA Result of PS Grade 6 Pretest

Table A1

CFA Result of PS Grade 6 Pretest

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
pre01	=	0.62		PA	+	0.78		e1	pre01	=	1.00		PA	+	1.00		e1
pre02	=	0.63	*	PA	+	0.78		e2	pre02	=	1.01	*	PA	+	1.00		e2
pre13	=	0.48	*	PA	+	0.88		e3	pre13	=	0.77	*	PA	+	1.00		e3
pre14	=	0.62	*	PA	+	0.78		e4	pre14	=	1.00	*	PA	+	1.00		e4
pre15	=	0.49	*	PA	+	0.87		e5	pre15	=	0.79	*	PA	+	1.00		e5
pre16	=	0.35	*	PA	+	0.94		e6	pre16	=	0.57	*	PA	+	1.00		e6
pre22	=	0.33	*	PA	+	0.94		e7	pre22	=	0.53	*	PA	+	1.00		e7
pre23	=	-0.06	*	PA	+	1.00		e8	pre23	=	-0.09	*	PA	+	1.00		e8
pre05	=	0.39		RA	+	0.92		e9	pre05	=	1.00		RA	+	1.00		e9
pre06	=	0.24	*	RA	+	0.97		e10	pre06	=	0.61	*	RA	+	1.00		e10
pre07	=	0.56	*	RA	+	0.83		e11	pre07	=	1.44	*	RA	+	1.00		e11
pre08	=	0.38	*	RA	+	0.93		e12	pre08	=	0.98	*	RA	+	1.00		e12
pre19	=	0.30	*	RA	+	0.95		e13	pre19	=	0.79	*	RA	+	1.00		e13
pre20	=	0.40	*	RA	+	0.92		e14	pre20	=	1.03	*	RA	+	1.00		e14
pre26	=	0.29	*	RA	+	0.96		e15	pre26	=	0.75	*	RA	+	1.00		e15
pre03	=	0.71		RNE	+	0.70		e16	pre03	=	1.00		RNE	+	1.00		e16
pre10	=	0.47	*	RNE	+	0.88		e17	pre10	=	0.65	*	RNE	+	1.00		e17
pre11	=	0.27	*	RNE	+	0.96		e18	pre11	=	0.38	*	RNE	+	1.00		e18
pre12	=	0.43	*	RNE	+	0.90		e19	pre12	=	0.60	*	RNE	+	1.00		e19
pre17	=	0.30	*	RNE	+	0.95		e20	pre17	=	0.43	*	RNE	+	1.00		e20
pre24	=	0.09	*	RNE	+	1.00		e21	pre24	=	0.13	*	RNE	+	1.00		e21
pre04	=	0.36		SE	+	0.93		e22	pre04	=	1.00		SE	+	1.00		e22
pre09	=	0.57	*	SE	+	0.82		e23	pre09	=	1.58	*	SE	+	1.00		e23
pre18	=	0.39	*	SE	+	0.92		e24	pre18	=	1.09	*	SE	+	1.00		e24

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
pre21	=	0.26	*	SE	+	0.97		e25	pre21	=	0.73	*	SE	+	1.00		e25
pre25	=	0.46	*	SE	+	0.89		e26	pre25	=	1.30	*	SE	+	1.00		e26
pre27	=	0.15	*	SE	+	0.99		e27	pre27	=	0.41	*	SE	+	1.00		e27
pre28	=	0.49	*	SE	+	0.87		e28	pre28	=	1.38	*	SE	+	1.00		e28

Appendix B:

CFA Result of PS Grade 6 Interim Measure

Table B1

CFA Results of PS Grade 6 Interim Measure

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
inter05	=	0.47		PA	+	0.88		e1	inter05	=	1.00		PA	+	1.00		e1
inter12	=	0.50	*	PA	+	0.87		e2	inter12	=	1.05	*	PA	+	1.00		e2
inter13	=	0.38	*	PA	+	0.93		e3	inter13	=	0.81	*	PA	+	1.00		e3
inter14	=	0.33	*	PA	+	0.94		e4	inter14	=	0.70	*	PA	+	1.00		e4
inter15	=	0.26	*	PA	+	0.97		e5	inter15	=	0.55	*	PA	+	1.00		e5
inter16	=	0.12	*	PA	+	0.99		e6	inter16	=	0.26	*	PA	+	1.00		e6
inter17	=	0.32	*	PA	+	0.95		e7	inter17	=	0.69	*	PA	+	1.00		e7
inter19A	=	0.45	*	PA	+	0.89		e8	inter19A	=	0.96	*	PA	+	1.00		e8
inter19B	=	0.53	*	PA	+	0.85		e9	inter19B	=	1.13	*	PA	+	1.00		e9
inter20	=	0.44	*	PA	+	0.90		e10	inter20	=	0.94	*	PA	+	1.00		e10
inter01	=	0.30		RNE	+	0.95		e11	inter01	=	1.00		RNE	+	1.00		e11
inter02	=	0.50	*	RNE	+	0.87		e12	inter02	=	1.68	*	RNE	+	1.00		e12
inter03	=	0.65	*	RNE	+	0.76		e13	inter03	=	2.19	*	RNE	+	1.00		e13
inter04	=	0.30	*	RNE	+	0.95		e14	inter04	=	1.01	*	RNE	+	1.00		e14
inter06	=	0.55	*	RNE	+	0.84		e15	inter06	=	1.85	*	RNE	+	1.00		e15
inter07	=	0.69	*	RNE	+	0.73		e16	inter07	=	2.31	*	RNE	+	1.00		e16
inter08	=	0.51	*	RNE	+	0.86		e17	inter08	=	1.72	*	RNE	+	1.00		e17
inter09	=	0.31	*	RNE	+	0.95		e18	inter09	=	1.04	*	RNE	+	1.00		e18
inter10	=	0.39	*	RNE	+	0.92		e19	inter10	=	1.32	*	RNE	+	1.00		e19
inter11	=	0.52	*	RNE	+	0.85		e20	inter11	=	1.75	*	RNE	+	1.00		e20
inter18A	=	0.34	*	RNE	+	0.94		e21	inter18A	=	1.16	*	RNE	+	1.00		e21
inter18B	=	0.53	*	RNE	+	0.85		e22	inter18B	=	1.78	*	RNE	+	1.00		e22

Appendix C:

CFA Result of PS Grade 6 Transfer Measure

Table C1

CFA Result of PS Grade 6 Transfer Measure

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
post17	=	0.45		PA	+	0.89		e1	post17	=	1.00		PA	+	1.00		e1
post20	=	0.27	*	PA	+	0.96		e2	post20	=	0.60	*	PA	+	1.00		e2
post21	=	0.04	*	PA	+	1.00		e3	post21	=	0.09	*	PA	+	1.00		e3
post23	=	0.32	*	PA	+	0.95		e4	post23	=	0.70	*	PA	+	1.00		e4
post24	=	0.39	*	PA	+	0.92		e5	post24	=	0.86	*	PA	+	1.00		e5
post02	=	0.27		RNE	+	0.96		e6	post02	=	1.00		RNE	+	1.00		e6
post04	=	0.36	*	RNE	+	0.93		e7	post04	=	1.32	*	RNE	+	1.00		e7
post06	=	0.45	*	RNE	+	0.89		e8	post06	=	1.68	*	RNE	+	1.00		e8
post10A	=	0.65	*	RNE	+	0.76		e9	post10A	=	2.41	*	RNE	+	1.00		e9
post10B	=	0.44	*	RNE	+	0.90		e10	post10B	=	1.63	*	RNE	+	1.00		e10
post12	=	0.50	*	RNE	+	0.86		e11	post12	=	1.86	*	RNE	+	1.00		e11
post13	=	0.52	*	RNE	+	0.85		e12	post13	=	1.93	*	RNE	+	1.00		e12
post26	=	0.25	*	RNE	+	0.97		e13	post26	=	0.92	*	RNE	+	1.00		e13
post01	=	0.26		SE	+	0.96		e14	post01	=	1.00		SE	+	1.00		e14
post03	=	-0.05	*	SE	+	1.00		e15	post03	=	-0.21	*	SE	+	1.00		e15
post05	=	0.25	*	SE	+	0.97		e16	post05	=	0.96	*	SE	+	1.00		e16
post07	=	0.44	*	SE	+	0.90		e17	post07	=	1.67	*	SE	+	1.00		e17
post08	=	0.48	*	SE	+	0.88		e18	post08	=	1.82	*	SE	+	1.00		e18
post09	=	-0.19	*	SE	+	0.98		e19	post09	=	-0.71	*	SE	+	1.00		e19
post11	=	0.54	*	SE	+	0.84		e20	post11	=	2.03	*	SE	+	1.00		e20
post14	=	0.43	*	SE	+	0.90		e21	post14	=	1.63	*	SE	+	1.00		e21
post15	=	0.39	*	SE	+	0.92		e22	post15	=	1.48	*	SE	+	1.00		e22
post16	=	0.32	*	SE	+	0.95		e23	post16	=	1.22	*	SE	+	1.00		e23
post18	=	0.26	*	SE	+	0.97		e24	post18	=	0.99	*	SE	+	1.00		e24

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
post19	=	0.26	*	SE	+	0.97		e25	post19	=	0.98	*	SE	+	1.00		e25
post22	=	0.35	*	SE	+	0.94		e26	post22	=	1.34	*	SE	+	1.00		e26
post25	=	0.32	*	SE	+	0.95		e27	post25	=	1.22	*	SE	+	1.00		e27
post27A	=	0.55	*	SE	+	0.83		e28	post27A	=	2.10	*	SE	+	1.00		e28
post27B	=	0.33	*	SE	+	0.94		e29	post27B	=	1.25	*	SE	+	1.00		e29
post27C	=	0.33	*	SE	+	0.94		e30	post27C	=	1.27	*	SE	+	1.00		e30

Appendix D:

CFA Results of PS Grade 7 Pretest

Table D1

CFA Result of PS Grade 7 Pretest

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
pre01	=	0.34		PA	+	0.94		e1	pre01	=	1.00		PA	+	1.00		e1
pre02	=	-0.17	*	PA	+	0.99		e2	pre02	=	-0.50	*	PA	+	1.00		e2
pre03	=	0.39	*	PA	+	0.92		e3	pre03	=	1.17	*	PA	+	1.00		e3
pre18	=	0.08	*	PA	+	1.00		e4	pre18	=	0.23	*	PA	+	1.00		e4
pre19	=	0.32	*	PA	+	0.95		e5	pre19	=	0.96	*	PA	+	1.00		e5
pre20	=	0.31	*	PA	+	0.95		e6	pre20	=	0.91	*	PA	+	1.00		e6
pre21	=	0.40	*	PA	+	0.92		e7	pre21	=	1.19	*	PA	+	1.00		e7
pre22	=	0.37	*	PA	+	0.93		e8	pre22	=	1.10	*	PA	+	1.00		e8
pre26	=	-0.05	*	PA	+	1.00		e9	pre26	=	-0.15	*	PA	+	1.00		e9
pre04	=	0.21		RNE	+	0.98		e10	pre04	=	1.00		RNE	+	1.00		e10
pre05	=	0.06	*	RNE	+	1.00		e11	pre05	=	0.28	*	RNE	+	1.00		e11
pre06	=	0.54	*	RNE	+	0.84		e12	pre06	=	2.64	*	RNE	+	1.00		e12
pre07	=	0.44	*	RNE	+	0.90		e13	pre07	=	2.16	*	RNE	+	1.00		e13
pre08	=	0.30	*	RNE	+	0.95		e14	pre08	=	1.46	*	RNE	+	1.00		e14
pre09	=	0.55	*	RNE	+	0.83		e15	pre09	=	2.70	*	RNE	+	1.00		e15
pre10	=	0.23	*	RNE	+	0.97		e16	pre10	=	1.12	*	RNE	+	1.00		e16
pre11	=	0.08	*	RNE	+	1.00		e17	pre11	=	0.37	*	RNE	+	1.00		e17
pre12	=	0.35	*	RNE	+	0.94		e18	pre12	=	1.71	*	RNE	+	1.00		e18
pre13	=	0.20	*	RNE	+	0.98		e19	pre13	=	0.97	*	RNE	+	1.00		e19
pre14	=	0.27	*	RNE	+	0.96		e20	pre14	=	1.29	*	RNE	+	1.00		e20
pre15	=	0.55	*	RNE	+	0.84		e21	pre15	=	2.66	*	RNE	+	1.00		e21
pre16	=	0.35	*	RNE	+	0.94		e22	pre16	=	1.69	*	RNE	+	1.00		e22
pre17	=	0.30	*	RNE	+	0.96		e23	pre17	=	1.44	*	RNE	+	1.00		e23
pre27	=	0.22	*	RNE	+	0.97		e24	pre27	=	1.08	*	RNE	+	1.00		e24

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
pre23	=	0.14		SE	+	0.99		e25	pre23	=	1.00		SE	+	1.00		e25
pre24	=	-0.02	*	SE	+	1.00		e26	pre24	=	-0.13	*	SE	+	1.00		e26
pre25	=	0.38	*	SE	+	0.92		e27	pre25	=	2.75	*	SE	+	1.00		e27

Appendix E:

CFA Result of PS Grade 7 Transfer Measure

Table E1

CFA Result of PS Grade 7 Transfer Measure

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
post01	=	0.08		PA	+	1.00		e1	post01	=	1.00		PA	+	1.00		e1
post03	=	0.25	*	PA	+	0.97		e2	post03	=	2.94	*	PA	+	1.00		e2
post11	=	0.58	*	PA	+	0.81		e3	post11	=	6.96	*	PA	+	1.00		e3
post17	=	0.42	*	PA	+	0.91		e4	post17	=	5.07	*	PA	+	1.00		e4
post06	=	0.52		RA	+	0.85		e5	post06	=	1.00		RA	+	1.00		e5
post07	=	0.32	*	RA	+	0.95		e6	post07	=	0.61	*	RA	+	1.00		e6
post08	=	0.35	*	RA	+	0.94		e7	post08	=	0.67	*	RA	+	1.00		e7
post21	=	0.41	*	RA	+	0.91		e8	post21	=	0.78	*	RA	+	1.00		e8
post24B	=	0.37	*	RA	+	0.93		e9	post24B	=	0.71	*	RA	+	1.00		e9
post26A	=	0.13	*	RA	+	0.99		e10	post26A	=	0.24	*	RA	+	1.00		e10
post26B	=	0.24	*	RA	+	0.97		e11	post26B	=	0.47	*	RA	+	1.00		e11
post02	=	0.30		RNE	+	0.95		e12	post02	=	1.00		RNE	+	1.00		e12
post05	=	0.61	*	RNE	+	0.79		e13	post05	=	2.03	*	RNE	+	1.00		e13
post12	=	0.47	*	RNE	+	0.88		e14	post12	=	1.55	*	RNE	+	1.00		e14
post14	=	0.49	*	RNE	+	0.87		e15	post14	=	1.60	*	RNE	+	1.00		e15
post15	=	0.53	*	RNE	+	0.85		e16	post15	=	1.76	*	RNE	+	1.00		e16
post23A	=	0.63		SE	+	0.78		e17	post23A	=	1.00		SE	+	1.00		e17
post23B	=	0.40	*	SE	+	0.92		e18	post23B	=	0.63	*	SE	+	1.00		e18
post23C	=	0.67	*	SE	+	0.75		e19	post23C	=	1.05	*	SE	+	1.00		e19
post24A	=	0.60	*	SE	+	0.80		e20	post24A	=	0.95	*	SE	+	1.00		e20
post04	=	0.32	*	SE	+	0.95		e21	post04	=	0.51	*	SE	+	1.00		e21
post09	=	0.36	*	SE	+	0.93		e22	post09	=	0.57	*	SE	+	1.00		e22
post10	=	0.53	*	SE	+	0.85		e23	post10	=	0.84	*	SE	+	1.00		e23
post13	=	0.37	*	SE	+	0.93		e24	post13	=	0.58	*	SE	+	1.00		e24

Manifest Variable Equations with Standardized Estimates									Manifest Variable Equations with Estimates								
post16	=	0.35	*	SE	+	0.94		e25	post16	=	0.55	*	SE	+	1.00		e25
post18	=	0.43	*	SE	+	0.90		e26	post18	=	0.68	*	SE	+	1.00		e26
post19	=	0.12	*	SE	+	0.99		e27	post19	=	0.19	*	SE	+	1.00		e27
post20	=	0.52	*	SE	+	0.86		e28	post20	=	0.82	*	SE	+	1.00		e28
post22	=	0.40	*	SE	+	0.91		e29	post22	=	0.64	*	SE	+	1.00		e29
post25	=	0.25	*	SE	+	0.97		e30	post25	=	0.39	*	SE	+	1.00		e30

Appendix F:
Item Analysis Results of PS Grade 6 Pretest

Table F1

Item Analysis Results of PS Grade 6 Pretest

		p-value		Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.922)	Alpha=.81
Item	Domain	0	1	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
PRE1	PA	0.24	0.77	0.75	-0.55	0.024	0.32	0.80
PRE2	PA	0.33	0.67	0.73	-0.24	0.021	0.33	0.79
PRE3	RNE	0.07	0.93	0.91	-0.04	0.021	0.34	0.80
PRE4	SE	0.02	0.98	0.77	-0.16	0.021	0.33	0.81
PRE5	RA	0.62	0.38	0.50	0.56	0.021	0.32	0.80
PRE6	RA	0.37	0.63	0.38	-0.12	0.021	0.33	0.81
PRE7	RA	0.10	0.90	0.78	-1.17	0.032	0.29	0.80
PRE8	RA	0.06	0.94	0.64	-1.54	0.041	0.26	0.80
PRE9	SE	0.08	0.92	0.84	-1.34	0.036	0.27	0.80
PRE10	RNE	0.20	0.80	0.59	-0.67	0.025	0.32	0.80
PRE11	RNE	0.40	0.60	0.51	-0.03	0.021	0.34	0.80
PRE12	RNE	0.18	0.82	0.55	-0.75	0.026	0.32	0.80
PRE13	PA	0.13	0.87	0.79	-1.00	0.029	0.30	0.80
PRE14	PA	0.22	0.78	0.78	-0.61	0.024	0.32	0.80
PRE15	PA	0.40	0.60	0.69	-0.03	0.021	0.34	0.80
PRE16	PA	0.36	0.65	0.57	-0.16	0.021	0.33	0.80
PRE17	RNE	0.37	0.64	0.59	-0.13	0.021	0.33	0.80
PRE18	SE	0.15	0.85	0.69	-0.89	0.028	0.31	0.80
PRE19	RA	0.26	0.74	0.47	-0.45	0.023	0.33	0.80
PRE20	RA	0.48	0.52	0.60	0.17	0.020	0.33	0.80
PRE21	SE	0.64	0.36	0.53	0.61	0.021	0.32	0.80
PRE22	PA	0.30	0.71	0.46	-0.34	0.022	0.33	0.81

		p-value		Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.922)	Alpha=.81
Item	Domain	0	1	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
PRE23	PA	0.87	0.13	0.06	3.60	0.152	0.05	0.81
PRE24	RNE	0.76	0.24	0.25	0.97	0.023	0.30	0.81
PRE25	SE	0.20	0.80	0.68	-0.70	0.025	0.32	0.80
PRE26	RA	0.59	0.41	0.44	0.46	0.020	0.33	0.81
PRE27	SE	0.87	0.13	0.44	1.45	0.029	0.26	0.81
PRE28	SE	0.27	0.73	0.74	-0.43	0.023	0.33	0.79

Appendix G:

Item Analysis Results of PS Grade 7 Pretest

Table G1

Item Analysis Results of PS Grade 7 Pretest

		p-value		Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.922)	Alpha=.73
Item	Domain	0	1	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
pre01	PA	0.74	0.26	0.558	0.70	0.024	0.32	0.72
pre02	PA	0.34	0.66	-0.022	-0.45	0.022	0.33	0.75
pre03	PA	0.27	0.73	0.61	-0.66	0.023	0.32	0.72
pre04	RNE	0.75	0.25	0.417	0.72	0.024	0.32	0.73
pre05	RNE	0.03	0.97	0.65	-2.26	0.061	0.18	0.73
pre06	RNE	0.47	0.53	0.69	-0.09	0.021	0.34	0.71
pre07	RNE	0.41	0.59	0.608	-0.24	0.021	0.33	0.72
pre08	RNE	0.17	0.83	0.584	-1.02	0.027	0.3	0.72
pre09	RNE	0.50	0.50	0.676	0.00	0.021	0.34	0.71
pre10	RNE	0.41	0.59	0.442	-0.23	0.021	0.33	0.73
pre11	RNE	0.77	0.23	0.281	0.78	0.025	0.31	0.74
pre12	RNE	0.43	0.57	0.517	-0.19	0.021	0.33	0.72
pre13	RNE	0.24	0.76	0.496	-0.77	0.024	0.31	0.73
pre14	RNE	0.69	0.31	0.453	0.53	0.022	0.33	0.73
pre15	RNE	0.59	0.41	0.661	0.25	0.021	0.33	0.71
pre16	RNE	0.78	0.22	0.544	0.83	0.025	0.31	0.72
pre17	RNE	0.75	0.25	0.494	0.72	0.024	0.32	0.73
pre18	PA	0.87	0.13	0.362	1.23	0.030	0.28	0.73
pre19	PA	0.63	0.38	0.504	0.34	0.022	0.33	0.72
pre20	PA	0.57	0.43	0.475	0.19	0.021	0.33	0.73
pre21	PA	0.29	0.71	0.599	-0.61	0.023	0.32	0.72
pre22	PA	0.80	0.20	0.639	0.92	0.026	0.31	0.72

		p-value		Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.922)	Alpha=.73
Item	Domain	0	1	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
pre23	SE	0.70	0.30	0.392	0.56	0.023	0.32	0.73
pre24	SE	0.76	0.25	0.183	0.74	0.024	0.32	0.74
pre25	SE	0.48	0.52	0.587	-0.05	0.021	0.34	0.72
pre26	PA	0.77	0.23	0.14	0.79	0.025	0.31	0.74
pre27	RNE	0.75	0.25	0.432	0.73	0.024	0.32	0.73

Appendix H:

