Abstract Title Page
Not included in page count.

Title: Linking Implementation Fidelity to Impacts in an RCT

Authors and Affiliations:

Fatih Unlu (Corresponding Author)
Abt Associates Inc.
55 Wheeler Street Cambridge, MA 02140
Tel: 617-520-2528, Fax: 617-386-7660, E-mail: fatih_unlu@abtassoc.com

Laurie Bozzi
Abt Associates Inc.
55 Wheeler Street Cambridge, MA 02140
Tel: 617-349-2485, Fax: 617-386-8334, E-mail: laurie_bozzi@abtassoc.com

Carolyn Layzer
Abt Associates Inc.
55 Wheeler Street Cambridge, MA 02140
Tel: 617-520-3597, Fax: 617-386-8500, E-mail: carolyn_layzer@abtassoc.com

Arthur Smith
Abt Associates Inc.
55 Wheeler Street Cambridge, MA 02140
Tel: 617-520-3671, Fax: 617-386-8466, E-mail: arthur_smith@abtassoc.com

Cristofer Price
Abt Associates Inc.
4550 Montgomery Avenue Suite 800 North Bethesda, MD 20814
Tel: 301-634-1852, Fax: 301-828-9807, E-mail: cristofer_price@abtassoc.com

Richard Hurtig
Professor & Starch Faculty Fellow
Dept. of Comm. Sciences & Disorders, 120B SHC, The University of Iowa, Iowa City,
IA 52242, Tel:319-335-8730, Email: Richard-Hurtig@uiowa.edu

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through the Grant R305G04145 to the University of Iowa, Abt Associates Inc. subcontractor. The opinions expressed are those of authors and do not represent views of the funding agency (Institute of Education Sciences, U.S. Department of Education).

Abstract Body

Background / Context: In experimental studies, researchers are often interested in secondary research questions that explore important aspects of main findings, such as whether or not program effects vary according to the level of fidelity in which the program has been implemented; or according to the dosage received by individuals participated in the program. Except in planned variation studies these kinds of questions cannot be addressed directly within the standard experimental framework because participants are not randomly assigned to specific fidelity or dosage levels, and fidelity and dosage are not observed in the control group. These problems limit the conventional approaches used to examine these questions such as comparing a portion of the treatment group (higher or lower implementers) to the entire control group or correlating implementation measures with outcomes only within the treatment group.

Purpose / Objective / Research Question / Focus of Study: We faced the limitations listed above in the experimental evaluation of the Breakthrough to Literacy (BTL) program in Chicago, where schools were randomly assigned to implement the program (treatment group) or not (control group). In this evaluation, stakeholders had asked the question of whether higher fidelity of implementation was related to greater impact of the intervention on student outcomes. This paper addresses this question and compares the methodology and results of three approaches for linking implementation fidelity to impacts.

Setting: The data used in this paper were originally collected in the study *Breakthrough to Literacy in the Chicago Public Schools: A Large-Scale Evaluation* (CLIMBERs). This was a five-year study investigating the impacts of the early elementary curriculum Breakthrough to Literacy (Wright Group/McGraw-Hill) in a sample of Chicago public preschool and kindergarten classrooms. This paper reports results from the kindergarten sample.

Population / Participants / Subjects: 44 schools were recruited and randomly assigned to either a treatment condition (BTL) or a "business-as-usual" control condition. The sample of study participants in the analysis reported here included all 133 kindergarten teachers from the original sample who had classroom observation and student assessment data from the spring of 2006 or 2007, or both (75 BTL and 58 control teachers). In addition, all kindergarten students in these teachers' classrooms who had assessment data from the same periods were included (1099 BTL and 787 control students). Exhibit 1 presents baseline characteristics of the analytic sample.

Intervention / Program / Practice: BTL is a comprehensive literacy curriculum designed to be incorporated into classroom instruction throughout the school day across topic areas and developmental activity centers. It is designed to support the development of children's vocabulary, oral language, and comprehension skills throughout the classroom curriculum with five "daily essential practices": listening to and discussing stories; reading; writing; individualized software instruction; and talking, reading and writing at home. BTL instruction is organized around a weekly theme that is centered on a book read aloud daily by the teacher. Each *Book of the Week*[1] is part of a theme and was deliberately chosen to be part of BTL for didactic reasons (e.g., promoting understanding of vocabulary and basic concepts such as colors and sizes). The *Book of the Week* is read aloud to the whole class by the teacher daily, and she or

[1] Published by Wright Group/McGraw-Hill.

he promotes active discussion with children, asking questions and making connections between the book and children's experiences and ideas. In addition, BTL also incorporates computer use into its activities. BTL software presents the *Book of the Week* as well as associated interactive activities and other books for students to use at their own pace. Kindergarten-age children are expected to spend about twelve to fifteen minutes a day using the individualized, self-paced BTL literacy activities on the computer. The software is also loaded with a tracking program that allows teachers to monitor their students' progress in a variety of early literacy sub-skill areas.