Sources of Transfer Measure Items

Grade 6 Transfer Measure Items-2009

Item #	Item Text (2008)	Format	Item Source	CA Standard	Powersource Domain	Item Data 2008	Revised item Text	Format
1	Solve: 6n = 36 a) 12 b) 2 c) 30 d) 6	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 11	AF 1.1	SE	91.1		
2	Charlie can type 32 words per minute. At this rate, how long would it take him in minutes to type 128 words? a) 1 b) 3 c) 4 d) 2	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 28	AF 2.3	RNE	88.5		
3	b = 14 + a. When a equals 7, what is the value of b?	SA	QCA, key stage 3, p. 4, #3	AF 1.2	SE	86.2		
4	Write the fraction 3/9 in its simplest form.	SA	QCA, key stage 3, p. 4, #7	NS 2.4	RNE	84.3	Explain how you would use the multiplicative identity to write 3/9 iijn its simplest form.	EX
5	What value of x makes the equation true? x - 9 = 32 a) 23 b) 41 c) 32 d) 9	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 10	AF 1.1	SE	82.6		
6	If 12/n = 36/21, then n equals: a) 3 b) 7 c) 36 d) 63	MC		NS 1.3	RNE	76.5		
7	There were two thousand people at a concert. Nine hundred and ninety-two of them were women. How many of the people were not women?	SA	QCA, key stage 3, p. 5, #19	NS 2.0 NS 2.3		68.9	There were 2,000 people at a concert. 992 of them were women. How many of the people were not women?	
8	How much change will John get back from $5.00 if he buys 2 notebooks that cost $1.80 each? a) $1.40 b) $2.40 c) $3.20 d) $3.60	MC	NAEP, grade 8, 2003, #7	NS 2.3	SE	67.7		
9	n is a number. When n is multiplied by 7, and 6 is then added, the result is 41. Which of these equations represents this relation? a) 7n + 6 = 41 b) 7n + – 6 = 41 c) 7n • 6 = 41 d) 7(n + 6) = 41	MC	TIMSS, grade 8, 1999, item number: B12	AF 1.0	SE	66.4		
10	Write a different fraction that is equivalent to three-fifths.	SA	QCA, key stage 3, p. 8, #5	NS 2.1	RNE	64.1	a) Write a different fraction that is equivalent to 3/5. b) Explain how you know the two fractions are equal.	EX
11	What is the value of x in the triangle? a) 65º b) 82º c) 90º d) 92º e) 98º	MC	NAEP, grade 8, 2003, #32	MG 2.2	SE	63.2		
12	In which list of fractions are all of the fractions equivalent? a) 12 , 24 , 46 b) 23 , 46 , 812 c) 25 , 410 , 850 d) 34 , 46 , 68	MC	TIMSS, grade 8, 1999, item number: N14	NS 1.1	RNE	59.9		

13	Which of the following ratios is equivalent to the ratio of 6 to 4? a) 12 to 18 b) 12 to 8 c) 8 to 6 d) 4 to 6 e) 2 to 3	MC	NAEP, grade 8, 2003, #58	NS 1.2	RNE	58.7		
14	The perimeter of a square is 36 inches. What is the length of one side of the square? a) 4 inches b) 6 inches c) 9 inches d) 18 inches	MC	NAEP, grade 8, 2003, #10	AF 3.1		57.4		
15	Sam's uncle is 21 years older than Sam. His uncle is 42. What equation could you use to solve for Sam's age, s ? a) s + 21 = 42 b) 4221 = s c) s - 21 = 42 d) s - 42 = 21	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 16	AF 1.1	SE	56.6		
16	What is the next step to solve this equation? x - 7 = 13 a) Subtract 7 from both sides b) Add x to both sides c) Add 7 to both sides d) Subtract 13 from both sides	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 14	AF 1.1	SE	55.7	In order to isolate x, what would be the next stepin the following equation? x - 7 = 13 a) Subtract 7 from both sides b) Add x to both sides c) Add 7 to both sides d) Subtract 13 from both sides	
17	Which of the following numerical expressions gives the area of the rectangle below? a) 4 • 6 b) 4 + 6 c) 2 (4 • 6) d) 2 (4 + 6) e) 4 + 6 + 4 + 6	MC	NAEP, grade 8, 2003, #34	AF 3.1	PA	43.4		
18	What is the value of p in the equation below ? 14p = 4 a) p = 4 b) p = 16 c) p = 4 14 c) p = 3 34	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 12	AF 1.1	SE	40.8		
19	If 3 + w = b, then w = a) 39 b) b • 3 c) b + 3 d) 3 – b e) b – 3	MC	NAEP, grade 8, 2003, #47	MR 3.3	SE	35.9		
20	16. Which of the following is equal to 6 (x + 6) ? a) x + 12 b) 6x + 6 c) 6x + 12 d) 6x + 36 e) 6x + 66	MC	NAEP, grade 8, 2005, #41	AF 1.3	PA	22.5		
21	What would be your answer if you were asked to multiply 8 • (x + 34) ? a) 8x + 34 b) 8 34x c) 8x + 6 d) x + 6	MC	Adapted from PISA item	AF 1.3	PA	19.6		

22			NAEP NQT v3.0, grade 8, 2005, p.8, #8	AF 1.2	SE	N/A	Which piece of information is NOT needed to solve the problem below. You do not have to solve the problem. "Carlos is planning to buy food for his 2 dogs. The food he buys must last for 4 weeks. Each dog eats 1 can of dog food and 3 dog biscuits every day. How many cans of dog food does Carlos need to buy?" a) Carlos has 2 dogs. b) The food must last 4 weeks. c) Each dog eats 1 can of dog food every day. d) Each dog eats 3 biscuits every day.	MC
23	Which of the following shows the distributive property being used correctly to simplify the expression: 3(4) + 3(2) a) 3(4)(2) b) 3(4 + 2) c) 4(3 + 2) d) 4(3) + 2(3)	MC	Adapted from 6th Grade Benchmark Test--3--Norwalk La Mirada--item 24	AF 1.3	PA	45.5		
24	Simplify using the distributive property. y (y – 6) =	SA	QCA, key stage 3, tier 6-8, paper 2, p. 15, #14	AF 1.3	PA	17.9	Explain how you would use the distributive property to rewrite this expression 3 (y - 6).	EX
25	For all numbers k, k + k + k + k + k can be written as a) k + 5 b) 5k c) k5 d) 5 (k + 1)	MC	TIMSS, grade 8, 1999, item number: p11	AF 1.0	SE	31.9	Explain why k + k + k + k + k is the same as 5k?	EX
26	Explain why the fraction 1/2/3/4 is equivalent to the fraction 2/3?	EX	Adapted from PISA item	NS 2.1	RNE	32.1 (0) 59.7 (1) 7.5 (2) 0.6 (3)		
27	The diagram shows triangle PQR. Work out the sizes of angles a, b, and c.	SA	QCA, key stage 3, tier 3-5, paper 1, p. 20, #20	MG 2.2	SE	a) 30.0 b) 14.1 c) 19.0		

Grade 7 Transfer Measure Items-2009						
Item #	**Item Text**	**Format**	**Item Source**	**CA Standard**	**Powersource Domain**	**Item Answer**
1	Which of these expressions is equivalent to n x n x n for all values of n? a) n/3 b) n+3 c) 3n d) n3 (change this to cubed)	MC	TIMSS 1999 8th grade, p.49, Item number P09	NS 1.2/NS 2.1	PA	d) n^3
2	Subtract: (3x/7)-(x/7)= (Note: the actual problem shows these as fractions without the parentheses) a) 2/7 b) 3 c) 2x d) x/7 e) 2x/7	MC	TIMSS 1999 8th grade, p.85, Item number M022185	NS 1.2	RNE & SE	3) 2x/7
3	Which of these is equal to (370x998)+(370x2)? a) 370x1,000 b) 372x998 c)740x998 d) 370x998x2	MC	TIMSS 1999 8th grade, p.178, Item number M032690	AF 1.3	PA	a) 370 x 1,000
4	Which written expression could be represented by 37-3n=5? a) The sum of 37 and 3 times a number is 5. b) The product of n and 37 decreased by 3 is 5. c) Three times a number decreased by 37 is 5. d) Thirty-seven decreased by 3 times a number is 5.	MC	Pre-Algebra Benchmark One (2007-2008); p. 5, #17	AF 1.1	SE	d) Thirty-seven decreased by 3 times a number is 5.
5	Here are four fractions. 3/4 1/8 1/3 3/5 (Note: written in fraction form) Look at the number line below. Write each fraction in the correct box. (Note: There are 4 empty boxes with arrows pointing to different areas along a number line. 2 of the boxes are between 0 and 0.5, the other 2 are between 0.5 and 1...take a look at the format of this question.) Sidenote: If our students don't know about decimals yet, the 0.5 on the number line could be written as 1/2.	SA	QCA 2005, key stage 3 (tier 4-6), p.8, #6	NS 1.1	RNE	The numbers should be written in boxes (from left to right): 1/8 1/3 3/5 3/4
6	In the figure, how many MORE small squares need to be shaded so that 4/5 of the small squares are shaded? a) 5 b) 4 c) 3 d) 2 e) 1 (Note: The figure shows a rectangle composed of 10 small squares in two rows of five squares. Only 3 small squares are currently shaded.)	MC	TIMSS 1999 8th grade, p.162, Item number M012001	NS 1.1 and NS 1.2? Since they have to know that 4/5=8/10 AND that 8-3=5	RA	a) 5
7	On the road shown above, the distance from Bay City to Exton is 60 miles. What is the distance from Bay City to Yardville? a) 45 miles b) 75 miles c) 90 miles d) 105 miles (See Figure. There's a line segment with equidistant notches. Exton is 4 notches from Bay City. Yardville is an additional 3 notches away.)	MC	NAEP NQT v3.0, grade 8, 2003, p.6, #19	NS 4.0 (Do 4.0 and 4.1 differ in terms of the number of steps needed? This requires only 1 step.)	RA	d) 105 miles
8	If there are 300 calories in 100 g of a certain food, how many calories are there in a 30 g portion of this food? a) 90 b) 100 c) 900 d) 1000 e) 9000	MC	TIMSS 1999 8th grade, p.2, Item number B08	NS 1.2	RA	a) 90
9	I think of a number. I multiply this number by 8, then subtract 66. The result is twice the number that I was thinking of. Which equation represents this situation? a) 8n-66=2n b) n+8-66=2+n c) 8n*66=2n d) 8+n*66=2+n	MC	Adapted from QCA 2005, key stage 3 (tier 4-6), p.25, #23	AF 1.1	SE	a) 8n-66=2n

10	If 4 times a number is 48, what is 1/3 of the number? a) 4 b) 8 c) 12 d) 16	MC	TIMSS 1999 8th grade, p.39, Item number D11	NS 1.2 if treated as an RNE problem (like #7 released question)	SE (or could be set up as a RNE problem using proportions)	a) 4
11	A garden has 14 rows. Each row has 20 plants. The gardener then plants 6 more rows with 20 plants in each row. Write an expression to show how many plants there are altogether.	SA	Adapted from TIMSS 1999 8th grade, p.182, Item number M032671	AF 1.1	PA	20(14+6) or 20(14) + 20(6)
12	Jim has 3/4 of a yard of string which he wishes to divide into pieces, each 1/8 of a yard long. How many pieces will he have? a) 3 b) 4 c) 6 d) 8	MC	NAEP NQT v3.0, grade 8, 2003, p.6, #17	NS 4.0 (Do 4.0 and 4.1 differ in terms of the number of steps needed? This requires only 1 step.)	RNE	c) 6
13	Fifteen boxes each containing 8 radios can be repacked in 10 larger boxes each containing how many radios? a) 8 b) 10 c) 12 d) 80 e) 120	MC	NAEP NQT v3.0, grade 8, 2003, p.13, #38	NS 4.1	SE	c) 12
14	(3/5)+(3/10 x 4/15)= Note: The fractions are typed out in fraction form & the first fraction doesn't have parentheses. a) 3/51 b) 1/6 c) 6/25 d) 11/25 e) 17/25	MC	TIMSS 1999 8th grade, p.154, Item number M022199	NS 1.2	RNE	e) 17/25
15	Robin and Jim took cherries from a basket. Robin took 1/3 of the cherries and Jim took 1/6 of the cherries. What fraction of the cherries remained in the basket? a) 1/2 b) 1/3 c) 1/6 d) 1/18	MC	TIMSS 1999 8th grade, p.25, Item number P15	AF 4.1	RNE	a) 1/2
16	The cost, *C*, of printing greeting cards consists of a fixed charge of 100 cents and a charge of 6 cents for each card printed. Which of these equations can be used to determine the cost of printing *n* cards? a) C=(100+6n) cents b) C=(106+n)cents c) C=(6+100n)cents d) C=(106n) cents e) C=(600n)cents	MC	TIMSS 1999 8th grade, p.39, Item number D10	AF 1.1	SE	a) C=(100+6n) cents
17	The fraction 2 1/4 means 2 + 1/4, which can also be written as (2 + 1/4). Show how you would use the distributive property to multiply 2 1/4 by 10.	SA	PISA, p. 1, #1 under "Distributed thoughts"	AF 1.3	PA	10(2 + 1/4) = 10(2) + 10(1/4)
18	A rectangular playground has a perimeter of 390 feet. The width of the playground is 75 feet. What is its length? a) 5.2 feet b) 97.5 feet c) 120 feet d) 130 feet 3) 240 feet	MC	NAEP NQT v3.0, grade 8, 2005, p.25, #40	MG 2.1 (but finding length rather than perimeter)	SE; if the student writes an equation using the Distributive Property, this would involve PA	c) 120 feet
19	Graham has twice as many books as Bob. Chan has six more books than Bob. If Bob has x books, which of the following represents the total number of books the three boys have? a) 3x+6 b) 3x+8 c) 4x+6 d) 5x+6 e) 8x+2	MC	TIMSS 1999 8th grade, p.89, Item number M022251	AF 1.1	SE	c) 4x+6

20	Daniel had 31 baseball cards. He gave the cards to his friends. Six of his friends received 3 cards each. Seven of his friends received 1 card each. The rest received 2 cards each. How many of his friends received exactly 2 cards from Daniel? Explain how you found your answer.	SA & EX	NAEP NQT v3.0, grade 8, 2005, p.31, #56	AF 4.2	SE	3

21	The screens of widescreen and standard televisions look different. They have different proportions. (See how this question is formatted.) Widescreen television- Ratio of height to width is **9:16**. Standard television- Ratio of height to width is **3:4**. Keri starts to draw scale drawings of the televisions. For each, the height is 4.5 cm. What should the **width** of each scale drawing be? Next to a drawing for the widescreen tv is the text, "The width of this scale drawing should be......cm" Next to a drawing for the standard television is the text, "The width of this scale drawing should be....cm"	SA	QCA 2005, key stage 3 (tier 4-6), p.24, #21	NS 4.0	RA	Widescreen = 8 cm; Standard = 6 cm
22	A painter had 25 L of paint. He used 2.5 L of paint every hour. He finished the job in 5.5 hours. How much paint did he have left? a) 10.25 L b) 11.25 L c) 12.75 L d) 13.75 L	MC	TIMSS 1999 8th grade, p.21, Item number N17	AF 4.2	SE	b) 11.25 L
23	John sold 60 magazines and Mark sold 80 magazines. The magazines were all sold for the same price. The total amount of money received for the magazines was $700. a) Write an equation to model this situation. b) Solve the equation to figure out how much money each boy made.	SA AND have to show equation and calculations	Adapted from TIMSS 1999 8th grade, p.31, Item number R15	a) AF 1.1; b) AF 4.1	SE (or could be set up as a RNE problem using proportions)	a) 60x+80x=700 with x being the cost of 1 magazine; OR 60/80=x/(700-x) if setting it up as an RNE and x is the amount John sold; b) John = $300, Mark = $400
24	A book publisher sent 140 copies of a certain book to a bookstore. The publisher packed the books in two types of boxes. One type of box held 8 copies of the book, and the other type of box held 12 copies of the book. The boxes were all full, and there were equal numbers of both types of boxes. a) How many boxes holding 12 books were sent to the bookstore? b) What fraction of the books sent to the bookstore were packed in the smaller boxes?	SA	TIMSS 1999 8th grade, p.32, Item number T02A	a) AF 4.1; b)NS 1.2	a) SE; b) RA	a) 7 b)56/140 or 14/35
25	In one week Jamal watched television for 26 hours. In that week: He watched television for the same length of time on Monday, Tuesday, Wednesday, and Thursday. On each Friday, Saturday and Sunday, he watched television for twice as long as on Monday. How long did he spend watching television on Saturday? Write your answers in hours and minutes.	SA	QCA 2005, key stage 3 (tier 6-8), p.14, #12 (Note: take a look at their question since the	AF 4.2	SE	5.2 hrs, 312 minutes

26	A biologist needs to estimate the size of the deer herd on a wildlife reserve. The biologist captures 150 deer, then tags and releases them. A week later, the biologist captures 50 deer and counts the number of tagged and the number of untagged deer. There are 15 tagged deer and 35 untagged deer in this group. The **ratio** of tagged to untagged deer in this group is the same as the **ratio** of tagged to untagged deer in the entire herd. a) If the number of deer in the herd is represented by the unknown "d", write an equation that shows the ratio of tagged deer to total deer in the captured group is equal to the ratio of tagged deer to total deer in the entire herd. b) How many untagged deer are in the total herd? Show your calculations.	SA AND have to give an equation and show calculations	PISA, p. 1, #1 under "Game reserve"	a) AF 1.1; b) NS 1.2	RA	1. (15/50)=(150/d); 2. student would need to show that (15/50)*(10/10)=150/500; Therefore, d=500. Since that is the estimated total number of deer, they'd have to show that untagged deer in the total heard = 500-150=350

Grade 8 Transfer Measure Items-2010 (based on v11 from Tamara)

Item #	Item Text	Format	Item Source	CA Standard	Powersource Domain	Item Answer	Rasch Difficulty	p-value	Notes
1	If the ratio 7 to 13 is the same as the ratio x to 52, what is the value of x? A) 7 B) 13 C) 28 D) 364	MC	1999 TIMSS 8th grade, p.38, Item number D08		RNE	C) 28			
2	Sam wanted to find three consecutive even numbers that add up to 84. He wrote the equation k + (k + 2) + (k + 4) = 84. What does the letter k represent? A) The least of the three even numbers B) The middle even number C) The greatest of the three even numbers D) The average of the three even numbers	MC	1999 TIMSS 8th grade, p.88, Item number M022002		SE	A) The least of the three even numbers			
3	Carla paid x zeds for 3 cartons of juice. What is the price in zeds of 1 carton of juice? A) x/3 B) 3/x C) 3 + x D) 3x	MC	1999 TIMSS 8th grade, p.90, Item number M032044		RNE	A) x/3			
4	If x = -3, what is the value of -3x? A) -9 B) -6 C) -1 D) 1 E) 9	MC	1999 TIMSS 8th grade, p.84, Item number M012042		PA	E) 9			
5	Which of the following is true when a, b, and c are different real numbers? A) a - b = b - a B) a(b - c) = b(c - a) C) b - c = c - b D) ab = ba E) ab - c = ab - b	MC	1999 TIMSS 8th grade, p.51, Item number R10		SE/PA	D) ab = ba			
6	The table shows some values of x and y, where x is proportional to y: (There's a table with 2 rows and 4 columns. In the first row, there's a x, 4, 8, Q. In the 2nd row, there's a y, 9, P, 45.) What are the values of P and Q? A) P = 40 and Q = 13 B) P = 18 and Q = 17 C) P = 20 and Q = 18 D) P = 40 and Q = 18 E) P = 18 and Q = 20	MC	1999 TIMSS 8th grade, p.46, Item number L15		RNE	E) P = 18 and Q = 20			
7	What is the value of 1 - 5 • (-2)? A) 11 B) 8 C) -8 D) -9	MC	1999 TIMSS 8th grade, p.167, Item number M032612		PA	A) 11			
8	If n is a negative integer, which of these is the largest number? A) 3 + n B) 3 • n C) 3 - n D) 3 ÷ n	MC	1999 TIMSS 8th grade, p. 168, Item # M032643		PA	C) 3 - n			
9	Write this expression as simply as possible. $9k^2/3k$=	MC	2005 QCA Key Stage 3 Tier 6-8, p.11, Item number 10		RNE	3k			
10	The number 0.01 can be written in many ways. a) Write the number 0.01 using words. For example, 10 would be written as "ten" and 35 would be written as "thirty-five". b) Write the number 0.01 as a fraction. c) Write the number 0.01 as a percent	SA	PISA questions under the "Fractions, Decimals, and Percents" section		RNE	a) one hundredths b) 1/100 c) 1%			

11	A scoop holds 1/5 kg of flour. How many scoops of flour are needed to fill a bag with 6 kg of flour? A) Answer: ___________________ B) Explain how you figured out the answer to this question in part A.	EX	We modified 1999 TIMSS 8th grade, p.153, Item number M022156 because we wanted to change it into an EX problem.		RNE	A) 30 scoops B) I set up an equation of equivalent fractions… 1 scoop/(1/5 kg) = x scoops/6 kg. Then I solved for x, the number of scoops of flour needed to fill a bag with 6 kg of flour.			
12	At a market, 7 oranges and 4 lemons cost 43 zeds, and 11 oranges and 12 lemons cost 79 zeds. Using *x* to represent the cost of an orange and *y* to represent the cost of a lemon, write two equations that could be used to find the values of *x* and *y* . Equation 1: _______________ Equation 2: _______________	SA/WP	1999 TIMSS 8th grade, p.97, Item number M032545		SE	Equation 1: 7x+4y=43 Equation 2: 11x+12y=79			
13	If y = 3x + 2, explain all the steps you must take to rewrite this equation so that x is expressed in terms of y.	EX	We modified 1999 TIMSS 8th grade, p.99, Item number M032046 to make it into an explanation problem.		SE	First, add the additive inverse of 2 to both sides of the equation. (You add -2 to get the 3x term on its own. You do this to both sides of the equation to keep the equation balanced.) Then multiply both sides of the equation by the multiplicative inverse of 3 so you can get x by itself. (Multiply 3 by 1/3 so the coefficient of x equals 1. You multiply both sides of the equation by 1/3 to keep the equation balanced.) You are left with x=(y-2)/3,			
14	If x - y = 5 and x/2= 3, what is the value of y? A) 6 B) 1 C) -1 D) -7	MC	1999 TIMSS 8th grade, p.95, Item number M032208		SE	B) 1			
15	Explain why 2x-3y+7x+5y can be simplified to 9x+2y.	EX	We modified 1999 TIMSS 8th grade, p.86, Item number M032036 to make it into an explanation problem.		SE?	2x and 7x are like terms since they represent groups of x so 2x and 7x can be combined to make 9x. -3y and 5y are like terms and can be combined to make 2y. Therefore, when all like terms are combined, 9x+2y is the result.			
16	The objects on the scale make it balance exactly. On the left pan there is a 1 kg weight (mass) and half a brick. On the right pan there is one brick. (There's a figure showing this.) What is the weight (mass) of one brick? A) 0.5 kg B) 1 kg C) 2 kg D) 3 kg	MC	1999 TIMSS 8th grade, p.98, Item number M012002		SE	C) 2 kg			
17	If a/b= 70, then a/2b= A) 35 B) 68 C) 72 D) 140	MC	1999 TIMSS 8th grade, p.96, Item number M032210		RNE	A) 35			

18	For the expression 3 + 15 ÷ 3 - 4 x 2, explain why adding 3 and 15 is not your first step when you simplify the expression.	EX	We modified 2003 NAEP Grade 8, p. 20 Item # 60 to make it into an explanation problem.		PA	Because order of operations dictates that division and multiplication are performed before addition and subtraction, adding 3 and 15 is not your first step; dividing 15 by 3 and multiplying -4 and 2 both occur before any addition.			
19	We want 4/5 of the squares to be shaded in the figure below. (There's a figure (2 rows, 5 columns) with 10 boxes. 3 of the boxes are shaded.) A) First, explain how to use the multiplicative identity property to figure out how many total small squares (out of 10) we want shaded. B) Total number of small squares (out of 10) we want shaded: _____ C) Explain how to find out how many MORE squares need to be shaded so that 4/5 of the small squares are shaded.	EX	We modified 1999 TIMSS 8th grade, p.162, Item number M012001 to make it into an explanation problem.		RNE	A) Because we have 10 boxes, we want to find out how many boxes out of 10 we want shaded. I used the multiplicative identity property to figure out that 4/5 is equivalent to 8/10 because 4/5 multiplied by 2/2 is 8/10. Multiplying by 2/2 is the same as multiplying by 1 so we don't change the underlying value of 4/5. B) 8 C) Since we want 8/10 of the squares to be shaded and we only have 3/10 of the squares shaded, we need to shade 8-3 = 5 squares to have 8 out of 10 squares shaded.			
20	In one week Jamal watched television for 26 hours. In that week: He watched television for the same length of time on Monday, Tuesday, Wednesday, and Thursday. On each Friday, Saturday and Sunday, he watched television for twice as long as on Monday. How long did he spend watching television on **Saturday**? Write your answers in hours and minutes.	SA	QCA 2005, key stage 3 (tier 6-8), p.14, #12 (Note: take a look at their question since the	Perhaps AF 4.2 since it requires more than 2 steps and involves some time conversion	SE	5.2 hrs, 12 minutes			Note: This appeared in the 2008-09 version of the 7th grade TM, but based on students' performance, we took it out of the 2009-10 version and added it here.
21	A biologist needs to estimate the size of the deer herd on a wildlife reserve. The biologist captures 150 deer, then tags and releases them. A week later, the biologist captures 50 deer and counts the number of tagged and the number of untagged deer. There are 15 tagged deer and 35 untagged deer in this group. The ratio of tagged to untagged deer in this group is the same as the ratio of tagged to untagged deer in the entire herd. a) If the number of deer in the herd is represented by the unknown *d*, write an equation that shows the ratio of tagged deer to total deer in the captured group is equal to the ratio of tagged deer to total deer in the entire herd. b) How many untagged deer are in the total herd? Show your calculations.	SA AND have to give an equation and show calculations	PISA, p. 1, #1 under "Game reserve"	a) AF 1.1; b) NS 1.2	RA	1. (15/50)=(150/d); 2. student would need to show how they solved for d. Since d=500 is the estimated total number of deer, they'd have to show that untagged deer in the total heard = 500-150=350.			Note: This appeared in the 2008-09 version of the 7th grade TM, but based on students' performance, we took it out of the 2009-10 version and added it here.

Appendix I:

Alignment of CA Standards and NCTM

6th Grade Standards	Domain	Checks for Understanding Item Number				Transfer Measure Item Number	NCTM Focal Points	Mathematical Standard for the Algebra Readiness Program
		RNE	SE	RA	PA			
Number Sense								
1.0 Students compare and order positive and negative fractions, decimals, and mixed numbers. Students solve problems involving fractions, ratios, proportions, and percentages:	**RNE, SE, RA**	RN-EX-12 RN-EX-15					NO1	
1.1 Compare and order positive and negative fractions, decimals, and mixed numbers and place them on a number line.	**RNE, RA**					#4, #31		**X**
1.2 Interpret and use ratios in different contexts (e.g., batting averages, miles per hour) to show the relative sizes of two quantities, using appropriate notations (a/b, a to b, a:b).	**RA**					#8	NO2	
1.3 Use proportions to solve problems (e.g., determine the value of N if 4/7 = N/ 21, find the length of a side of a polygon similar to a known polygon). Use cross-multiplication as a method for solving such problems, understanding it as the multiplication of both sides of an equation by a multiplicative inverse.	**RNE, SE, RA**	RN-BT-6 RN-BT-2 RN-EX-6ab RN-BT-5 RN-BT-6 RN-EX-18ab RN-BT-4	SE-BT-12 SE-FS-1 SE-BT-20		,	#7	NO2	
1.4 Calculate given percentages of quantities and solve problems involving discounts at sales, interest earned, and tips.								**X**
2.0 Students calculate and solve problems involving addition, subtraction, multiplication, and division:	**RNE, SE, RA**	RN-BT-7 RN-BT-8 RN-BT-9 RN-BT-15 RN-BT-16 RN-BT-17	SE-BT-6 SE-BT-7 SE-BT-8 SE-BT-9 SE-BT-12 SE-BT-20 SE-EX-26 SE-EX-19ab SE-EX-27 SE-FS-1 SE-FS-3			#1, #3	NO1, NO2	**X**
2.1 Solve problems involving addition, subtraction, multiplication, and division of positive fractions and explain why a particular operation was used for a given situation.	**RNE, SE, RA**	RN-EX-6ab RN-EX-13 RN-BT-15 RN-EX-17 RN-EX-18ab RN-EX-28ab	SE-EX-19ab SE-EX-26 SE-EX-27 SE-EX-28 SE-EX-29			#5	NO1	**X**
2.2 Explain the meaning of multiplication and division of positive fractions and perform the calculations (e.g., 5/8 ÷ 15/16 = 5/8 x 16/15 = 2/3).	**RNE, SE, RA**						NO1	**X**
2.3 Solve addition, subtraction, multiplication, and division problems, including those arising in concrete situations, that use positive and negative integers and combinations of these operations.	**RNE, SE, RA**					#3, #18	NO1	
2.4 Determine the least common multiple and the greatest common divisor of whole numbers; use them to solve problems with fractions (e.g., to find a common denominator to add two fractions or to find the reduced form for a fraction).		RN-EX-28ab				#2		
Algebra and Functions								
1.0 Students write verbal expressions and sentences as algebraic expressions and equations; they evaluate algebraic expressions, solve simple linear equations, and graph and interpret their results:	**PA, SE, RA**		SE-WP-11abc SE-WP-12abc SE-WP-14		PA-BT-32 PA-BT-33 PA-WP-2 PA-WP-3 PA-EX-8	#9, #32	Alg1	**X**

6th Grade Standards	Domain	Checks for Understanding Item Number				Transfer Measure Item Number	NCTM Focal Points	Mathematical Standard for the Algebra Readiness Program
1.1 Write and solve one-step linear equations in one variable.	SE, RA		SE-WP-11abc, SE-WP-12abc				Alg1	X
1.2 Write and evaluate an algebraic expression for a given situation, using up to three variables.	SE, RA					#6		
1.3 Apply algebraic order of operations and the commutative, associative, and distributive properties to evaluate expressions; and justify each step in the process.	PA, SE, RA				PA-FS-1, PA-EX-11ab, PA-EX-12ab, PA-WP-2, PA-WP-3, PA-FS-2	#16, #17, #30	Alg2	
1.4 Solve problems manually by using the correct order of operations or by using a scientific calculator.								
2.0 Students analyze and use tables, graphs, and rules to solve problems involving rates and proportions:	RA						NO2	
2.1 Convert one unit of measurement to another (e.g., from feet to miles, from centimeters to inches).	RA							
2.2 Demonstrate an understanding that rate is a measure of one quantity per unit value of another quantity.	RA							
2.3 Solve problems involving rates, average speed, distance, and time.	RA	RN-WP-6, RN-WP-7					NO2	
3.0 Students investigate geometric patterns and describe them algebraically:	RA							
3.1 Use variables in expressions describing geometric quantities (e.g., P = 2w + 2l, A = 1/2bh, C = pd - the formulas for the perimeter of a rectangle, the area of a triangle, and the circumference of a circle, respectively).	RA					#19, #20		
3.2 Express in symbolic form simple relationships arising from geometry.								
Measurement and Geometry								
1.0 Students deepen their understanding of the measurement of plane and solid shapes and use this understanding to solve problems:							Geo	
1.1 Understand the concept of a constant such as p; know the formulas for the circumference and area of a circle.								
1.2 Know common estimates of p (3.14; 22/7) and use these values to estimate and calculate the circumference and the area of circles; compare with actual measurements.								
1.3 Know and use the formulas for the volume of triangular prisms and cylinders (area of base x height); compare these formulas and explain the similarity between them and the formula for the volume of a rectangular solid.							Geo	
2.0 Students identify and describe the properties of two-dimensional figures:	SE, RA							
2.1 Identify angles as vertical, adjacent, complementary, or supplementary and provide descriptions of these terms.								
2.2 Use the properties of complementary and supplementary angles and the sum of the angles of a triangle to solve problems involving an unknown angle.	SE, RA		SE-EX-30			#21, #22, #33		
2.3 Draw quadrilaterals and triangles from given information about them (e.g., a quadrilateral having equal sides but no right angles, a right isosceles triangle).								
Statistics, Data Analysis, and Probability								
1.0 Students compute and analyze statistical measurements for data sets:								

6th Grade Standards	Domain	Checks for Understanding Item Number				Transfer Measure Item Number	NCTM Focal Points	Mathematical Standard for the Algebra Readiness Program
1.1 Compute the range, mean, median, and mode of data sets.								
1.2 Understand how additional data added to data sets may affect these computations of measures of central tendency.								
1.3 Understand how the inclusion or exclusion of outliers affects measures of central tendency.								
1.4 Know why a specific measure of central tendency (mean, median) provides the most useful information in a given context.								
2.0 Students use data samples of a population and describe the characteristics and limitations of the samples:								
2.1 Compare different samples of a population with the data from the entire population and identify a situation in which it makes sense to use a sample.								
2.2 Identify different ways of selecting a sample (e.g., convenience sampling, responses to a survey, random sampling) and which method makes a sample more representative for a population.								
2.3 Analyze data displays and explain why the way in which the question was asked might have influenced the results obtained and why the way in which the results were displayed might have influenced the conclusions reached.								
2.4 Identify data that represent sampling errors and explain why the sample (and the display) might be biased.								
2.5 Identify claims based on statistical data and, in simple cases, evaluate the validity of the claims.								
3.0 Students determine theoretical and experimental probabilities and use these to make predictions about events:								
3.1 Represent all possible outcomes for compound events in an organized way (e.g., tables, grids, tree diagrams) and express the theoretical probability of each outcome.								
3.2 Use data to estimate the probability of future events (e.g., batting averages or number of accidents per mile driven).								
3.3 Represent probabilities as ratios, proportions, decimals between 0 and 1, and percentages between 0 and 100 and verify that the probabilities computed are reasonable; know that if P is the probability of an event, 1- P is the probability of an event not occurring.								
3.4 Understand that the probability of either of two disjoint events occurring is the sum of the two individual probabilities and that the probability of one event following another, in independent trials, is the product of the two probabilities.								
3.5 Understand the difference between independent and dependent events.								
Mathematical Reasoning								
1.0 Students make decisions about how to approach problems:	**RNE, PA, SE, RA**		SE-WP-11abc SE-WP-12abc SE-FS-1 SE-FS-3		PA-FS-1, PA-FS-2,			**X**

6th Grade Standards	Domain	Checks for Understanding Item Number				Transfer Measure Item Number	NCTM Focal Points	Mathematical Standard for the Algebra Readiness Program
1.1 Analyze problems by identifying relationships, distinguishing relevant from irrelevant information, identifying missing information, sequencing and prioritizing information, and observing patterns.	**RNE, PA, SE, RA**	RN-EX-12 RN-EX-15			PA-BT-1, PA-BT-13,PA-BT-20, PA-BT-28, PA-BT-31,			**X**
1.2 Formulate and justify mathematical conjectures based on a general description of the mathematical question or problem posed.	**RNE, PA, SE, RA**							**X**
1.3 Determine when and how to break a problem into simpler parts.	**RNE, PA, SE, RA**				RN-BT-8, RN-BT-9, RN-BT-16, RN-BT-17			**X**
2.0 Students use strategies, skills, and concepts in finding solutions:	**RNE, PA, SE, RA**					#10 (maybe), #12, #13(maybe)		**X**
2.1 Use estimation to verify the reasonableness of calculated results.								**X**
2.2 Apply strategies and results from simpler problems to more complex problems.	**RNE, PA, SE, RA**							**X**
2.3 Estimate unknown quantities graphically and solve for them by using logical reasoning and arithmetic and algebraic techniques.	**SE, RA**				RN-EX-1ab, RN-EX-16ab			**X**
2.4 Use a variety of methods, such as words, numbers, symbols, charts, graphs, tables, diagrams, and models, to explain mathematical reasoning.	**RNE, PA, SE, RA**	RN-EX-1ab RN-EX-6ab RN-EX-12 RN-EX-13 RN-EX-15 RN-EX-16ab RN-EX-17 RN-EX-18ab	SE-EX-19ab SE-EX-23ab SE-WP-11abc SE-WP-12abc SE-FS-1 SE-FS-3		PA-BT-24 PA-BT-27 PA-BT-32 PA-BT-33 PA-EX-8 PA-EX-11ab PA-EX-12ab PA-WP-2 PA-WP-3		Alg1	**X**
2.5 Express the solution clearly and logically by using the appropriate mathematical notation and terms and clear language; support solutions with evidence in both verbal and symbolic work.	**RNE, PA, SE, RA**	RN-EX-16ab	SE-EX-19ab SE-EX-23ab SE-EX-26 SE-EX-28 SE-WP-11abc SE-WP-12abc		PA-BT-24 PA-BT-27 PA-EX-8 RN-EX-1ab,			**X**
2.6 Indicate the relative advantages of exact and approximate solutions to problems and give answers to a specified degree of accuracy.								**X**
2.7 Make precise calculations and check the validity of the results from the context of the problem.	**RNE, PA, SE, RA**							**X**
3.0 Students move beyond a particular problem by generalizing to other situations:	**RNE, PA, SE, RA**				PA-WP-2 PA-WP-3 PA-EX-8			**X**
3.1 Evaluate the reasonableness of the solution in the context of the original situation.	**RNE, PA, SE, RA**	RN-EX-6ab RN-EX-13 RN-EX-17 RN-EX-18ab	SE-EX-26 SE-EX-27 SE-EX-28 SE-EX-29					**X**
3.2 Note the method of deriving the solution and demonstrate a conceptual understanding of the derivation by solving similar problems.	**RNE, PA, SE, RA**							**X**
3.3 Develop generalizations of the results obtained and the strategies used and apply them in new problem situations.	**RNE, PA, SE, RA**					#25		**X**

Appendix J:

Item Analysis Results of PS Grade 6 Posttest

Table J1

Item Analysis Results of POWERSOURCE© Grade 7 Posttest

		p-value				Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.931)	Alpha=.81
Item	Domain	0	1	2	3	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
post01	SE	0.06	0.94			0.69	-1.80	0.040	0.23	0.83
post02	RNE	0.10	0.91			0.59	-1.54	0.034	0.26	0.83
post03	SE	0.96	0.04			-0.02	2.12	0.050	0.2	0.83
post04	RNE	0.60	0.23	0.13	0.04	0.57	0.90	0.014	0.59	0.83
post05	SE	0.17	0.83			0.46	-1.09	0.027	0.29	0.83
post06	RNE	0.24	0.76			0.64	-0.80	0.024	0.31	0.82
post07	SE	0.23	0.77			0.63	-0.86	0.025	0.31	0.82
post08	SE	0.30	0.70			0.62	-0.61	0.023	0.32	0.82
post09	SE	0.97	0.03			-0.34	2.24	0.055	0.18	0.83
post10A	RNE	0.29	0.71			0.80	-0.64	0.023	0.32	0.82
post10B	RNE	0.63	0.19	0.16	0.02	0.66	1.09	0.014	0.57	0.82
post11	SE	0.34	0.66			0.68	-0.47	0.022	0.33	0.82
post12	RNE	0.24	0.76			0.67	-0.83	0.025	0.31	0.82
post13	RNE	0.40	0.60			0.64	-0.28	0.022	0.33	0.82
post14	SE	0.42	0.58			0.55	-0.24	0.022	0.33	0.82
post15	SE	0.47	0.53			0.54	-0.09	0.021	0.34	0.82
post16	SE	0.54	0.46			0.50	0.12	0.022	0.33	0.83
post17	PA	0.50	0.50			0.59	0.01	0.021	0.34	0.82
post18	SE	0.60	0.40			0.47	0.29	0.022	0.33	0.83
post19	SE	0.82	0.16	0.01		0.61	1.62	0.025	0.39	0.82
post20	PA	0.84	0.23			0.56	0.85	0.025	0.31	0.82
post21	PA	0.77	0.20			0.36	0.98	0.026	0.3	0.83

		p-value				Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.931)	Alpha=.81
Item	Domain	0	1	2	3	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
post22	SE	0.80	0.46			0.49	0.12	0.022	0.33	0.83
post23	PA	0.54	0.43			0.51	0.22	0.022	0.33	0.83
post24	PA	0.77	0.22			0.66	0.89	0.025	0.31	0.82
post25	SE	0.33	0.66	0.01		0.46	1.31	0.022	0.34	0.83
post26	RNE	0.89	0.10			0.67	1.50	0.033	0.26	0.83
post27A	SE	0.90	0.40			0.73	0.30	0.022	0.33	0.82
post27B	SE	0.60	0.14			0.73	1.29	0.030	0.28	0.82
post27C	SE	0.87	0.17			0.68	1.11	0.027	0.29	0.82

Appendix K:

Grade 7 Transfer Measure 2008/2009

PowerSource

Answer each question below.

1. Which of these expressions is equivalent to $n \cdot n \cdot n$ for all values of n?

a) $\frac{n}{3}$

b) $n + 3$

c) $3n$

d) n^3

2. Subtract: $\frac{3x}{7} - \frac{x}{7} =$

a) $\frac{3}{7}$

b) 3

c) $2x$

d) $\frac{x}{7}$

e) $\frac{2x}{7}$

3. Which of these is equal to $(370 \cdot 998) + (370 \cdot 2)$?

a) $370 \cdot 1{,}000$

b) $372 \cdot 998$

c) $740 \cdot 998$

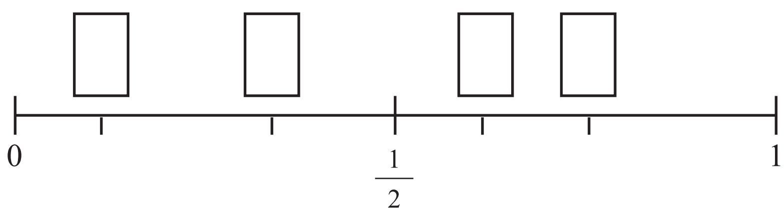
d) $370 \cdot 998 \cdot 2$

4. Which written expression could be represented by $37 - 3n = 5$?

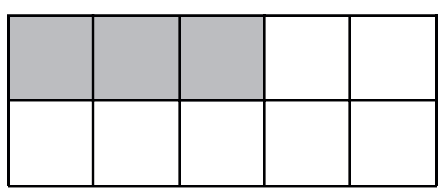
a) The sum of 37 and 3 times a number is 5.
b) The product of *n* and 37 decreased by 3 is 5.
c) Three times a number decreased by 37 is 5.
d) Thirty-seven decreased by 3 times a number is 5.

5. Here are four fractions: $\frac{3}{4}$, $\frac{1}{8}$, $\frac{1}{3}$ and $\frac{3}{5}$.

Look at the number line below. Write each fraction in the correct box.

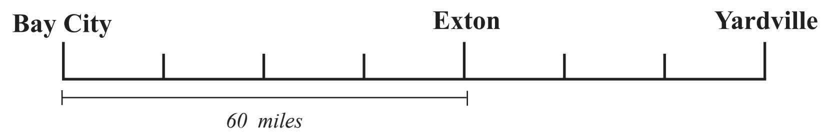


6. In the figure, how many MORE small squares need to be shaded so that $\frac{4}{5}$ of the small squares are shaded?



a) 5
b) 4
c) 3
d) 2
e) 1

7.



On the road shown above, the distance from Bay City to Exton is 60 miles. What is the distance from Bay City to Yardville?

a) 45 miles
b) 120 miles
c) 90 miles
d) 105 miles

8. There are 300 calories in 100 g of a certain food, how many calories are there in a 30 g portion of this food?

a) 90
b) 100
c) 900
d) 30
e) 10

9. I think of a number. I multiply this number by 8, then subtract 66. The result is twice the number that I was thinking of. Which equation represents this situation?

a) $8n - 66 = 2n$
b) $n + 8 - 66 = 2 + n$
c) $8n \bullet 66 = 2n$
d) $8 + n \bullet 66 = 2 + n$

10. If 4 times a number is 48, what is $\frac{1}{3}$ of the number?

a) 4
b) 8
c) 12
d) 16

11. A garden has 14 rows. Each row has 20 plants. The gardener then plants x more rows with 20 plants in each row.

Use the distributive property to write an expression to show how many plants there are altogether.