Research Design: In the CLIMBERs study, 44 schools were recruited and randomly assigned to either a treatment condition (BTL) or a "business-as-usual" control condition. To examine whether the impacts of BTL varied by the level of implementation fidelity, we needed a reliable method for predicting which control group teachers would likely have implemented BTL at a high level had they been assigned to the intervention. This could be considered the *potential* implementation fidelity, which is not observable. For this purpose, we used three approaches which we will refer to as the "cut-off method", the "matching method", and the "interaction approach". These approaches are based on the work of Schochet & Burghardt (2007) and Peck (2003) and use the following common steps: construction of an implementation index measuring the fidelity of program implementation in the treatment group; fitting a regression model to treatment teachers' data that predicts their implementation fidelity as a function of baseline school, teacher, and student characteristics; and calculation of predicted implementation ratings for both BTL and control teachers using this model. At the completion of these steps, each treatment teacher has a predicted and an actual observed implementation rating, and each control teacher has a predicted implementation rating.

The three methods differ in how the predicted implementation ratings are employed. Specifically, *the cut-off method* divides BTL and control teachers into two subgroups using the predicted rating: high predicted implementers and low predicted implementers. The impact of BTL is then estimated separately for each subgroup and the resulting impacts are compared. As pointed out by Schochet and Burghardt (2007), this approach is expected to produce unbiased impact estimates for the two subgroups since both the observable and unobservable characteristics are expected to be balanced across the BTL and control units within each subgroup, which is ensured by random assignment and the partitioning of the two groups based on an index that is a function of only pre-treatment exogenous measures. Whether the unbiased impact estimates for the predicted "high" and "low" subgroups reflect the relationship between implementation fidelity and impacts, however, depends on how well the covariates used in the prediction regression can distinguish high implementers from low implementers.

The matching method matches each BTL teacher with a control teacher with the most similar predicted implementation rating. The matched BTL and control teacher-pairs are then divided into high and low implementer groups using the actual implementation rating of the BTL teachers and separate impacts are estimated for each subgroup. Unlike the cut-off approach, matching approach yield subgroup impact estimates that may be biased unless matching yields perfectly balanced BTL and matched control teacher pairs in terms of both observable and unobservable characteristics. This depends on how well the covariates used to predict the implementation ratings capture being a high implementer or not. An advantage of this method

over the cut-off method is that the classification of BTL teachers as higher or lower implementers is without any error since it is based on the actual implementation ratings.

The interaction approach entails using the interaction of the predicted implementation rating with the BTL indicator in the impact model to examine whether the impact varies with predicted implementation fidelity. Both the predicted implementation rating and the BTL indicator are exogenous measures, as the former is a function of baseline exogenous measures and the latter is the result of random assignment. Hence, the interaction of the two variables is also exogenous and the coefficient estimate for the interaction variable is an unbiased estimate. As before, whether that estimate truly pertains to the degree of implementation fidelity depends on the predictive power of the independent variables in the regression model.

Data Collection and Analysis: Baseline measures for students included the *Peabody Picture Vocabulary Test, 3rd Edition* (PPVT-III; Dunn & Dunn, 1997) and the Print Knowledge subtest of the *Test of Preschool Early Literacy* (as Pre-CTOPPP; Lonigan, Wagner, Torgesen, Rashotte, 2004; and as TOPEL, 2007) while post-test outcome measures included three subtests of the Woodcock *Reading Mastery Test – Revised/Normative Update* (WRMT-R/NU; Woodcock, 1998) – Letter Identification, Word Identification, and Word Attack, and the *Expressive One-Word Picture Vocabulary Test – Third Edition* (EOWPVT; Brownell, 2000).

Classroom observation data were collected using the *Observation Measures of Language and Literacy Instruction* (OMLIT) from both treatment and control classrooms in the spring of 2006 and 2007. OMLIT data were then used to create single-item indicators measuring teachers' use of instructional practices that were closely aligned to BTL. The item indicators were derived from an intervention fidelity checklist created by the researchers with input from the BTL developers. The BTL implementation measure was created in an iterative process in which we compared OMLIT data items with the fidelity checklist and incorporated feedback from the developers in order to find the ideal compromise between the available OMLIT data and the original BTL implementation measure. The resulting BTL implementation index had seven conceptual dimensions that mapped onto the original BTL fidelity checklist's five categories (which were reads & discusses; writes; individualized software instruction; classroom culture and management routines; and reads, writes, and talks at home) as seen in Exhibits 2 and 3.

Impacts for the full sample and subgroups were estimated using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) to account for clustering. Two-level models (teachers nested within schools) were used for teacher outcomes while three-level models (students nested within classrooms within schools) were used for student outcomes. Blocking variables, baseline measures (whenever possible), student and classroom attributes were utilized as covariates to improve precision of the impact estimates.