12. Jim has $\frac{3}{4}$ of a yard of string which he wishes to divide into pieces, each $\frac{1}{8}$ of a yard long. How many pieces will he have?

a) 3
b) 4
c) 6
d) 8

13. Fifteen boxes each containing 8 radios can be repacked in 10 larger boxes each containing how many radios?

a) 8
b) 10
c) 12
d) 80
e) 120

14. $\frac{3}{5} + \left(\frac{3}{10} \cdot \frac{4}{15}\right) =$

a) $\frac{3}{51}$

b) $\frac{1}{6}$

c) $\frac{6}{25}$

d) $\frac{11}{25}$

e) $\frac{17}{25}$

15. Robin and Jim took cherries from a basket. Robin took $\frac{1}{3}$ of the cherries and Jim took $\frac{1}{6}$ of the cherries. What fraction of the cherries remained in the basket?

a) $\frac{1}{2}$

b) $\frac{1}{3}$

c) $\frac{1}{6}$

d) $\frac{1}{18}$

16. The cost, *c,* of printing business cards consists of a fixed charge of 100 cents and a charge of 6 cents of each card printed. Which of these equations can be used to determine the cost of printing *n* cards?

a) $c = (100 + 6n)$
b) $c = (106 + n)$
c) $c = (6 + 100n)$
d) $c = (106n)$
e) $c = (600n)$

17. The fraction $2\frac{1}{4}$ means $2 + \frac{1}{4}$, which can also be written as $(2 + \frac{1}{4})$. Show how you would use the distributive property to multiply $2\frac{1}{4}$ by 10.

18. A rectangular playground has a **perimeter** of 390 feet. The width of the playground is 75 feet. What is its length?

a) 5.2 feet
b) 97.5 feet
c) 120 feet
d) 130 feet
e) 240 feet

19. Graham has twice as many books as Bob. Chan has six more books than Bob. If Bob has b books, which of the following represents the total number of books the **three** boys (Graham, Bob and Chan) have?

a) $3b + 6$
b) $3b + 8$
c) $4b + 6$
d) $5b + 6$
e) $8b + 2$

20. Daniel had 31 baseball cards. He gave the cards to his friends. Six of his friends received 3 cards each. Seven of his friends received 1 card each. The rest received 2 cards each. How many of his friends received exactly 2 cards from Daniel? Explain how you found your answer.

21. The screens of widescreen and standard televisions look different. Widescreen television ratio of height to width is 9:16. Standard television ratio of height to width is 3:4. Keri starts to draw scale drawings of the televisions. For each, the height is 4.5 cm. What should the width of each scale drawing be?

Widescreen

The width of this scale drawing should be _____cm

Standard

The width of this scale drawing should be _____cm

22. A painter had 25 L of paint. He used 2.5 L of paint every hour. He finished the job in 5.5 hours. How much paint did he have left?

a) 10.25 L
b) 11.25 L
c) 12.75 L
d) 13.75 L

23. John sold 60 magazines and Mark sold 80 magazines. The magazines were all sold for the same price. The total amount of money received for the magazines was $700.

a) Write an equation to find the cost of a magazine.

b) Solve the equation to find out how much each magazine cost.

c) How much money did **each** boy make?

24. A book publisher sent 140 copies of a book to a bookstore. The publisher packed the books in two types of boxes. On type of box held 8 copies of the book, and the other type of box held 12 copies of the book. The boxes were all full, and there were **equal** numbers of both types of boxes.

a) How many full boxes of 12 books were there?

b) What fraction of the books were packed in the smaller boxes?

25. In one week Jamal watched television for 26 hours. In that week: He watched television for the same length of time on Monday, Tuesday, Wednesday and Thursday. On each Friday, Saturday and Sunday, he watched television for twice as long as on Monday. How long did he spend watching television on **Saturday**? Write your answer in hours and minutes.

26. A biologist needs to estimate the size of the deer herd on a wildlife reserve. The biologist captures 150 deer, then tags and releases them. A week later, the biologist captures 50 deer and counts the number tagged and the number of untagged deer. There are 15 tagged deer and 35 untagged deer in this group.

a) If the number of deer in the herd is represented by the unknown *d*, write an equation that shows the ratio of tagged deer to total deer in the captured group is equal to the ratio of tagged deer to total deer in the entire herd.

b) How many untagged deer are in the total herd? Show your calculations.

Appendix L:

Item Analysis Results of PS Grade 7 Posttest

Table L1

Item Analysis Results of PS Grade 7 Posttest

		p-value			Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.924)	Alpha=.85
Item	Domains	0	1	2	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
post01	PA	0.09	0.91		0.4	-1.647	0.039	0.25	0.85
post02	RNE	0.48	0.52		0.5	-0.063	0.023	0.34	0.85
post03	PA	0.75	0.25		0.4	0.792	0.027	0.31	0.85
post04	SE	0.40	0.61		0.5	-0.322	0.024	0.33	0.85
post05	RNE	0.55	0.45		0.8	0.143	0.023	0.33	0.84
post06	RA	0.53	0.47		0.7	0.085	0.023	0.34	0.84
post07	RA	0.65	0.35		0.5	0.459	0.025	0.33	0.85
post08	RA	0.52	0.48		0.5	0.046	0.023	0.34	0.85
post09	SE	0.21	0.79		0.7	-0.969	0.028	0.3	0.85
post10	SE	0.52	0.48		0.7	0.042	0.023	0.34	0.84
post11	PA	0.75	0.25		0.7	0.805	0.027	0.31	0.84
post12	RNE	0.44	0.56		0.6	-0.187	0.023	0.33	0.85
post13	SE	0.54	0.46		0.5	0.119	0.023	0.33	0.85
post14	RNE	0.63	0.37		0.6	0.377	0.024	0.33	0.85
post15	RNE	0.51	0.50		0.7	0.006	0.023	0.34	0.84
post16	SE	0.38	0.63		0.5	-0.381	0.024	0.33	0.85
post17	PA	0.86	0.14		0.6	1.301	0.033	0.28	0.85
post18	SE	0.59	0.41		0.6	0.267	0.024	0.33	0.85
post19	SE	0.82	0.18		0.3	1.118	0.030	0.29	0.85
post20	SE	0.60	0.23	0.16	0.7	0.612	0.018	0.51	0.85
post21	RA	0.93	0.07		0.9	1.854	0.044	0.22	0.85
post22	SE	0.66	0.34		0.5	0.479	0.025	0.33	0.85

		p-value			Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.924)	Alpha=.85
Item	Domains	0	1	2	$r_{poly.}$	b	*SE*(b)	Item reliability	If deleted
post23A	SE	0.75	0.25		0.8	0.803	0.027	0.31	0.84
post23B	SE	0.85	0.16		0.6	1.225	0.032	0.28	0.85
post23C	SE	0.72	0.28		0.8	0.678	0.026	0.32	0.84
post24A	SE	0.84	0.16		0.8	1.187	0.031	0.29	0.85
post24B	RA	0.94	0.06		0.8	1.905	0.046	0.22	0.85
post25	SE	0.97	0.03		0.8	2.500	0.068	0.15	0.85
post26A	RA	0.98	0.02		0.8	2.770	0.084	0.12	0.85
post26B	RA	0.97	0.03		0.8	2.376	0.063	0.17	0.85

Appendix M:

Grade 6 Transfer Measure 2008/2009

Powersource

Answer each question below. Circle your answer.

1. What do you need to add to eighty-three to make one hundred?

2. Write the fraction $\frac{3}{9}$ in its simplest form.

3. There were two thousand people at a concert. Nine hundred and ninety-two of them were women. How many of the people were not women?

4. Write a fraction that is less than $\frac{4}{9}$.

5. Write a fraction that has a denominator of 100 **and** is equivalent to $\frac{7}{20}$.

6. What value of x makes the equation true?

$x - 9 = 32$

a) 23
b) 41
c) 32
d) 9

7. Solve: $6n = 36$

a) 12
b) 2
c) 30
d) 6

8. What is the next step to solve this equation?

$x - 7 = 13$

a) Subtract 7 from both sides
b) Add x to both sides
c) Add 7 to both sides
d) Subtract 13 from both sides

9. Write a different fraction that is equivalent to three-fifths.

10. $b = 14 + a$. When a equals 7, what is the value of b?

11. If $\frac{12}{n} = \frac{36}{21}$, then n equals:

a) 3
b) 7
c) 36
d) 63

12. Which of the following ratios is equivalent to the ratio of 6 to 4?

a) 12 to 18
b) 12 to 8
c) 8 to 6
d) 4 to 6
e) 2 to 3

13. Charlie can type 32 words per minute. At this rate, how long would it take him, in minutes, to type 128 words?

a) 1
b) 3
c) 4
d) 2

14. Sam's uncle is 21 years older than Sam. His uncle is 42. What equation could you use to solve for Sam's age, *s* ?

a) $s + 21 = 42$
b) $\frac{42}{21} = s$
c) $s - 21 = 42$
d) $s - 42 = 21$

15. Which of the following shows the distributive property being used correctly to simplify the expression: 3(4) + 3(2)

a) 3(4)(2)
b) 3(4 + 2)
c) 4(3 + 2)
d) 4(3) + 2(3)

16. What is the value of p in the equation below ?

$$\frac{1}{4}p = 4$$

a) $p = 4$

b) $p = 16$

c) $p = 4\frac{1}{4}$

d) $p = 3\frac{3}{4}$

17. For all numbers k,

$k + k + k + k + k$ can be written as

a) $k + 5$
b) $5k$
c) k^5
d) $5\ (k + 1)$

18. Which of the following is equal to $6\ (x + 6)$?

a) $x + 12$
b) $6x + 6$
c) $6x + 12$
d) $6x + 36$
e) $6x + 66$

19. Simplify using the distributive property.

$y\,(y - 6) =$

20. How much change will John get back from $5.00 if he buys 2 notebooks that cost $1.80 each?

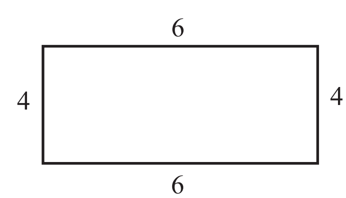
a) $1.40
b) $2.40
c) $3.20
d) $3.60

21. The perimeter of a square is 36 inches. What is the length of one side of the square?

a) 4 inches
b) 6 inches
c) 9 inches
d) 18 inches

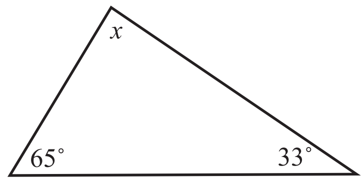
22. Which of the following numerical expressions gives the area of the rectangle below?

a) $4 \bullet 6$
b) $4 + 6$
c) $2\,(4 \bullet 6)$
d) $2\,(4 + 6)$
e) $4 + 6 + 4 + 6$



23. What is the value of x in the triangle?

a) 65º
b) 82º
c) 90º
d) 92º
e) 98º



24. If $3 + w = b$, then $w =$

a) $\frac{3}{9}$

b) $b \cdot 3$
c) $b + 3$
d) $3 - b$
e) $b - 3$

25. In which list of fractions are all of the fractions equivalent?

a) $\frac{1}{2}, \frac{2}{4}, \frac{4}{6}$

b) $\frac{2}{3}, \frac{4}{6}, \frac{8}{12}$

c) $\frac{2}{5}, \frac{4}{10}, \frac{8}{50}$

d) $\frac{3}{4}, \frac{4}{6}, \frac{6}{8}$

26. n is a number. When n is multiplied by 7, and 6 is then added, the result is 41. Which of these equations represents this relation?

a) $7n + 6 = 41$
b) $7n + {}^{-}6 = 41$
c) $7n \bullet 6 = 41$
d) $7(n + 6) = 41$

27. Explain why the fraction $\dfrac{\frac{1}{2}}{\frac{3}{4}}$ is equivalent to the fraction $\frac{2}{3}$?

28. What would be your answer if you were asked to multiply $8 \bullet (x + \frac{3}{4})$?

a) $8x + \frac{3}{4}$

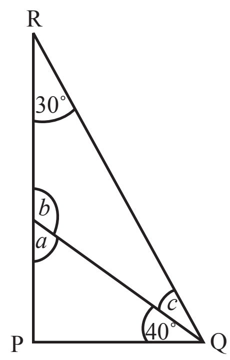
b) $8\frac{3}{4}x$

c) $8x + 6$

d) $x + 6$

29. The diagram shows triangle PQR.

What are the sizes of the angles *a*, *b*, and *c* ?



Appendix N:

Item Analysis Results of PS Grade 6 Intertest

Table N1

Item Analysis Results of PS Grade 6 Intertest

		p-value				Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.918)	Alpha=.83
Item	Domain	0	1	2	3	rpoly.	b	*SE*(b)	Item reliability	If deleted
inter01	RNE	0.12	0.88			0.53	-1.43	0.03	0.27	0.82
inter02	RNE	0.19	0.81			0.74	-1.05	0.03	0.30	0.82
inter03	RNE	0.39	0.61			0.76	-0.33	0.02	0.33	0.81
inter04	RNE	0.26	0.74			0.47	-0.78	0.03	0.31	0.82
inter05	PA	0.27	0.73			0.65	-0.73	0.02	0.32	0.82
inter06	RNE	0.21	0.79			0.78	-0.96	0.03	0.30	0.81
inter07	RNE	0.41	0.59			0.80	-0.27	0.02	0.33	0.81
inter08	RNE	0.44	0.57			0.63	-0.19	0.02	0.33	0.82
inter09	RNE	0.86	0.14	0.00		0.56	2.68	0.03	0.29	0.82
inter10	RNE	0.39	0.61			0.55	-0.34	0.02	0.33	0.82
inter11	RNE	0.51	0.49			0.64	0.03	0.02	0.34	0.82
inter12	PA	0.41	0.59			0.63	-0.27	0.02	0.33	0.82
inter13	PA	0.46	0.54			0.49	-0.13	0.02	0.33	0.82
inter14	PA	0.59	0.41			0.48	0.28	0.02	0.33	0.82
inter15	PA	0.73	0.27			0.45	0.74	0.02	0.32	0.82
inter16	PA	0.85	0.15			0.38	1.25	0.03	0.28	0.83
inter17	PA	0.89	0.11			0.66	1.48	0.03	0.26	0.82
inter18A	RNE	0.36	0.64			0.62	-0.41	0.02	0.33	0.82
inter18B	RNE	0.49	0.03	0.49		0.80	0.00	0.01	0.58	0.82
inter19A	PA	0.55	0.45			0.62	0.14	0.02	0.33	0.82

		p-value				Polyserial correlation	Rasch difficulty		IRT reliability (test Reli.=.918)	Alpha=.83
Item	Domain	0	1	2	3	rpoly.	b	*SE*(b)	Item reliability	If deleted
inter19B	PA	0.73	0.06	0.16	0.04	0.80	0.98	0.01	0.61	0.82
inter20	PA	0.71	0.28	0.01		0.62	1.55	0.02	0.38	0.82

Appendix O:
Descriptive Statistics by Content Domain

Descriptive Statistics of Subdomain: PA

We also calculated descriptive statistics of subscores. There are 9 items measured PA on the pretest, while 4 PA items are on the posttest in Grade 7. PA items on the posttest seem more difficult than those on the pretest. On average, students got less than half of the items on the pretest and posttest correct. POWERSOURCE© students did not perform far higher than control group students on PA posttest items. The observed difference on the posttest is 0.18 in the B-S design, while the difference is 0.05 in the W-S design. The treatment group students have higher scores both on pretest and posttest than control group students. Specifically, PA posttest mean for POWERSOURCE© and control students in B-S design are 1.21 and 1.65, respectively, and the difference is 0.44, which is about a 0.38 pooled-standard deviation. This difference is even larger, 0.77 (0.64 pooled-*SD*), in W-S design.

Table O1

Descriptive Statistics of PA Scores on Pretest (Grade 7)

Design			Pretest PA			
		N	Mean	*SD*	Min	Max
Between	Control	689	3.73	1.53	0	9
	Treatment	567	3.57	1.58	0	8
Within	Control	527	4.13	1.68	0	9
	Treatment	810	4.02	1.68	0	9

Table O2

Descriptive Statistics of PA Scores on Posttest (Grade 7)

Design			Posttest PA			
		N	Mean	*SD*	Min	Max
Between	Control	689	1.42	0.76	0	4
	Treatment	567	1.60	0.91	0	4
Within	Control	527	1.61	0.81	0	4
	Treatment	810	1.66	0.92	0	4

For Grade 6 PA, POWERSOURCE© students did perform better than control group students both on the interim and posttests. On the interim test, POWERSOURCE© students in W-S design outperformed control students by more than 1.5 points which is close to a 0.6

pooled standard deviation. Likewise, in W-S design, the difference in posttest score is as large as a 0.6 pooled standard deviation. However, on both the interim test and posttest, the observed means for POWERSOURCE© students and control students are very similar.

Table O3

Descriptive Statistics of PA Scores on Pretest (Grade 6)

Design			Pretest PA			
		N	Mean	*SD*	Min	Max
Between	Control	656	5.27	1.75	0	8
	Treatment	1,034	4.87	1.87	0	8
Within	Control	593	5.30	1.75	0	8
	Treatment	755	5.78	1.50	0	8

Table O4

Descriptive Statistics of PA Scores on Interim Test (Grade 6)

Design			Interim test PA			
		N	Mean	*SD*	Min	Max
Between	Control	656	3.75	2.37	0	12
	Treatment	1,034	4.15	2.58	0	12
Within	Control	593	3.49	2.05	0	13
	Treatment	755	5.19	3.02	0	13

Table O5

Descriptive Statistics of PA Scores on Posttest (Grade 6)

Design			Posttest PA			
		N	Mean	*SD*	Min	Max
Between	Control	691	1.55	1.30	0	5
	Treatment	1,055	1.67	1.26	0	5
Within	Control	569	1.20	1.11	0	5
	Treatment	768	1.94	1.59	0	5

Descriptive Statistics of Subscore: RNE

There are 15 RNE items on pretest and 5 items on posttest for Grade 7. However, there are 6, 12, and 8 RNE items for Grade 6, respectively on pretest, interim test, and posttest. Dissimilar to PA, Grade 7 RNE mean pretest score for POWERSOURCE© students is very close to one for control students. Note that students in W-S design have a point higher pretest score than those in the control group. Posttest means both in B-S design and in W-S design are very similar for each of the two groups. These observed posttest differences in B-S design and in W-S design are, respectively, 0.04 and 0.01.

Table O6

Descriptive Statistics of RNE Scores on Pretest (Grade 7)

Design			Pretest RNE			
		N	Mean	*SD*	Min	Max
Between	Control	689	6.84	2.71	0	15
	Treatment	567	6.93	2.45	1	14
Within	Control	527	8.28	2.73	0	15
	Treatment	810	8.02	3.10	0	15

Table O7

Descriptive Statistics of RNE Scores on Posttest (Grade 7)

Design			Posttest RNE			
		N	Mean	*SD*	Min	Max
Between	Control	689	2.13	1.53	0	5
	Treatment	567	2.17	1.42	0	5
Within	Control	527	2.68	1.55	0	5
	Treatment	810	2.67	1.52	0	5

For the Grade 6 RNE, the differences on pretest are 0.1 in B-S design and 0.2 in W-S design. These differences are considered as very small, given the size of standard deviation (approximately 1.2). In terms of outcomes, it seems that are there some significant changes. For example, on pretest score in B-S design, POWERSOURCE© students had lower scores than control students, but they outperformed on the posttest. In addition, the observed mean difference on posttest between the two groups in W-S design is 0.73 (a 0.3 pooled standard deviation).

Table O8

Descriptive Statistics of RNE Scores on Pretest (Grade 6)

Design			Pretest RNE			
		N	Mean	*SD*	Min	Max
Between	Control	656	3.91	1.27	0	6
	Treatment	1,034	3.80	1.27	0	6
Within	Control	593	4.24	1.16	0	6
	Treatment	755	4.48	1.16	0	6

Table O9

Descriptive Statistics of RNE Scores on Interim Test (Grade 6)

Design			Interim test RNE			
		N	Mean	*SD*	Min	Max
Between	Control	656	8.03	3.30	0	13
	Treatment	1,034	7.65	3.32	0	13
Within	Control	593	8.28	2.86	0	13
	Treatment	755	8.46	3.18	0	13

Table O10

Descriptive Statistics of RNE Scores on Posttest (Grade 6)

Design			Posttest RNE			
		N	Mean	*SD*	Min	Max
Between	Control	691	4.65	2.21	0	11
	Treatment	1,055	4.88	2.31	0	13
Within	Control	569	5.05	2.08	0	10
	Treatment	768	5.78	2.54	0	12

Descriptive Statistics of Subdomain: SE

We included 3 SE items on pretest and 14 on posttest for Grade 7. The pretest scores are very similar across four different groups, control and POWERSOURCE© groups in B-S and W-S designs (see Tables O11 and O12). The POWERSOURCE© students in W-S students have a higher mean posttest score than control students by .21; whereas, control students outperformed POWERSOURCE© students in B-S design by 0.45 point. Note that both on pretest and on posttest, the overall performance of students on these SE items are seemingly low because the average scores are only 1/3 of the possible maximum scores.

Table O11

Descriptive Statistics of SE Scores on Pretest (Grade 7)

Design			Pretest SE			
		N	Mean	*SD*	Min	Max
Between	Control	689	1.05	0.80	0	3
	Treatment	567	0.97	0.80	0	3
Within	Control	527	1.20	0.87	0	3
	Treatment	810	1.18	0.87	0	3

Table O12

Descriptive Statistics of SE Scores on Posttest (Grade 7)

Design			Posttest SE			
		N	Mean	*SD*	Min	Max
Between	Control	689	5.31	3.15	0	14
	Treatment	567	4.86	2.91	0	14
Within	Control	527	5.53	3.12	0	14
	Treatment	810	5.74	3.21	0	15

Tables O13 and O14 present the descriptive statistics for Grade 6 SE scores. In B-S design, control group students' pretest score was 0.24 points higher than treatment students. Yet, in W-S design POWERSOURCE© students scored 0.38 points higher than control students. This pattern also holds for posttest scores. The observed mean score difference was 0.4 points (favorable for POWERSOURCE© students in the W-S design); whereas, it was 0.6 points favorable for control students in the B-S design.

Table O13

Descriptive Statistics of SE Scores on Pretest (Grade 6)

Design			Pretest *SE*			
		N	Mean	*SD*	Min	Max
Between	Control	806	4.91	1.21	0	7
	Treatment	1,050	4.67	1.24	0	7
Within	Control	579	4.80	1.16	0	7
	Treatment	745	5.18	1.27	0	7

Table O14

Descriptive Statistics of SE Scores on Posttest (Grade 6)

Design			Posttest *SE*			
		N	Mean	*SD*	Min	Max
Between	Control	806	8.32	3.01	1	15
	Treatment	1,050	7.71	2.97	0	16
Within	Control	579	8.07	2.72	1	15
	Treatment	745	8.50	3.26	0	16

Appendix P:

Additional Descriptive Statistics for Grades 6 and 7

Table P1

Grade 7 Descriptive Statistics of Pretest Scores by District and Treatment

District	Design	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	Control	147	12.74	4.52	5	22
		Treatment	73	10.56	3.16	5	18
CA-1	Within	Control	260	13.23	3.32	4	23
		Treatment	380	15.11	4.70	5	27
CA-2	Within	Control	70	11.11	3.77	2	23
		Treatment	297	11.04	3.29	3	22
CA-3	Within	Control	57	11.53	3.71	0	18
		Treatment	33	12.18	3.47	6	19
CA-4	Between	Control	445	10.71	3.53	0	21
		Treatment	417	11.31	3.71	1	23
CA-5	Between	Control	97	14.04	4.07	6	24
		Treatment	77	13.17	3.38	7	22
	Within	Control	140	16.41	4.88	0	24
		Treatment	100	12.87	5.74	0	23

Table P2

Grade 7 Descriptive Statistics of Posttest scores by District and Treatment

District	Design	Treatment	*N*	Posttest Mean	*SD*	Min	Max
AZ-1	Between	Control	147	14.18	6.09	2	28
		Treatment	73	11.14	6.02	2	24
CA-1	Within	Control	260	9.77	4.23	1	23
		Treatment	380	12.09	6.10	0	29
CA-2	Within	Control	70	9.46	5.42	0	23
		Treatment	297	11.20	5.59	2	26
CA-3	Within	Control	57	13.12	5.25	2	26
		Treatment	33	9.00	5.15	3	23
CA-4	Between	Control	445	9.43	5.32	1	26
		Treatment	417	9.34	4.67	1	26
CA-5	Between	Control	97	8.59	4.19	3	22
		Treatment	77	12.17	5.76	3	26
	Within	Control	140	14.36	6.19	3	27
		Treatment	100	12.39	5.08	4	26

Table P3

Grade 7 Descriptive Statistics of Pretest Scores by School

District	Design	School	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	#1	Control	147	12.74	4.52	5	22
		#2	Treatment	59	9.83	2.97	5	18
		#3	Treatment	14	13.64	1.82	11	17
CA-1	Within	#4	Control	209	13.26	3.35	4	23
			Treatment	72	12.17	3.47	5	19
		#5	Control	51	13.08	3.21	7	20
			Treatment	308	15.80	4.69	5	27
CA-2	Within	#6	Control	27	12.33	4.44	5	23
			Treatment	206	10.73	3.21	3	22
		#7	Control	43	10.35	3.09	2	16
			Treatment	91	11.75	3.38	4	22
CA-3	Within	#8	Control	57	11.53	3.71	0	18
			Treatment	33	12.18	3.47	6	19
CA-4	Between	#9	Control	157	11.35	3.99	4	21
		#10	Control	233	10.34	3.17	0	21
		#11	Control	55	10.45	3.40	4	21
		#12	Treatment	288	11.06	3.65	1	20
		#13	Treatment	129	11.86	3.81	3	23
CA-5	Between	#14	Treatment	13	13.92	2.81	9	18
		#15	Treatment	33	12.64	3.27	7	20
		#16	Treatment	31	13.42	3.71	7	22
		#17	Control	19	11.79	2.35	6	15
		#18	Control	78	14.59	4.23	7	24
	Within	#19	Control	24	10.29	2.35	7	17
			Treatment	25	14.28	4.13	6	21
		#20	Control	75	18.75	4.39	0	24
			Treatment	52	11.40	6.79	0	22
		#21	Control	41	15.73	3.26	8	22
			Treatment	23	14.65	3.47	10	23

Table P4

Grade 7 Descriptive Statistics of Posttest Scores by School

District	Design	School	Treatment	*N*	Posttest Mean	*SD*	Min	Max
AZ-1	Between	#1	Control	147	14.18	6.09	2	28
		#2	Treatment	59	9.32	4.99	2	24
		#3	Treatment	14	18.79	3.42	11	24
CA-1	Within	#4	Control	209	10.16	4.20	1	23
			Treatment	72	10.60	6.53	1	24
		#5	Control	51	8.16	4.02	2	19
			Treatment	308	12.44	5.95	0	29
CA-2	Within	#6	Control	27	9.44	5.96	0	23
			Treatment	206	10.04	4.75	2	25
		#7	Control	43	9.47	5.13	3	22
			Treatment	91	13.80	6.45	2	26
CA-3	Within	#8	Control	57	13.12	5.25	2	26
			Treatment	33	9.00	5.15	3	23
CA-4	Between	#9	Control	157	10.82	5.47	1	26
		#10	Control	233	8.66	5.10	1	24
		#11	Control	55	8.71	5.07	1	22
		#12	Treatment	288	9.40	4.64	1	26
		#13	Treatment	129	9.20	4.75	1	22
CA-5	Between	#14	Treatment	13	12.23	5.13	4	19
		#15	Treatment	33	10.70	5.29	4	20
		#16	Treatment	31	13.71	6.24	3	26
		#17	Control	19	8.26	3.23	4	16
		#18	Control	78	8.67	4.40	3	22
	Within	#19	Control	24	6.67	2.46	3	13
			Treatment	25	11.88	5.95	4	22
		#20	Control	75	17.55	4.87	7	27
			Treatment	52	12.23	4.81	5	26
		#21	Control	41	13.02	5.45	4	23
			Treatment	23	13.30	4.77	6	23

Table P5

Grade 7 Descriptive Statistics of Pretest Scores by Teacher

District	Design	School	Teacher ID	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	#1	133	Control	88	10.30	3.07	5	20
			139	Control	59	16.39	3.84	6	22
		#2	137	Treatment	59	9.83	2.97	5	18
		#3	120	Treatment	14	13.64	1.82	11	17
CA-1	Within	#4	218	Treatment	44	13.52	3.38	7	19
			229	Control	135	13.61	3.22	4	23
			237	Control	74	12.62	3.51	6	22
			243	Treatment	28	10.04	2.40	5	15
		#5	202	Treatment	22	16.32	2.71	11	21
			235	Treatment	56	21.80	2.23	15	27
			238	Treatment	141	16.57	2.83	9	23
			239	Control	51	13.08	3.21	7	20
			240	Treatment	89	10.67	2.95	5	17
CA-2	Within	#6	308	Control	27	12.33	4.44	5	23
			313	Treatment	94	11.30	3.02	3	19
			314	Treatment	32	10.03	3.07	4	18
			315	Treatment	80	10.34	3.41	5	22
		#7	301	Treatment	50	11.78	3.22	4	18
			304	Control	43	10.35	3.09	2	16
			311	Treatment	41	11.71	3.59	4	22
CA-3	Within	#8	411	Treatment	33	12.18	3.47	6	19
			413	Control	57	11.53	3.71	0	18

District	Design	School	Teacher ID	Treatment	*N*	Pretest Mean	*SD*	Min	Max
CA-4	Between	#9	801	Control	62	9.90	2.13	5	15
			803	Control	56	9.52	2.73	4	15
			804	Control	36	16.94	3.04	9	21
			860	Control	3	8.33	1.53	7	10
		#10	817	Control	26	14.50	3.40	8	21
			854	Control	62	10.44	2.88	3	17
			865	Control	72	9.99	2.45	4	16
			867	Control	46	9.20	3.10	0	16
			868	Control	27	9.04	2.08	5	13
		#11	833	Control	51	10.71	3.31	4	21
			838	Control	4	7.25	3.30	5	12
		#12	811	Treatment	57	8.88	3.44	1	16
			841	Treatment	29	10.45	3.89	5	19
			842	Treatment	36	9.61	2.92	3	14
			844	Treatment	69	14.36	2.94	7	20
			845	Treatment	97	10.72	2.77	6	19
		#13	819	Treatment	26	13.15	3.92	7	21
			823	Treatment	29	13.55	3.32	4	19
			843	Treatment	74	10.74	3.60	3	23
CA-5	Between	#14	905	Treatment	13	13.92	2.81	9	18
		#15	906	Treatment	33	12.64	3.27	7	20
		#16	907	Treatment	31	13.42	3.71	7	22
		#17	913	Control	19	11.79	2.35	6	15
		#18	914	Control	49	13.31	3.91	7	23
			915	Control	29	16.76	3.89	8	24
	Within	#19	908	Treatment	25	14.28	4.13	6	21
			911	Control	24	10.29	2.35	7	17
		#20	909	Treatment	52	11.40	6.79	0	22
			916	Control	75	18.75	4.39	0	24
		#21	903	Control	41	15.73	3.26	8	22
			910	Treatment	23	14.65	3.47	10	23

Table P6

Grade 7 Descriptive Statistics of Posttest Scores by Teacher

District	Design	School	Teacher ID	Treatment	*N*	Post Mean	*SD*	Min	Max
AZ-1	Between	#1	133	Control	88	11.06	3.84	3	20
			139	Control	59	18.83	5.86	2	28
		#2	137	Treatment	59	9.32	4.99	2	24
		#3	120	Treatment	14	18.79	3.42	11	24
CA-1	Within	#4	218	Treatment	44	12.73	7.16	1	24
			229	Control	135	10.09	4.25	1	23
			237	Control	74	10.30	4.13	2	21
			243	Treatment	28	7.25	3.37	3	18
		#5	202	Treatment	22	13.05	4.38	6	20
			235	Treatment	56	20.07	4.55	6	29
			238	Treatment	141	12.81	4.24	3	23
			239	Control	51	8.16	4.02	2	19
			240	Treatment	89	6.92	3.04	0	16
CA-2	Within	#6	308	Control	27	9.44	5.96	0	23
			313	Treatment	94	10.37	5.06	2	24
			314	Treatment	32	8.53	4.00	2	17
			315	Treatment	80	10.26	4.58	4	25
		#7	301	Treatment	50	17.28	5.36	7	26
			304	Control	43	9.47	5.13	3	22
			311	Treatment	41	9.56	5.00	2	23
CA-3	Within	#8	411	Treatment	33	9.00	5.15	3	23
			413	Control	57	13.12	5.25	2	26
CA-4	Between	#9	801	Control	62	8.68	3.92	2	24
			803	Control	56	9.64	5.08	1	26
			804	Control	36	17.03	3.19	10	22
			860	Control	3	2.67	1.15	2	4
		#10	817	Control	26	18.35	4.11	10	24
			854	Control	62	8.06	3.84	1	18
			865	Control	72	8.51	4.15	1	22
			867	Control	46	5.93	2.77	2	17
			868	Control	27	5.70	1.96	1	9
		#11	833	Control	51	9.16	4.96	1	22

District	Design	School	Teacher ID	Treatment	*N*	Post Mean	*SD*	Min	Max
			838	Control	4	3.00	2.16	1	6
			811	Treatment	57	8.37	4.18	1	21
			841	Treatment	29	7.72	4.10	3	20
		#12	842	Treatment	36	8.06	4.29	2	20
			844	Treatment	69	13.41	4.80	2	26
			845	Treatment	97	8.14	3.33	2	20
			819	Treatment	26	11.58	5.44	3	22
		#13	823	Treatment	29	9.17	4.96	2	21
			843	Treatment	74	8.38	4.16	1	21
CA-5	Between	#14	905	Treatment	13	12.23	5.13	4	19
		#15	906	Treatment	33	10.70	5.29	4	20
		#16	907	Treatment	31	13.71	6.24	3	26
		#17	913	Control	19	8.26	3.23	4	16
		#18	914	Control	49	9.63	4.85	3	22
			915	Control	29	7.03	2.92	4	18
	Within	#19	908	Treatment	25	11.88	5.95	4	22
			911	Control	24	6.67	2.46	3	13
		#20	909	Treatment	52	12.23	4.81	5	26
			916	Control	75	17.55	4.87	7	27
		#21	903	Control	41	13.02	5.45	4	23
			910	Treatment	23	13.30	4.77	6	23

Table P7

Grade 6 Descriptive Statistics of Pretest Scores by District and Treatment

District	Design	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	Control	167	18.95	3.87	7	25
		Treatment	223	15.57	4.54	0	26
CA-1	Within	Control	361	19.13	4.02	0	27
		Treatment	494	21.19	3.82	0	28
CA-2	Within	Control	218	18.46	4.12	5	25
		Treatment	251	18.51	4.08	5	27
CA-3	Between	Control	58	14.53	5.10	4	25
		Treatment	84	17.69	4.69	6	26
CA-4	Between	Control	494	19.12	3.86	0	27
		Treatment	705	18.28	4.19	0	27
CA-5	Between	Control	87	21.03	3.97	10	28
		Treatment	38	19.21	5.06	7	25

Table P8

Grade 6 Descriptive Statistics of Interim Test Scores by District and Treatment

District	Design	Treatment	*N*	Interimtest Mean	*SD*	Min	Max
AZ-1	Between	Control	178	13.53	4.57	2	25
		Treatment	231	11.87	4.77	1	23
CA-1	Within	Control	361	11.29	3.91	2	21
		Treatment	488	15.48	5.33	0	25
CA-2	Within	Control	208	12.79	4.44	2	26
		Treatment	280	10.24	4.62	0	22
CA-3	Between	Control	55	11.96	4.66	2	22
		Treatment	95	12.04	5.22	2	24
CA-4	Between	Control	374	9.70	4.51	0	23
		Treatment	684	11.55	5.36	0	25
CA-5	Between	Control	84	15.14	5.57	4	25
		Treatment	45	14.58	5.73	3	23

Table P9

Grade 6 Descriptive Statistics of Posttest Scores by District and Treatment

District	Design	Treatment	*N*	Posttest Mean	*SD*	Min	Max
AZ-1	Between	Control	167	16.99	4.92	5	27
		Treatment	223	15.13	5.10	3	31
CA-1	Within	Control	361	13.60	4.59	2	26
		Treatment	494	17.35	6.57	2	32
CA-2	Within	Control	218	15.51	4.67	4	27
		Treatment	251	14.28	5.78	1	32
CA-3	Between	Control	58	14.93	5.87	3	28
		Treatment	84	16.60	5.56	3	33
CA-4	Between	Control	494	13.39	4.88	1	27
		Treatment	705	13.56	5.49	0	30
CA-5	Between	Control	87	17.91	6.04	4	29
		Treatment	38	17.26	5.24	3	28

Table P10

Grade 6 Descriptive Statistics of Pretest Scores by School

District	Design	School	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	#1	Control	167	18.95	3.87	7	25
		#2	Treatment	171	15.63	4.32	4	26
		#3	Treatment	52	15.37	5.26	0	24
CA-1	Within	#4	Control	105	19.41	3.05	8	25
			Treatment	186	20.46	3.63	6	28
		#5	Control	160	18.26	4.75	0	25
			Treatment	100	20.61	4.42	0	27
		#6	Control	96	20.29	3.22	12	27
			Treatment	208	22.13	3.48	9	28
CA-2	Within	#7	Control	129	18.71	3.93	9	25
			Treatment	192	18.52	4.17	6	27
		#8	Control	89	18.10	4.38	5	25
			Treatment	59	18.46	3.84	5	25
CA-3	Between	#9	Control	55	14.36	5.15	4	25
		#10	Treatment	38	18.58	5.02	6	26
		#11	Control	3	17.67	3.06	15	21
		#12	Treatment	46	16.96	4.31	6	24

District	Design	School	Treatment	*N*	Pretest Mean	*SD*	Min	Max
CA-4	Between	#13	Control	117	19.30	4.02	8	26
		#14	Control	206	18.23	3.61	7	25
		#15	Control	112	20.76	3.41	11	27
		#16	Control	26	17.88	3.56	10	26
		#17	Treatment	333	18.05	4.37	0	26
		#18	Treatment	108	19.71	2.97	13	26
		#19	Treatment	74	16.81	3.93	5	26
		#20	Control	33	19.52	4.64	0	26
		#21	Treatment	190	18.45	4.31	6	27
CA-5	Between	#22	Treatment	19	21.63	2.71	17	25
		#23	Control	87	21.03	3.97	10	28
		#24	Treatment	19	16.79	5.74	7	25

Table P11

Grade 6 Descriptive Statistics of Interim Test Scores by School

District	Design	School	Treatment	*N*	Interim test Mean	*SD*	Min	Max
AZ-1	Between	#1	Control	178	13.53	4.57	2	25
		#2	Treatment	183	11.96	4.85	1	22
		#3	Treatment	48	11.52	4.45	2	23
CA-1	Within	#4	Control	112	11.52	3.81	2	20
			Treatment	187	13.97	5.29	0	24
		#5	Control	155	10.78	3.58	2	21
			Treatment	106	15.00	5.67	3	25
		#6	Control	94	11.87	4.45	3	21
			Treatment	195	17.19	4.66	5	25
CA-2	Within	#7	Control	127	13.33	4.92	3	26
			Treatment	222	10.71	4.52	0	22
		#8	Control	81	11.94	3.41	2	19
			Treatment	58	8.45	4.60	1	18
CA-3	Between	#9	Control	55	11.96	4.66	2	22
		#10	Treatment	51	13.90	5.22	2	24
		#11	Treatment	44	9.89	4.36	3	18
CA-4	Between	#12	Control	123	11.28	4.86	0	23
		#13	Control	227	8.77	4.09	1	21
		#14	Control	24	10.33	4.32	4	20
		#15	Treatment	318	11.25	5.34	0	24
		#16	Treatment	118	12.98	5.06	2	23
		#17	Treatment	50	7.02	3.11	1	15

District	Design	School	Treatment	*N*	Interim test Mean	*SD*	Min	Max
		#18	Treatment	198	12.33	5.39	2	25
CA-5	Between	#19	Treatment	25	16.72	4.46	3	22
		#20	Control	84	15.14	5.57	4	25
		#21	Treatment	20	11.90	6.12	3	23

Table P12

Grade 6 Descriptive Statistics of Posttest Scores by School

District	Design	School	Treatment	*N*	Posttest Mean	*SD*	Min	Max
AZ-1	Between	#1	Control	167	16.99	4.92	5	27
		#2	Treatment	171	14.67	4.76	3	27
		#3	Treatment	52	16.63	5.88	5	31
CA-1	Within	#4	Control	105	13.94	4.31	3	25
			Treatment	186	15.53	6.11	2	31
		#5	Control	160	12.63	4.46	2	24
			Treatment	100	16.37	6.72	3	30
		#6	Control	96	14.83	4.79	4	26
			Treatment	208	19.46	6.31	4	32
CA-2	Within	#7	Control	129	16.39	4.72	5	27
			Treatment	192	14.95	5.70	1	32
		#8	Control	89	14.25	4.32	4	23
			Treatment	59	12.10	5.52	1	26
CA-3	Between	#9	Control	55	14.82	6.00	3	28
		#10	Treatment	38	16.76	6.65	5	33
		#11	Control	3	17.00	1.00	16	18
		#12	Treatment	46	16.46	4.55	3	27
CA-4	Between	#13	Control	117	14.15	5.03	3	25
		#14	Control	206	11.42	3.91	1	24
		#15	Control	112	15.66	5.13	5	27
		#16	Control	26	12.46	4.64	5	22
		#17	Treatment	333	13.14	5.46	0	27
		#18	Treatment	108	15.80	4.68	7	29
		#19	Treatment	74	11.03	4.44	1	20
		#20	Control	33	16.03	4.01	9	24
		#21	Treatment	190	14.02	5.83	1	30
CA-5	Between	#22	Treatment	19	19.16	3.72	11	24
		#23	Control	87	17.91	6.04	4	29
		#24	Treatment	19	15.37	5.92	3	28

Table P13

Grade 6 Descriptive Statistics of Pretest Scores by Teacher

District	Design	School	Teacher ID	Treatment	*N*	Pretest Mean	*SD*	Min	Max
AZ-1	Between	#1	118	Control	88	19.50	4.19	9	25
			128	Control	52	17.69	3.58	7	24
			129	Control	27	19.56	2.67	14	24
		#2	124	Treatment	100	14.68	3.85	6	24
			125	Treatment	51	17.90	4.53	5	26
			132	Treatment	20	14.60	3.99	4	22
		#3	120	Treatment	17	20.94	2.08	17	24
			130	Treatment	35	12.66	4.04	0	20
CA-1	Within	#4	207	Treatment	45	19.87	3.31	10	25
			210	Control	81	19.23	3.19	8	25
			218	Treatment	48	20.17	2.94	12	26
			225	Treatment	28	24.29	2.21	20	28
			236	Treatment	65	19.45	3.78	6	25
		#5	237	Control	24	20.00	2.48	16	25
			208	Control	52	18.94	3.54	10	24
			216	Treatment	25	24.64	1.87	21	27
			230	Treatment	75	19.27	4.21	0	26
			232	Control	60	19.10	3.69	7	25
			233	Control	48	16.46	6.40	0	23
		#6	201	Control	48	19.85	2.96	12	24
			202	Treatment	118	23.52	2.88	14	28
			203	Treatment	39	19.05	4.21	9	25
			205	Control	48	20.73	3.43	12	27
			209	Treatment	51	21.25	2.15	15	25
CA-2	Within	#7	305	Treatment	56	17.20	4.21	8	25
			306	Treatment	83	19.64	4.02	6	27
			309	Control	85	19.08	3.93	10	25
			310	Control	44	18.00	3.85	9	25
			316	Treatment	53	18.17	3.93	8	25
		#8	303	Control	89	18.10	4.38	5	25
			312	Treatment	59	18.46	3.84	5	25

District	Design	School	Teacher ID	Treatment	*N*	Pretest Mean	*SD*	Min	Max
		#9	406	Control	55	14.36	5.15	4	25
		#10	401	Treatment	20	16.75	5.87	6	25
			402	Treatment	18	20.61	2.81	16	26
CA-3	Between	#11	405	Control	3	17.67	3.06	15	21
			408	Treatment	9	12.44	3.47	6	16
		#12	409	Treatment	17	15.71	3.41	10	20
			410	Treatment	20	20.05	2.84	13	24
		#13	802	Control	93	18.59	4.09	8	26
			830	Control	24	22.04	2.20	17	26
			812	Control	45	19.71	3.52	11	25
			813	Control	52	18.23	3.09	11	24
		#14	814	Control	23	19.22	4.09	7	24
			815	Control	48	18.73	3.18	8	24
			868	Control	38	15.24	2.94	9	21
			831	Control	30	18.63	3.34	11	24
		#15	832	Control	41	19.98	2.77	12	24
			833	Control	41	23.10	2.64	13	27
		#16	839	Control	26	17.88	3.56	10	26
			806	Treatment	82	18.57	3.66	10	26
			807	Treatment	58	17.00	3.52	8	23
			808	Treatment	11	6.18	3.92	0	12
CA-4	Between	#17	809	Treatment	31	17.81	3.79	8	23
			810	Treatment	30	19.43	2.82	13	26
			821	Treatment	29	18.79	4.09	9	26
			834	Treatment	92	19.07	4.00	8	26
			818	Treatment	49	19.76	2.97	13	25
		#18	821	Treatment	31	19.35	3.13	13	26
			866	Treatment	28	20.04	2.87	13	25
			835	Treatment	24	18.71	3.63	7	26
		#19	836	Treatment	28	18.07	2.51	12	23
			837	Treatment	22	13.14	3.30	5	20
		#20	816	Control	33	19.52	4.64	0	26
			825	Treatment	43	18.16	4.90	6	27
		#21	826	Treatment	56	18.45	4.55	6	25
			828	Treatment	41	18.07	3.47	7	26
			846	Treatment	50	19.00	4.19	10	26
		#22	902	Treatment	19	21.63	2.71	17	25
CA-5	Between	#23	904	Control	87	21.03	3.97	10	28
		#24	901	Treatment	19	16.79	5.74	7	25

Table P14

Grade 6 Descriptive Statistics of Interim Test Scores by Teacher

District	Design	School	Teacher ID	Treatment	*N*	Interim test Mean	*SD*	Min	Max
AZ-1	Between	#1	118	Control	92	14.63	4.69	5	25
			128	Control	55	12.02	3.46	3	19
			129	Control	26	14.69	3.46	6	21
			138	Control	5	3.80	1.30	2	5
		#2	124	Treatment	106	11.08	4.84	1	22
			125	Treatment	55	12.13	4.73	2	21
			132	Treatment	22	15.73	3.25	9	22
		#3	120	Treatment	18	15.11	3.20	10	23
			130	Treatment	30	9.37	3.65	2	16
CA-1	Within	#4	207	Treatment	43	12.42	4.74	3	22
			210	Control	85	11.00	3.69	2	20
			218	Treatment	50	13.76	4.88	5	23
			225	Treatment	29	19.79	2.81	13	24
			236	Treatment	65	12.57	5.11	0	20
			237	Control	27	13.15	3.79	3	20
		#5	208	Control	52	10.77	3.65	2	17
			216	Treatment	28	21.32	2.02	18	25
			230	Treatment	78	12.73	4.76	3	22
			232	Control	58	10.31	3.80	3	21
			233	Control	45	11.40	3.18	4	17
		#6	201	Control	49	11.12	3.87	3	19
			202	Treatment	114	18.84	3.91	5	25
			203	Treatment	37	12.27	3.99	5	20
			205	Control	45	12.69	4.93	3	21
			209	Treatment	44	17.05	4.04	8	25
CA-2	Within	#7	305	Treatment	66	11.15	4.82	0	22
			306	Treatment	90	10.93	4.47	0	20
			309	Control	91	13.81	5.12	3	26
			310	Control	36	12.11	4.22	3	23
			316	Treatment	66	9.95	4.26	2	21
		#8	303	Control	81	11.94	3.41	2	19
			312	Treatment	58	8.45	4.60	1	18

District	Design	School	Teacher ID	Treatment	*N*	Interim test Mean	*SD*	Min	Max
CA-3	Between	#9	406	Control	55	11.96	4.66	2	22
		#10	401	Treatment	22	12.64	5.62	2	21
			402	Treatment	29	14.86	4.78	6	24
		#11	408	Treatment	19	7.37	3.90	3	15
			409	Treatment	25	11.80	3.71	5	18
CA-4	Between	#12	802	Control	99	10.48	4.59	1	22
			830	Control	23	15.17	3.54	7	23
			860	Control	1	0.00	.	0	0
		#13	812	Control	49	10.41	4.42	1	21
			813	Control	49	7.67	3.40	2	15
			814	Control	45	8.31	3.81	3	17
			815	Control	52	10.23	4.32	2	21
			868	Control	32	6.22	2.39	3	12
		#14	839	Control	24	10.33	4.32	4	20
		#15	806	Treatment	81	11.43	4.91	1	21
			807	Treatment	52	7.52	4.76	0	19
			808	Treatment	8	7.75	2.71	4	11
			809	Treatment	29	12.48	5.14	5	22
			810	Treatment	32	13.34	5.48	3	23
			821	Treatment	30	10.80	4.05	3	17
			834	Treatment	86	12.60	5.50	1	24
		#16	818	Treatment	59	14.05	4.62	2	23
			821	Treatment	30	11.50	5.12	4	23
			866	Treatment	29	12.34	5.53	2	22
		#17	835	Treatment	27	7.74	3.55	1	15
			837	Treatment	23	6.17	2.29	2	12
		#18	825	Treatment	50	10.82	5.85	2	25
			826	Treatment	58	11.38	5.13	3	22
			828	Treatment	39	13.82	4.73	2	23
			846	Treatment	50	13.94	5.02	4	22
			847	Treatment	1	4.00	.	4	4
CA-5	Between	#19	902	Treatment	25	16.72	4.46	3	22
		#20	904	Control	84	15.14	5.57	4	25
		#21	901	Treatment	20	11.90	6.12	3	23

Table P15

Grade 6 Descriptive Statistics of Posttest Scores by Teacher

District	Design	School	Teacher ID	Treatment	*N*	Posttest Mean	*SD*	Min	Max
AZ-1	Between	#1	118	Control	88	19.30	4.59	6	27
			128	Control	52	14.00	3.62	5	20
			129	Control	27	15.26	4.43	8	24
		#2	124	Treatment	100	13.66	4.52	3	27
			125	Treatment	51	15.37	5.02	6	26
			132	Treatment	20	17.95	3.39	10	24
		#3	120	Treatment	17	21.82	4.08	15	31
			130	Treatment	35	14.11	4.91	5	24
CA-1	Within	#4	207	Treatment	45	12.58	4.69	5	27
			210	Control	81	13.32	4.35	3	25
			218	Treatment	48	14.42	4.66	4	22
			225	Treatment	28	25.04	3.45	19	31
			236	Treatment	65	14.31	4.89	2	25
			237	Control	24	16.04	3.47	11	25
		#5	208	Control	52	11.62	4.39	2	21
			216	Treatment	25	23.96	3.59	18	30
			230	Treatment	75	13.84	5.51	3	26
			232	Control	60	12.88	4.98	3	24
			233	Control	48	13.42	3.64	5	20
		#6	201	Control	48	13.73	4.28	5	23
			202	Treatment	118	22.25	5.78	7	32
			203	Treatment	39	13.51	4.91	4	26
			205	Control	48	15.94	5.07	4	26
			209	Treatment	51	17.55	4.30	7	24
CA-2	Within	#7	305	Treatment	56	14.96	6.06	2	30
			306	Treatment	83	15.96	5.68	1	28
			309	Control	85	16.47	5.02	5	27
			310	Control	44	16.23	4.11	6	26
			316	Treatment	53	13.36	5.06	5	32
		#8	303	Control	89	14.25	4.32	4	23
			312	Treatment	59	12.10	5.52	1	26
CA-3	Between	#9	406	Control	55	14.82	6.00	3	28
		#10	401	Treatment	20	15.85	7.11	5	27
			402	Treatment	18	17.78	6.13	9	33
		#11	405	Control	3	17.00	1.00	16	18
		#12	408	Treatment	9	10.67	4.00	3	16
			409	Treatment	17	16.65	2.71	11	21

District	Design	School	Teacher ID	Treatment	*N*	Posttest Mean	*SD*	Min	Max
			410	Treatment	20	18.90	3.73	12	27
CA-4	Between	#13	802	Control	93	12.92	4.65	3	24
			830	Control	24	18.92	3.34	12	25
		#14	812	Control	45	12.64	5.12	1	24
			813	Control	52	10.90	3.18	4	17
			814	Control	23	10.87	3.53	4	16
			815	Control	48	12.73	3.76	5	20
			868	Control	38	9.37	2.32	4	15
		#15	831	Control	30	14.60	3.91	9	24
			832	Control	41	12.61	4.36	5	21
			833	Control	41	19.49	4.18	11	27
		#16	839	Control	26	12.46	4.64	5	22
		#17	806	Treatment	82	13.39	5.13	3	24
			807	Treatment	58	10.50	4.58	3	20
			808	Treatment	11	3.73	2.00	0	7
			809	Treatment	31	14.97	4.71	4	25
			810	Treatment	30	14.73	5.51	4	26
			821	Treatment	29	14.55	4.61	5	23
			834	Treatment	92	14.12	5.43	0	27
		#18	818	Treatment	49	17.06	4.16	9	29
			821	Treatment	31	15.16	4.92	7	24
			866	Treatment	28	14.29	4.84	7	28
		#19	835	Treatment	24	15.08	3.02	10	20
			836	Treatment	28	10.07	4.03	1	19
			837	Treatment	22	7.82	2.59	4	13
		#20	816	Control	33	16.03	4.01	9	24
		#21	825	Treatment	43	12.74	5.24	3	24
			826	Treatment	56	13.27	5.89	4	30
			828	Treatment	41	17.00	5.92	1	29
			846	Treatment	50	13.50	5.50	5	29
CA-5	Between	#22	902	Treatment	19	19.16	3.72	11	24
		#23	904	Control	87	17.91	6.04	4	29
		#24	901	Treatment	19	15.37	5.92	3	28

Appendix Q:
Complete Statistical Model Used

Statistical Model: 2-Level Hierarchical Model with School Fixed Effects

The level-1 (between-student; within-teacher) model specifies the relationship between student score on the transfer measure and his or her pretest score as a covariate. The transfer measure total score, Y_{ij} is outcome for student *i* in teacher *j*. The pretest score for student *i* in teacher *j* ($Pretest_{ij}$) is centered around its mean. By virtue of this centering method, β_{0j} is unadjusted transfer measure mean for teacher *j*, and β_{1j} is pretest-outcome slope for teacher *j*.

Level-1 (between-student; within-teacher) model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Pretest_{ij} - Pretest_{.j}) + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \qquad (1)$$

The level-2 (between-teacher) model includes treatment indicator variable (control group teacher = 0 and POWERSOURCE© teacher =1), design indicator variable (within-school design = 0 and between school design =1), and pretest mean. Note that we also include school flag variables in order to estimate school specific effects, which takes into account intra-class correlation in school level.

Level-2 (beween-teacher) model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Trt_j + \gamma_{02}Design_j + \gamma_{03}Trt_j \times Design_j + \gamma_{04}(Pretest_{.j} - Pretest_{..}) +$$
$$\gamma_{0\_k}S\_{k} + \ldots + \gamma_{0\_kn\text{-}3}S\_{kn\text{-}3} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00}) \qquad (2a)$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Trt_j + u_{1j} \qquad u_{1j} \sim N(0, \tau_{11}) \qquad (2b)$$

γ_{00} represent the expected mean for control group in W-S design holding other variable constant including school specific effects. γ_{01} and γ_{02} are main effects of treatment and design, respectively. γ_{03} captures the interaction effect between treatment condition and design. If this coefficient is statistically significant, it indicates that treatment effect is different depending upon designs. γ_{0_k} through $\gamma_{0\_kn\text{-}3}$ are school specific fixed effects. Note that there are *k*-3 school fixed effects, where *k* is total number of schools, because there are four baseline groups: treatment and control in B-S design, and treatment and control in W-S design. γ_{11} captures the difference in pretest-outcome slope between treatment and control group.

Appendix R:

Estimates for Fixed Effects and the Variance

Table R1

HM Result: Transder Measure Total Score (Grade 7)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	10.36	1.22	30	8.50	<.0001
Treatment, γ_{01}	1.19	0.88	30	1.34	0.190
Design, γ_{02}	1.41	1.61	30	0.87	0.390
Treatment*Design, γ_{03}	-2.51	2.61	30	-0.96	0.345
Pretest Mean, γ_{04}	1.16	0.13	30	8.84	<.0001
s03, γ_{0_03}	0.71	1.76	30	0.40	0.689
s04, γ_{0_04}	-0.38	1.40	30	-0.27	0.790
s06, γ_{0_06}	-0.13	2.99	30	-0.04	0.966
s07, γ_{0_07}	1.82	2.99	30	0.61	0.546
s08, γ_{0_08}	-1.58	1.80	30	-0.88	0.388
s09, γ_{0_09}	-1.52	1.94	30	-0.78	0.440
s10, γ_{0_10}	-0.92	1.45	30	-0.64	0.528
s11, γ_{0_11}	0.67	2.40	30	0.28	0.783
s12, γ_{0_12}	-0.61	2.49	30	-0.25	0.807
s13, γ_{0_13}	1.92	1.76	30	1.09	0.285
s14, γ_{0_14}	2.24	1.53	30	1.46	0.154
s15, γ_{0_15}	-2.81	1.47	30	-1.91	0.066
s16, γ_{0_16}	2.09	3.00	30	0.70	0.491
s17, γ_{0_17}	7.17	3.09	30	2.32	0.027
s18, γ_{0_18}	-6.21	1.83	30	-3.39	0.002
s19, γ_{0_19}	1.19	1.81	30	0.66	0.516
s20, γ_{0_20}	-1.09	1.86	30	-0.59	0.561
s21, γ_{0_21}	-2.45	2.36	30	-1.04	0.308
Model for pretest slope					
Intercept, γ_{10}	0.56	0.06	2538	8.88	<.0001
Treatment, γ_{11}	0.17	0.09	2538	1.96	0.051

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	3.70	1.09	3.40	0.000
Pretest slope, u_{1j}	0.06	0.02	3.17	0.001
Level-1 error for B-S, e_{ij_1}	14.54	0.59	24.49	<.0001
Level-1 error for W-S, e_{ij_2}	16.14	0.63	25.44	<.0001

Appendix S:
Subdomain HLM Analysis Results for Grade 7

Table S1

HM result: Transfer measure RNE Subdomain (Grade 7)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	2.83	0.28	30	9.97	<.0001
Treatment, γ_{01}	-0.01	0.20	30	-0.05	0.961
Design, γ_{02}	-0.31	0.37	30	-0.84	0.410
Treatment*Design, γ_{03}	-0.34	0.64	30	-0.53	0.600
Pretest Mean, γ_{04}	0.44	0.05	30	8.52	<.0001
s03, γ_{0_03}	-0.05	0.41	30	-0.13	0.897
s04, γ_{0_04}	-0.13	0.33	30	-0.4	0.690
s06, γ_{0_06}	0.04	0.72	30	0.06	0.955
s07, γ_{0_07}	0.70	0.72	30	0.98	0.335
s08, γ_{0_08}	-0.79	0.43	30	-1.86	0.072
s09, γ_{0_09}	-0.32	0.46	30	-0.69	0.493
s10, γ_{0_10}	-0.63	0.34	30	-1.87	0.072
s11, γ_{0_11}	0.24	0.59	30	0.4	0.688
s12, γ_{0_12}	0.15	0.61	30	0.24	0.810
s13, γ_{0_13}	0.46	0.40	30	1.13	0.266
s14, γ_{0_14}	0.31	0.35	30	0.89	0.381
s15, γ_{0_15}	-0.87	0.35	30	-2.51	0.018
s16, γ_{0_16}	0.14	0.71	30	0.19	0.850
s17, γ_{0_17}	1.31	0.76	30	1.73	0.093
s18, γ_{0_18}	-1.37	0.43	30	-3.14	0.004
s19, γ_{0_19}	-0.29	0.42	30	-0.7	0.492
s20, γ_{0_20}	-0.74	0.45	30	-1.65	0.110
s21, γ_{0_21}	-0.30	0.56	30	-0.54	0.594
Model for pretest slope					
Intercept, γ_{10}	0.18	0.02	2538	7.47	<.0001
Treatment, γ_{11}	0.06	0.03	2538	1.88	0.061

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	0.18	0.06	3.08	0.0010
Pretest slope, u_{1j}	0.01	0.00	2.55	0.0050
Level-1 error for B-S, e_{ij_1}	1.47	0.06	24.55	<.0001
Level-1 error for W-S, e_{ij_2}	1.56	0.06	25.43	<.0001

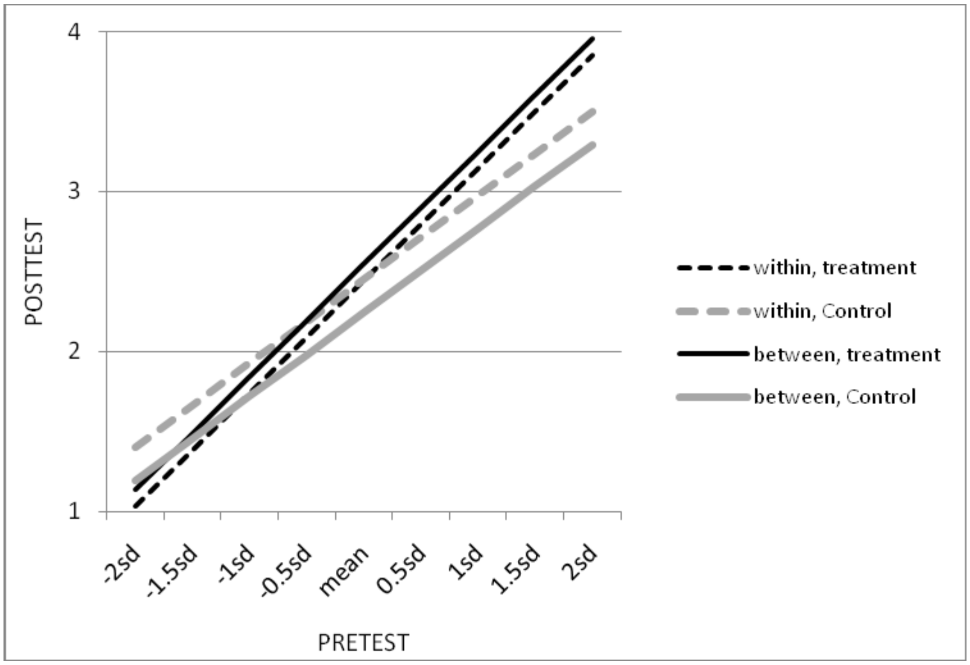


Figure S1. HM result for grade 7 (RNE transfer measure subscore): Fitted relationships between pretest RNE subscore and posttest RNE subscore by design and treatment condition.

Table S2

HM Result: Transfer Measure PA Subdomain (Grade 7)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	1.39	0.14	30	9.67	<.0001
Treatment, γ_{01}	0.14	0.10	30	1.34	0.190
Design, γ_{02}	0.08	0.19	30	0.44	0.662
Treatment*Design, γ_{03}	0.07	0.33	30	0.22	0.824
Pretest Mean, γ_{04}	0.33	0.05	30	6.94	<.0001
s03, γ_{0_03}	0.43	0.21	30	2.10	0.045
s04, γ_{0_04}	0.00	0.17	30	-0.03	0.978
s06, γ_{0_06}	0.02	0.38	30	0.04	0.965
s07, γ_{0_07}	0.14	0.38	30	0.38	0.706
s08, γ_{0_08}	0.06	0.22	30	0.27	0.790
s09, γ_{0_09}	-0.06	0.24	30	-0.25	0.802
s10, γ_{0_10}	0.05	0.17	30	0.29	0.774
s11, γ_{0_11}	-0.09	0.31	30	-0.30	0.764

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
s12, γ_{0_12}	-0.11	0.32	30	-0.36	0.724
s13, γ_{0_13}	0.22	0.20	30	1.09	0.286
s14, γ_{0_14}	0.33	0.18	30	1.81	0.081
s15, γ_{0_15}	0.02	0.16	30	0.15	0.880
s16, γ_{0_16}	0.09	0.37	30	0.26	0.800
s17, γ_{0_17}	1.42	0.41	30	3.49	0.002
s18, γ_{0_18}	-0.41	0.21	30	-1.96	0.060
s19, γ_{0_19}	0.33	0.20	30	1.59	0.122
s20, γ_{0_20}	0.14	0.22	30	0.65	0.520
s21, γ_{0_21}	-0.08	0.29	30	-0.26	0.795
Model for pretest slope					
Intercept, γ_{10}	0.08	0.02	2538	5.14	<.0001
Treatment, γ_{11}	0.02	0.02	2538	0.73	0.467

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	0.04	0.01	2.81	0.0030
Pretest slope,u_{1j}	0.00	0.00	0.26	0.3960
Level-1 error for B-S, e_{ij_1}	0.56	0.02	24.73	<.0001
Level-1 error for W-S, e_{ij_2}	0.65	0.03	25.38	<.0001

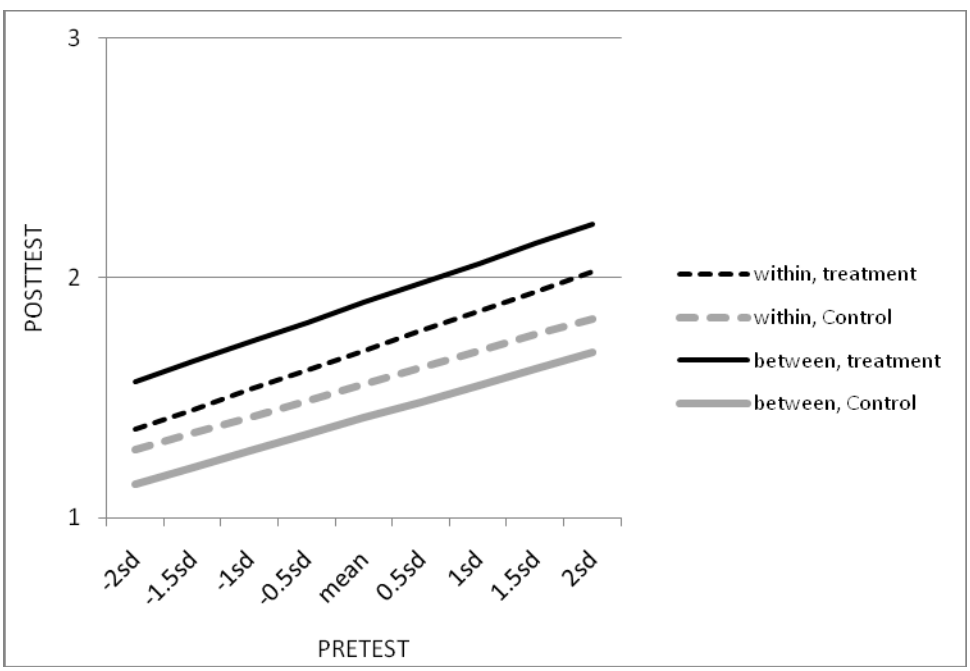


Figure S2. HM result for grade 7(PA transfer measure subscore): Fitted relationships between pretest PA subscore and posttest PA subscore by design and treatment condition.

Table S3

HM Result: Transfer Measure SE Subdomain

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	3.76	0.78	30	4.80	<.0001
Treatment, γ_{01}	0.95	0.61	30	1.56	0.128
Design, γ_{02}	2.35	1.04	30	2.26	0.031
Treatment*Design, γ_{03}	-2.21	1.79	30	-1.24	0.225
Pretest Mean, γ_{04}	3.75	0.68	30	5.51	<.0001
s03, γ_{0_03}	0.68	1.12	30	0.61	0.549
s04, γ_{0_04}	-0.23	0.91	30	-0.25	0.803
s06, γ_{0_06}	2.11	2.06	30	1.02	0.314
s07, γ_{0_07}	1.35	2.04	30	0.66	0.513
s08, γ_{0_08}	0.74	1.20	30	0.62	0.542
s09, γ_{0_09}	-2.80	1.27	30	-2.20	0.035
s10, γ_{0_10}	0.94	0.91	30	1.04	0.308
s11, γ_{0_11}	0.46	1.65	30	0.28	0.783
s12, γ_{0_12}	0.26	1.71	30	0.15	0.880
s13, γ_{0_13}	0.24	1.13	30	0.21	0.834
s14, γ_{0_14}	1.97	0.98	30	2.01	0.054
s15, γ_{0_15}	0.06	0.88	30	0.07	0.947
s16, γ_{0_16}	0.53	2.00	30	0.27	0.792
s17, γ_{0_17}	4.74	2.14	30	2.22	0.034
s18, γ_{0_18}	-2.65	1.16	30	-2.29	0.029
s19, γ_{0_19}	1.86	1.09	30	1.70	0.099
s20, γ_{0_20}	2.51	1.15	30	2.19	0.037
s21, γ_{0_21}	-2.65	1.57	30	-1.68	0.103
Model for pretest slope					
Intercept, γ_{10}	0.385	0.1368	2538	2.82	0.0049
Treatment, γ_{11}	0.2523	0.1886	2538	1.34	0.1811

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	2.03	0.57	3.54	0.000
Pretest slope, u_{1j}	0.21	0.10	2.14	0.016
Level-1 error for B-S, e_{ij_1}	6.26	0.26	24.55	<.0001
Level-1 error for W-S, e_{ij_2}	6.75	0.27	25.38	<.0001

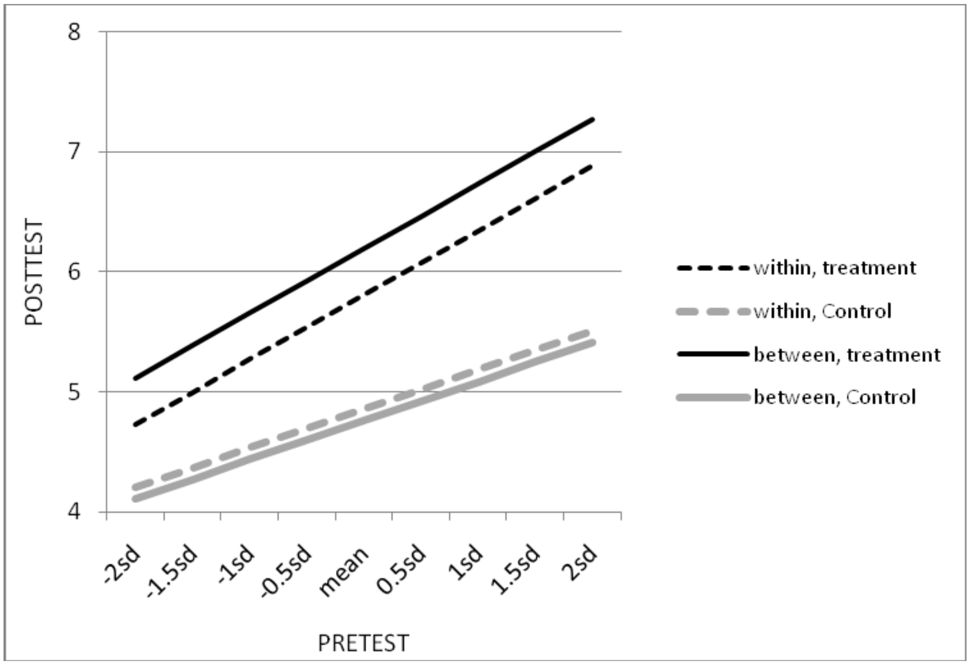


Figure S3. HM result for grade 7 (SE transfer measure subscore): Fitted relationships between pretest SE subscore and posttest SE subscore by design and treatment condition.

Appendix T:

Fixed Effects and the Variance Components in the Model, Grade 6

Table T1

HM Result: Transfer Measure Total Score (Grade 6)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	16.09	0.64	44	25.04	<.0001
Treatment, γ_{01}	-0.44	0.55	44	-0.81	0.425
Design, γ_{02}	-1.50	1.37	44	-1.09	0.280
Treatment*Design, γ_{03}	0.74	1.62	44	0.45	0.651
Interim measure Mean, γ_{04}	1.14	0.07	44	16.63	<.0001
s03, γ_{0_03}	0.39	1.48	44	0.26	0.793
s05, γ_{0_05}	0.30	1.35	44	0.23	0.823
s08, γ_{0_08}	-1.92	0.76	44	-2.53	0.015
s09, γ_{0_09}	-0.26	1.80	44	-0.15	0.884
s10, γ_{0_10}	-1.03	1.05	3014	-0.98	0.328
s11, γ_{0_11}	-0.43	1.14	3014	-0.38	0.703
s12, γ_{0_12}	-2.19	0.80	44	-2.74	0.009
s13, γ_{0_13}	-0.10	1.38	44	-0.07	0.945
s14, γ_{0_14}	-1.14	1.02	44	-1.12	0.270
s15, γ_{0_15}	0.00	1.40	44	0.00	1.000
s17, γ_{0_17}	-1.73	0.80	44	-2.16	0.036
s18, γ_{0_18}	0.89	1.36	44	0.66	0.515
s19, γ_{0_19}	0.13	1.18	44	0.11	0.910
s20, γ_{0_20}	2.92	1.33	44	2.19	0.034
s21, γ_{0_21}	-1.73	1.63	44	-1.06	0.293
s22, γ_{0_22}	0.03	1.70	44	0.02	0.985
s23, γ_{0_23}	0.30	1.66	44	0.18	0.858
s24, γ_{0_24}	-1.39	1.12	44	-1.25	0.219
Model for interim measure slope					
Intercept, γ_{10}	0.54	0.05	3014	10.51	<.0001
Treatment, γ_{11}	0.14	0.06	3014	2.21	0.027

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	1.25	0.37	3.39	0.0003
Interim measure slope, u_{1j}	0.04	0.01	3.89	<.0001
Level-1 error for B-S, e_{ij_1}	12.49	0.43	28.89	<.0001
Level-1 error for W-S, e_{ij_2}	14.85	0.58	25.43	<.0001

Appendix U:
Subdomains as an Outcome Variable, Grade 6

Table U1

HM Result: RNE Transfer Measure Subscore (Grade 6)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	5.04	0.25	44	20.48	<.0001
Treatment, γ_{01}	0.46	0.21	44	2.21	0.033
Design, γ_{02}	-0.72	0.53	44	-1.36	0.180
Treatment*Design, γ_{03}	0.46	0.64	44	0.72	0.477
Interim measure Mean, γ_{04}	0.67	0.05	44	13.62	<.0001
s03, γ_{0_03}	0.57	0.58	44	0.98	0.332
s05, γ_{0_05}	0.62	0.52	44	1.18	0.245
s08, γ_{0_08}	-0.08	0.29	44	-0.27	0.787
s09, γ_{0_09}	0.47	0.72	44	0.65	0.519
s10, γ_{0_10}	-0.23	0.43	3014	-0.54	0.590
s11, γ_{0_11}	-0.33	0.46	3014	-0.72	0.472
s12, γ_{0_12}	-0.40	0.31	44	-1.29	0.203
s13, γ_{0_13}	0.24	0.54	44	0.45	0.654
s14, γ_{0_14}	-0.55	0.39	44	-1.40	0.170
s15, γ_{0_15}	0.22	0.57	44	0.38	0.703
s17, γ_{0_17}	-0.11	0.31	44	-0.37	0.710
s18, γ_{0_18}	0.30	0.55	44	0.54	0.591
s19, γ_{0_19}	-0.35	0.47	44	-0.74	0.461
s20, γ_{0_20}	0.66	0.54	44	1.22	0.231
s21, γ_{0_21}	-0.86	0.66	44	-1.29	0.202
s22, γ_{0_22}	0.22	0.65	44	0.33	0.742
s23, γ_{0_23}	-0.16	0.68	44	-0.23	0.818
s24, γ_{0_24}	-0.35	0.45	44	-0.78	0.441
Model for interim measure slope					
Intercept, γ_{10}	0.29	0.03	3014	9.61	<.0001
Treatment, γ_{11}	0.08	0.04	3014	2.05	0.0401

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	0.17	0.06	2.84	0.0020
Interim measure slope,u_{1j}	0.01	0.00	2.85	0.0020
Level-1 error for B-S, e_{ij_1}	3.06	0.11	28.91	<.0001
Level-1 error for W-S, e_{ij_2}	3.57	0.14	25.39	<.0001

Appendix V:
Interim Measure Estimates of Fixed Effect and the Variance Components

Table V1

HM Result: Interim Measure Total Score (Grade 6)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	11.04	0.91	47	12.15	<.0001
Treatment, γ_{01}	0.96	0.72	2966	1.32	0.185
Design, γ_{02}	4.26	2.04	47	2.08	0.043
Treatment*Design, γ_{03}	-2.94	2.34	2966	-1.26	0.209
Pretest Mean, γ_{04}	0.81	0.10	47	7.82	<.0001
s03, γ_{0_03}	-3.57	2.23	2966	-1.60	0.110
s05, γ_{0_05}	-5.95	2.01	47	-2.96	0.005
s08, γ_{0_08}	0.52	1.10	47	0.47	0.641
s09, γ_{0_09}	-4.41	2.63	47	-1.68	0.100
s10, γ_{0_10}	-1.14	1.44	2966	-0.79	0.427
s11, γ_{0_11}	-0.97	1.53	2966	-0.64	0.525
s12, γ_{0_12}	0.86	1.15	47	0.75	0.459
s13, γ_{0_13}	-1.81	2.14	47	-0.85	0.402
s14, γ_{0_14}	-0.60	1.49	47	-0.40	0.688
s15, γ_{0_15}	-4.07	1.78	2966	-2.29	0.022
s17, γ_{0_17}	1.00	1.16	47	0.86	0.394
s18, γ_{0_18}	1.10	1.54	2966	0.71	0.475
s19, γ_{0_19}	1.64	1.68	47	0.98	0.333
s20, γ_{0_20}	0.44	1.86	47	0.23	0.816
s21, γ_{0_21}	2.29	2.32	47	0.98	0.330
s22, γ_{0_22}	-1.92	2.62	47	-0.73	0.467
s23, γ_{0_23}	0.27	2.32	47	0.12	0.909
s24, γ_{0_24}	-0.69	1.54	47	-0.45	0.655
Model for pretest slope					
Intercept, γ_{10}	0.55	0.06	2966	8.82	<.0001
Treatment, γ_{11}	0.12	0.08	2966	1.47	0.141

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	2.99	0.75	3.97	<.0001
Pretest slope, u_{1j}	-0.04	0.12	-0.33	0.7420
Level-1 error for B-S, e_{ij_1}	13.90	0.49	28.40	<.0001
Level-1 error for W-S, e_{ij_2}	12.22	0.48	25.52	<.0001

Appendix W:

Estimates for Fixed Effects and the Variance Components in the Model, Grade 6 Transfer Measure

Table W1

HM Result: Transfer Measure Total Score (Grade 6)

Fixed effects	Coefficient	*SE*	*df*	*t*-value	*p*-value
Model for class mean					
Intercept, γ_{00}	15.57	1.01	45	15.47	<.0001
Treatment, γ_{01}	0.52	0.86	45	0.61	0.545
Design, γ_{02}	3.55	2.19	45	1.62	0.112
Treatment*Design, γ_{03}	-2.90	2.58	45	-1.13	0.266
Pretest Mean, γ_{04}	1.09	0.11	45	9.48	<.0001
s03, γ_{0_03}	-4.94	2.46	45	-2.01	0.050
s05, γ_{0_05}	-7.54	2.16	45	-3.50	0.001
s06, γ_{0_06}	-1.20	3.56	45	-0.34	0.738
s07, γ_{0_07}	-5.74	2.34	45	-2.45	0.018
s08, γ_{0_08}	-2.14	1.20	45	-1.78	0.082
s09, γ_{0_09}	-5.72	2.81	45	-2.03	0.048
s10, γ_{0_10}	-2.41	1.63	3107	-1.48	0.139
s11, γ_{0_11}	-2.24	1.72	3107	-1.30	0.192
s12, γ_{0_12}	-2.12	1.25	45	-1.70	0.097
s13, γ_{0_13}	-2.98	2.28	45	-1.31	0.198
s14, γ_{0_14}	-2.67	1.60	45	-1.67	0.103
s15, γ_{0_15}	-3.92	1.89	45	-2.07	0.044
s16, γ_{0_16}	-4.12	2.82	45	-1.46	0.151
s17, γ_{0_17}	-1.28	1.26	45	-1.02	0.314
s18, γ_{0_18}	1.40	1.94	45	0.72	0.476
s19, γ_{0_19}	1.45	1.88	45	0.77	0.443
s20, γ_{0_20}	3.09	2.07	45	1.49	0.142
s21, γ_{0_21}	-1.06	2.57	45	-0.41	0.683
s22, γ_{0_22}	-2.78	2.80	45	-0.99	0.326
s23, γ_{0_23}	0.91	2.54	45	0.36	0.723
s24, γ_{0_24}	-2.09	1.76	45	-1.19	0.241
Model for pretest slope					
Intercept, γ_{10}	0.52	0.06	3107	8.27	<.0001
Treatment, γ_{11}	0.13	0.08	3107	1.53	0.126

Random Effects	Variance component	*SE*	*z*-value	*p*-value
Class mean, u_{0j}	3.62	0.89	4.07	<.0001
Pretest slope, u_{1j}	0.08	0.02	4.02	<.0001
Level-1 error for B-S, e_{ij_1}	15.72	0.53	29.77	<.0001
Level-1 error for W-S, e_{ij_2}	17.56	0.69	25.29	<.0001

Appendix X:
Teacher Evaluation of Student Work

Descriptive Results for Subquestions of Task 3

Differences in pre and post scores for the Teacher Evaluation of Student Work were evaluated with the Kruskal-Wallis test. This test does not allow for a comparison of means, but does allow researchers to compare pre and post scores when the dependent variable is ordinal.

Descriptive results for each subquestion (i.e., a, b, and c) of Task 3 are found in the tables below:

		PRE_3a					
Group		0	1	2	3	4	Total
Treatment	Freq.	5	6	19	19	3	52
	Row %	9.62	11.54	36.54	36.54	5.77	
Control	Freq.	4	3	11	8	1	27
	Row %	14.81	11.11	40.74	29.63	3.7	
							79

		POST_3a					
Group		0	1	2	3	4	Total
Treatment	Freq.	2	5	34	26	3	70
	Row %	2.86	7.14	48.57	37.14	4.29	
Control	Freq.	2	2	24	17	2	47
	Row %	4.26	4.26	51.06	36.17	4.26	
							117

Group		PRE_3b 0	1	2	3	4	Total
Treatment	Freq.	6	31	9	5	1	52
	Row %	11.54	59.62	17.31	9.62	1.92	
Control	Freq.	4	13	5	4	1	27
	Row %	14.81	48.15	18.52	14.81	3.7	
							79

Group		POST_3b 0	1	2	3	4	Total
Treatment	Freq.	1	42	8	18	1	70
	Row %	1.43	60	11.43	25.71	1.43	
Control	Freq.	3	32	5	7	0	47
	Row %	6.38	68.09	10.64	14.89	0	
							117

Group		PRE_3c 0	1	2	3	4	Total
Treatment	Freq.	10	15	7	15	5	52
	Row %	19.23	28.85	13.46	28.85	9.62	
Control	Freq.	4	8	5	10	0	27
	Row %	14.81	29.63	18.52	37.04	0	
							79

Group		POST_3c 0	1	2	3	4	Total
Treatment	Freq.	1	18	17	30	4	70
	Row %	1.43	25.71	24.29	42.86	5.71	
Control	Freq.	3	11	5	24	4	47
	Row %	6.38	23.4	10.64	51.06	8.51	
							117

Appendix Y:

Grade 7 Interim Transfer Measure 2009/2010

Student ID # ________

PowerSource

Fill in the best answer for each question.

1. Which of these expressions is equivalent to $x \cdot x \cdot x \cdot x$ for all values of x?

   Ⓐ $x + 4$

   Ⓑ x^4

   Ⓒ $4x$

   Ⓓ $\frac{x}{4}$

2. Fill in the missing number: $6(a + 7) = (6 \cdot a) + (\square \cdot 7)$

3. $x \cdot (4 \cdot 5)$ has the same value as:

   Ⓐ $(x \cdot 4) \cdot 5$

   Ⓑ $4x + 5$

   Ⓒ $x \cdot 4 \cdot x \cdot 5$

   Ⓓ $x + 4 + 5$

4. $\frac{3}{11} \cdot 3$ has the same value as:

Ⓐ $\frac{6}{11}$

Ⓑ $\frac{1}{11}$

Ⓒ $\frac{9}{11}$

Ⓓ $\frac{3}{11}$

5. Which of the following fractions is equal to 1?

Ⓐ $\frac{0}{0}$

Ⓑ $\frac{\frac{5}{7}}{\frac{5}{7}}$

Ⓒ $\frac{(5 + 1)}{(5 - 1)}$

Ⓓ All of the above are equal to 1

6. If you add $\frac{4}{7} + \frac{5}{7}$, why is the answer not $\frac{9}{14}$? Explain in words:

7. Simplify:

$$\frac{4}{9} + \frac{5}{b} =$$

Note: $b \neq 0$

8. Keri has a rope that is $\frac{2}{3}$ feet long. She wants to divide the rope into pieces that are each $\frac{1}{6}$ feet long. How many pieces will she have?

- (A) 2
- (B) 4
- (C) 6
- (D) 9

9. Michael and Kevin took jellybeans from a jar. Michael took $\frac{3}{5}$ of the jellybeans and Kevin took $\frac{1}{10}$ of the jellybeans. What fraction of the jellybeans remained in the jar?

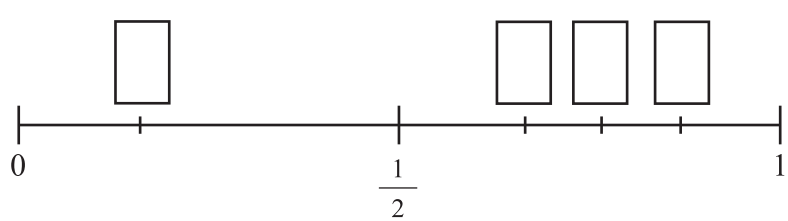
Ⓐ $\frac{3}{10}$

Ⓑ $\frac{7}{10}$

Ⓒ $\frac{1}{5}$

Ⓓ $\frac{4}{5}$

10. Here are four fractions: $\frac{2}{3}$, $\frac{4}{5}$, $\frac{1}{6}$, and $\frac{3}{4}$. Look at the number line below. Write each fraction in the correct box.



11. Which of the following shows the distributive property being used to rewrite the expression 5(7) + 5(4) ?

Ⓐ (5)(7)(4)

Ⓑ 7(5 + 4)

Ⓒ 7(5) + 4(5)

Ⓓ 5(7 + 4)

12. Which of these is equal to (436 • 795) + (436 • 5)?

Ⓐ 436 • 795 • 5

Ⓑ 436 • 795

Ⓒ 872 • 800

Ⓓ 436 • 800

13. Rhonda decided to decorate her t-shirt with sequins. She sewed 4 rows of sequins onto her shirt. Each row contains 8 sequins. She decided to add more sequins so she sewed 3 additional rows of 8 sequins on her shirt.

Write an expression to show how many sequins there are altogether.

14. The fraction $7\frac{3}{4}$ means $7 + \frac{3}{4}$, which can also be written as $(7 + \frac{3}{4})$.

Show how you would use the distributive property to multiply $7\frac{3}{4}$ by 5.

15. A student was asked to add the fractions, $\frac{5}{7} + \frac{2}{3}$.
The first step of her work is shown here:

Step 1: $\frac{5}{7} \cdot \frac{3}{3} + \frac{2}{3} \cdot \frac{7}{7}$

a) Is this work correct so far? ________________________________

b) Explain in words why you think it is correct or incorrect: ___________

____

____

____

____

____

PowerSource

16. A student solved the problem $3 \cdot 11\frac{2}{3}$ in the following way:

Step 1: $3 \cdot (11 + \frac{2}{3})$

Step 2: $(3 \cdot 11) + (3 \cdot \frac{2}{3})$

Step 3: $33 + 2$

Step 4: 35

a) Name the property the student used to get from Step 1 to Step 2: ______________

____

b) Explain why that property works: ____

____

____

____

____

____

Appendix Z:
Grade 7 Transfer Measure, Revised Version

PowerSource

Answer each question below.

1. I think of a number. I multiply this number by 8, then subtract 66. The result is twice the number that I was thinking of. Which equation represents this situation?

Ⓐ $8n - 66 = 2n$

Ⓑ $n + 8 - 66 = 2 + n$

Ⓒ $8n \cdot 66 = 2n$

Ⓓ $8 + n \cdot 66 = 2 + n$

2. The cost, *c,* of printing business cards consists of a fixed charge of 100 cents and a charge of 6 cents for each card printed. Which of these equations can be used to determine the cost of printing *n* cards?

Ⓐ $c = (100 + 6n)$

Ⓑ $c = (106 + n)$

Ⓒ $c = (6 + 100n)$

Ⓓ $c = (106n)$

Ⓔ $c = (600n)$

3. Which written expression could be represented by $37 - 3n = 5$?

- (A) The sum of 37 and 3 times a number is 5.
- (B) The product of *n* and 37 decreased by 3 is 5.
- (C) Three times a number decreased by 37 is 5.
- (D) Thirty-seven decreased by 3 times a number is 5.

4. Jim has $\frac{3}{4}$ of a yard of string which he wishes to divide into pieces, each $\frac{1}{8}$ of a yard long. How many pieces will he have?

- (A) 3
- (B) 4
- (C) 6
- (D) 8

5. Subtract: $\frac{3x}{7} - \frac{x}{7} =$

- (A) $\frac{3}{7}$
- (B) 3
- (C) $2x$
- (D) $\frac{x}{7}$
- (E) $\frac{2x}{7}$

6. Robin and Jim took cherries from a basket. Robin took $\frac{1}{3}$ of the cherries and Jim took $\frac{1}{6}$ of the cherries. What fraction of the cherries remained in the basket?

Ⓐ $\frac{1}{2}$

Ⓑ $\frac{1}{3}$

Ⓒ $\frac{1}{6}$

Ⓓ $\frac{1}{18}$

7. Four times a number is 48. Explain how you can find $\frac{1}{3}$ of the number.

8. There are 300 calories in 100 g of a certain food. How many calories are there in a 30 g portion of this food?

Ⓐ 90

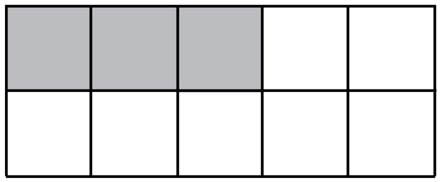
Ⓑ 100

Ⓒ 900

Ⓓ 30

Ⓔ 10

9. In the figure, how many MORE small squares need to be shaded so that $\frac{4}{5}$ of the small squares are shaded?



Ⓐ 5

Ⓑ 4

Ⓒ 3

Ⓓ 2

Ⓔ 1

10. Fifteen boxes each containing 8 radios can be repacked in 10 larger boxes each containing how many radios?

Ⓐ 8

Ⓑ 10

Ⓒ 12

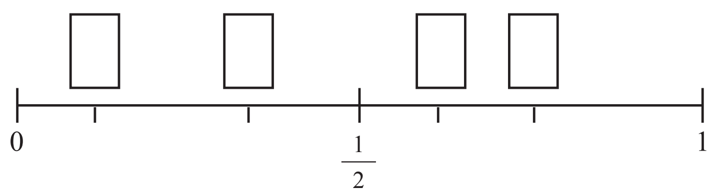
Ⓓ 80

Ⓔ 120

11. Here are four fractions: $\frac{3}{4}$, $\frac{1}{8}$, $\frac{1}{3}$ and $\frac{3}{5}$.

a) Explain how you can use the multiplicative identity property to help determine which fraction is the largest.

b) Look at the number line below. Write each fraction in the correct box.



12. A rectangular playground has a **perimeter** of 390 feet. The width of the playground is 75 feet.

A student wrote the following equation to solve the problem: $2(h + 75) = 390$
His first step is shown here:

Step 1: $2h + 150 = 390$

Explain how you would solve the equation in step 1 to find the height. Be sure to explain all your steps:

13. $\frac{3}{5} + \left(\frac{3}{10} \bullet \frac{4}{15}\right) =$

Ⓐ $\frac{3}{51}$

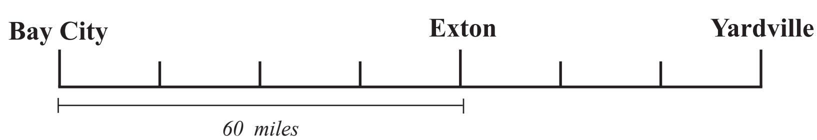
Ⓑ $\frac{1}{6}$

Ⓒ $\frac{6}{25}$

Ⓓ $\frac{11}{25}$

Ⓔ $\frac{17}{25}$

14.



On the road shown above, the distance from Bay City to Exton is 60 miles. What is the distance from Bay City to Yardville?

Ⓐ 45 miles

Ⓑ 120 miles

Ⓒ 90 miles

Ⓓ 105 miles

15. A painter had 25 L of paint. He used 2.5 L of paint every hour. He finished the job in 5.5 hours. How much paint did he have left?

- (A) 10.25 L
- (B) 11.25 L
- (C) 12.75 L
- (D) 13.75 L

16. Daniel had 31 baseball cards. He gave the cards to his friends. Six of his friends received 3 cards each. Seven of his friends received 1 card each. The rest received 2 cards each. How many of his friends received exactly 2 cards from Daniel? Explain how you found your answer.

17. Which of these is equal to (370 • 998) + (370 • 2)?

- (A) 370 • 1,000
- (B) 372 • 998
- (C) 740 • 998
- (D) 370 • 998 • 2

18. A garden has 14 rows. Each row has 20 plants. The gardener then plants *x* more rows with 20 plants in each row.

Use the distributive property to write an expression to show how many plants there are altogether.

19. John sold 60 magazines and Mark sold 80 magazines. The magazines were all sold for the same price. The total amount of money received for the magazines was $700.

a) Write an equation to find the cost of a magazine.

b) Solve the equation to find out how much each magazine cost.

c) How much money did **each** boy make?

20. Graham has twice as many books as Bob. Chan has six more books than Bob. If Bob has b books, which of the following represents the total number of books the **three** boys (Graham, Bob and Chan) have?

- (A) $3b + 6$
- (B) $3b + 8$
- (C) $4b + 6$
- (D) $5b + 6$
- (E) $8b + 2$

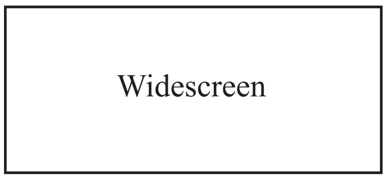
21. The fraction $2\frac{1}{4}$ means $2 + \frac{1}{4}$, which can also be written as $(2 + \frac{1}{4})$. Explain how you would use the distributive property to multiply $2\frac{1}{4}$ by 10.

22. A book publisher sent 140 copies of a book to a bookstore. The publisher packed the books in two types of boxes. On type of box held 8 copies of the book, and the other type of box held 12 copies of the book. The boxes were all full, and there were **equal** numbers of both types of boxes.

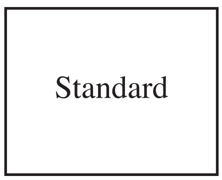
a) How many full boxes of 12 books were there?

b) What fraction of the books were packed in the smaller boxes?

23. The screens of widescreen and standard televisions look different. Widescreen television ratio of height to width is 9:16. Standard television ratio of height to width is 3:4. Keri starts to draw scale drawings of the televisions. For each, the height is 4.5 cm. What should the width of each scale drawing be?



The width of this scale drawing should be _____cm



The width of this scale drawing should be _____cm

Appendix AA:
Grade 8 Pretest

Student ID # ________

PowerSource Pretest

Fill in the best answer to each question.

1. The product of a number (x) and 15 is 37. Which equation shows this relationship?

Ⓐ $15x = 37$

Ⓑ $x + 15 = 37$

Ⓒ $x - 15 = 37$

Ⓓ $\frac{x}{15} = 37$

2. Which of the following equations shows the inverse property of multiplication?

Ⓐ $7 \cdot 7 = 49$

Ⓑ $7 \cdot 1 = 7$

Ⓒ $7 \cdot \frac{1}{7} = 1$

Ⓓ $7 \cdot 0 = 0$

3. Solve for x:

$2x + 3 = 10 + x$

Ⓐ $\frac{7}{3}$

Ⓑ $\frac{10}{4}$

Ⓒ 10

Ⓓ 7

4. What is the value of y if $-5y + 4 = -11$?

Ⓐ $y = -2$

Ⓑ $y = -3$

Ⓒ $y = 2$

Ⓓ $y = 3$

5. Manuel paid $28.52 for 4 books. All of the books were the same price. What was the cost of each book?

Ⓐ $7.10

Ⓑ $7.13

Ⓒ $9.30

Ⓓ $9.36

6. What value of x makes this equation true?

$$\frac{x}{7} - 4 = -1$$

Ⓐ 28

Ⓑ 21

Ⓒ 14

Ⓓ 7

7. Which has the same value as $\frac{4x}{7}$?

Ⓐ $\frac{8}{14}$

Ⓑ $\frac{6x}{14}$

Ⓒ $\frac{11x}{14}$

Ⓓ $\frac{12x}{21}$

8. Which of the following shows the distributive property?

Ⓐ $(3 + 9) + 4 = 4 + (3 + 9)$

Ⓑ $(27 + 3) + 0 = 27 + 3$

Ⓒ $(27 + 3) + 4 = 27 + (3 + 4)$

Ⓓ $3(9 + 4) = 27 + 12$

9. Simplify the fraction completely:

$$\frac{4}{5} + \frac{3}{10} - \frac{1}{2}$$

Ⓐ 6

Ⓑ $1\frac{1}{10}$

Ⓒ $1\frac{1}{5}$

Ⓓ $\frac{3}{5}$

10. Which of the following is NOT true?

Ⓐ Adding zero to any nonzero number will not change the value of the nonzero number.

Ⓑ Multiplying zero to any nonzero number will not change the value of the nonzero number.

Ⓒ Multiplying one to any nonzero number will not change the value of the nonzero number.

Ⓓ Dividing any nonzero number by one will not change the value of the nonzero number.

11. There were 20 fish at the pet store. Bryan bought $\frac{3}{5}$ of the fish. How many fish did Bryan buy?

(A) 5

(B) 12

(C) 15

(D) 17

12. Which expression is equivalent to $4a - 4b$?

(A) $2(2a - b)$

(B) $4a - b$

(C) $4(a - b)$

(D) $4ab$

13. Steve mixed together the following ingredients to make blueberry muffins: 2 cups of milk, $\frac{1}{10}$ cup of vegetable oil, 3 cups of flour, $\frac{2}{5}$ cup of sugar $\frac{1}{2}$ cup blueberries. What is the total amount of all five ingredients?

(A) 5 cups

(B) $5\frac{2}{5}$ cups

(C) $5\frac{3}{10}$ cups

(D) 6 cups

14. If $a = 13$ and $b = 4$, then $ab + 2b - 2a =$

- (A) 60
- (B) 52
- (C) 40
- (D) 34

15. Which has the same value as $(230 \cdot 440) + (230 \cdot 2) + (230 \cdot 8)$?

- (A) $230 \cdot 440$
- (B) $690 \cdot 450$
- (C) $690 \cdot 440$
- (D) $230 \cdot 450$

16. Which of the following shows a way the distributive property could be used to multiply $3\frac{3}{4}$ by 10?

- (A) $3 + (\frac{3}{4})(10)$
- (B) $\frac{3}{4} + 3(10)$
- (C) $10(4 - \frac{1}{4})$
- (D) $10(4 + \frac{1}{4})$

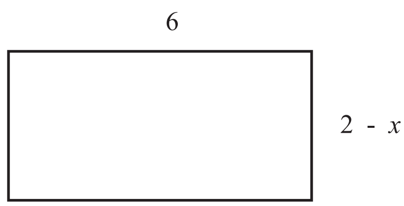
17. A library bookshelf has 4 rows filled with books. Each row has 6 books. The librarian decides to add books to the remaining empty rows of the bookshelf. The librarian places 6 books on each of these 3 empty rows. Which expression represents the total number of books on the bookshelf?

- (A) $4(6 + 3)$
- (B) $4(6) + 3$
- (C) $6(4) + 3$
- (D) $6(4 + 3)$

18. Solve for x: $\frac{3}{4} + \frac{x}{6} - \frac{2}{3} = \frac{1}{6}$

- (A) -5
- (B) $\frac{1}{2}$
- (C) 2
- (D) $\frac{2}{3}$

19. Find the area of this rectangle:



Ⓐ $8 - x$

Ⓑ $12 - 6x$

Ⓒ $12 - 2x$

Ⓓ $8 - 6x$

20. John had \$430 dollars. He spent $\frac{3}{5}$ of his money. How much money did he have left?

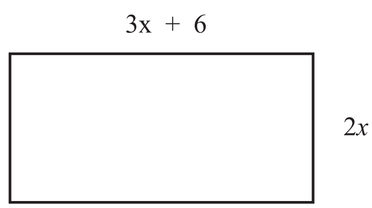
Ⓐ \$258

Ⓑ \$86

Ⓒ \$344

Ⓓ \$172

21. Which of the following expressions is equivalent to the perimeter of the rectangle shown below?



Ⓐ $2\,(11x)$

Ⓑ $2\,(5x + 3)$

Ⓒ $2\,(3x + 8)$

Ⓓ $2\,(5x + 6)$

22. A car was driven 55 miles per hour for 4 hours, then 58 miles per hour for 2 hours. What is the total distance the car was driven?

Ⓐ 220 miles

Ⓑ 274 miles

Ⓒ 330 miles

Ⓓ 336 miles

23. In a game, 5 blue pebbles can be traded for 2 red pebbles. How many red pebbles would you get for 15 blue pebbles?

Ⓐ 6

Ⓑ 10

Ⓒ 12

Ⓓ 15

24. Clark earns $42 for 4 hours of work. At that rate, how long would he have to work to earn $840?

Ⓐ 20 hours

Ⓑ 40 hours

Ⓒ 80 hours

Ⓓ 96 hours

25. On a farm, 78 square feet of grass is enough to feed 3 cows. How many square feet of grass is needed to feed 7 cows?

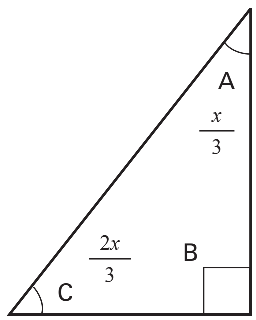
Ⓐ 26 square feet

Ⓑ 82 square feet

Ⓒ 116 square feet

Ⓓ 182 square feet

26. The triangle shown here is a right triangle. What is the measure of angle A?



Ⓐ 90 degrees

Ⓑ 60 degrees

Ⓒ 30 degrees

Ⓓ 10 degrees

27. Julia can type 60 words per minute. How many words can she type in 20 seconds?

Ⓐ 20 words

Ⓑ 30 words

Ⓒ 80 words

Ⓓ 120 words

28. Grace took $\frac{2}{3}$ of the cookies from the cookie jar and Chris took 1/6 of the cookies from the cookie jar. What fraction of cookies remained in the cookie jar?

(A) $\frac{1}{3}$

(B) $\frac{2}{3}$

(C) $\frac{1}{6}$

(D) $\frac{5}{6}$

29. The following chart describes how quickly the students in an acting class can memorize pages of sentences. Who memorizes pages of sentences the slowest?

Student	Memorization Rate
Jessica	10 pages per hour
Grayson	3 pages every half hour
John	1 page every 40 seconds
Dawn	6 pages every 20 minutes

(A) Jessica
(B) Grayson
(C) John
(D) Dawn

Appendix BB:

Grade 8 Transfer Measure 2009/2010

Student ID # ________

PowerSource

Fill in the best answer to each question.

1. If the ratio 7 to 13 is the same as the ratio x to 52, what is the value of x?

   Ⓐ 7

   Ⓑ 13

   Ⓒ 28

   Ⓓ 364

TM-105

2. Sam wanted to find three consecutive even numbers that add up to 84. He wrote the equation $k + (k + 2) + (k + 4) = 84$. What does the letter k represent?

   Ⓐ The least of the three even numbers

   Ⓑ The middle even number

   Ⓒ The greatest of the three even numbers

   Ⓓ The average of the three even numbers

TM-106

3. Carla paid x zeds for 3 cartons of juice. What is the price in zeds of 1 carton of juice?

Ⓐ $\frac{x}{3}$

Ⓑ $\frac{3}{x}$

Ⓒ $3 + x$

Ⓓ $3x$

4. If $x = -3$, what is the value of $-3x$?

Ⓐ -9

Ⓑ -6

Ⓒ -1

Ⓓ 1

Ⓔ 9

5. Which of the following is true when a, b, and c are different real numbers?

Ⓐ $a - b = b - a$

Ⓑ $a(b - c) = b(c - a)$

Ⓒ $b - c = c - b$

Ⓓ $ab = ba$

Ⓔ $ab - c = ab - b$

6. The table shows some values of x and y, where x is proportional to y:

x	4	8	Q
y	9	P	45

What are the values of P and Q?

Ⓐ $P = 40$ and $Q = 13$

Ⓑ $P = 18$ and $Q = 17$

Ⓒ $P = 20$ and $Q = 18$

Ⓓ $P = 40$ and $Q = 18$

Ⓔ $P = 18$ and $Q = 20$

7. What is the value of 1 - 5 • (-2)?

Ⓐ 11

Ⓑ 8

Ⓒ -8

Ⓓ -9

8. If n is a negative integer, which of these is the largest number?

Ⓐ $3 + n$

Ⓑ $3 \bullet n$

Ⓒ $3 - n$

Ⓓ $3 \div n$

9. Write this expression as simply as possible.

$$\frac{9k^2}{3k} =$$

TM-113

10. The number 0.01 can be written in many ways.

a) Write the number 0.01 using words. For example, 10 would be written as "ten" and 35 would be written as "thirty-five".

b) Write the number 0.01 as a fraction.

c) Write the number 0.01 as a percent.

TM-114

11. A scoop holds $\frac{1}{5}$ kg of flour. How many scoops of flour are needed to fill a bag with 6 kg of flour?

TM-115

a) Answer:____________________

b) Explain how you figured out the answer to part a:

12. At a market, 7 oranges and 4 lemons cost 43 zeds, and 11 oranges and 12 lemons cost 79 zeds. Using x to represent the cost of an orange and y to represent the cost of a lemon, write two equations that could be used to find the values of x and y.

Equation 1: ____

Equation 2: ____

TM-116

13. If $y = 3x + 2$, explain all the steps you must take to rewrite this equation so that x is expressed in terms of y.

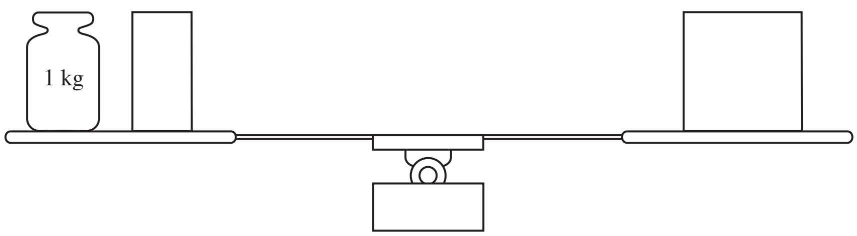
14. If $x - y = 5$ and $\frac{x}{2} = 3$, what is the value of y?

- (A) 6
- (B) 1
- (C) -1
- (D) -7

15. Explain why $2x - 3y + 7x + 5y$ can be simplified to $9x + 2y$.

TM-119

16. The objects on the scale make it balance exactly. On the left pan there is a 1 kg weight (mass) and half a brick. On the right pan there is one brick.



TM-120

What is the weight (mass) of one brick?

- (A) 0.5 kg
- (B) 1 kg
- (C) 2 kg
- (D) 3 kg

17. If $\frac{a}{b} = 70$, then $\frac{a}{2b} =$

Ⓐ 35

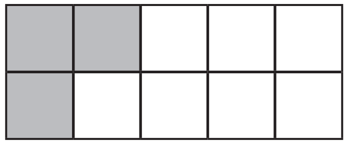
Ⓑ 68

Ⓒ 72

Ⓓ 140

18. For the expression $3 + 15 \div 3 - 4 \bullet 2$, explain why adding 3 and 15 is not your first step when you simplify the expression.

19. We want $\frac{4}{5}$ of the small squares to be shaded in the figure below.



a) First, explain how to use the multiplicative identity property to figure out how many total small squares (out of 10) we want shaded.

b) Total number of small squares (out of 10) we want shaded:_____________

c) Explain how to find out how many MORE squares need to be shaded so that $\frac{4}{5}$ of the small squares are shaded.

20. In one week Jamal watched television for 26 hours. In that week: He watched television for the same length of time on Monday, Tuesday, Wednesday, and Thursday. On each Friday, Saturday and Sunday, he watched television for twice as long as on Monday. How long did he spend watching television on **Saturday**? Write your answer in hours and minutes.

TM-124

21. A biologist needs to estimate the size of the deer herd on a wildlife reserve. The biologist captures 150 deer, then tags and releases them. A week later, the biologist captures 50 deer and counts the number tagged and the number of untagged deer. There are 15 tagged deer and 35 untagged deer in this group. The ratio of tagged to untagged deer in this group is the same as the ratio of tagged to untagged deer in the entire herd.

a) If the number of deer in the herd is represented by the unknown *d*, write an equation that shows the ratio of tagged deer to total deer in the captured group is equal to the ratio of tagged deer to total deer in the entire herd.

b) How many untagged deer are in the total herd? Show your calculations.

TM-125