Findings / Results: We fit an OLS model to 75 BTL teachers' data and predicted their BTL index rating as a function of their baseline teaching quality and classroom and school characteristics as presented in Exhibit 4. The resulting model is then used to predict BTL and control teachers predicted BTL implementation ratings.

Panel A in Exhibit 5 presents the full-sample impact estimates, the standard errors of the impact estimates, the corresponding p-values, and the effect sizes to serve as a reference and inform the results for the subgroup analyses. As seen, none of these estimates are statistically significant and the largest is 0.11 of a standard deviation (SD). Panel B of Exhibit 5 presents the impact estimates and associated statistics for the predicted higher and lower implementing subgroups from the cut-off approach, which yielded 37 BTL and 18 control teachers in the higher group and 38 BTL and 40 control teachers in the lower group. These results show that while none of the estimated impacts for the high implementers are significant at the usual p=0.05 level, two were significant at the p=0.10 level (Expressive Vocabulary, p-value=0.08; Word Attack, p-value=0.061) and the effect sizes for three measures were larger than 0.20 SDs, ranging between 0.23 and 0.27 (Expressive Vocabulary 0.26 SD, Word Attack 0.27 SD., and Letter ID 0.23 SD). The impact estimates for the lower implementers, on the hand, were all small (two are negative and two are positive) and far from being significant with effect sizes between -0.13 and 0.04.

Panel C of Exhibit 5 shows results for the higher and lower implementing subgroups yielded by the matching method. These results were parallel to those described above. We also investigated whether the estimated impacts for the higher implementers were different than those for the lower implementers across the two methods (presented in Exhibit 6), which suggest that higher implementers had a higher average impact than lower implementers and that the difference between the impacts in effect size units ranges between 0.19 and 0.48 of a standard deviation where two differences were statistically significant at the p=0.10 level.

Finally, the results from the interaction method are presented in Exhibit 7, which includes impact estimates, associated standard errors and p-value as well as the corresponding effect sizes of the coefficient on the BTL indicator and its interaction with the predicted rating. The statistics included in Panel A are from the impact model estimated using the full-sample and they serve as a reference for the statistics presented in Panel B, which shows coefficient estimates on the BTL indicator, the predicted implementation rating (denoted by "PRI"), and the interaction between these two terms as well as the BTL impact and the corresponding effect sizes calculated at the 25^{th} and 75^{th} percentile of the predicted rating in the BTL group. Exhibit 7 shows that the coefficient estimate on the interaction term is positive for all measures and it is significant at the p= 0.10 level for Expressive Vocabulary and Word ID. For Expressive Vocabulary, the coefficient estimate suggests that a one-point difference in the implementation rating (which corresponds to roughly 40% of a standard deviation of the measure) is associated with almost a 1 NCE score difference in the estimated impact, which corresponds to an effect size of 0.06 SDs.

Conclusions: In the BTL context, the three approaches used to examine the relationship between impacts and fidelity of program implementation yielded similar substantive conclusions and point to a positive association between the impact of BTL and the implementation fidelity of BTL. We believe these are promising approaches for studying the relationships between aspects of the implementation of a program and its effects. These approaches are especially attractive because they rely primarily on exogenous baseline characteristics, rather than actual implementation measures, which are endogenous. The cut-off method has the added attraction that the results are relatively easy to convey and interpret. In future studies utilizing RCTs, researchers could include plans for conducting such exploratory analyses and developing data collection instruments to measure detailed baseline characteristics to facilitate these analyses.

Appendices

Not included in page count.

Appendix A. References

Brownell, R. (2000). Expressive One-Word Picture Vocabulary Test – Third Edition (EOWPVT). Novato, CA: Academic Therapy Publications.

Dunn, L.M., & Dunn, L.M. (1997). Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III). Circle Pines, MN: American Guidance Service, Inc.

Goodson, B. D., Layzer, C. J., Smith, W. C., & Rimdzius, T. (2004). Observation Measures of Language and Literacy Instruction (OMLIT). Developed as part of the Even Start Classroom Literacy Interventions and Outcomes Study, under contract ED-01-CO-0120, as administered by the Institute of Education Sciences, U.S. Department of Education.

Goodson, B.D., Layzer, J.I., & Layzer, C.J. (2004). Quality of Early Childhood Care Settings (QUEST). Developed as part of the National Study of Child Care for Low-Income Families, under contract to the Agency for Children and Families, U.S. Department of Health and Human Services.

Lonigan, C.J., Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (2007). Test of Preschool Early Literacy (TOPEL). Austin, TX: PRO-ED, Inc.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation*, 2(24), 157-187.

Schochet, P. Z., & J. Burghard. (2007). Using propensity scoring to estimate program-related subgroup impacts in experimental program evaluations. *Evaluation Review*, 31, 95-120.

Woodcock, R.W. (1998). Woodcock *Reading Mastery Test – Revised/Normative Update* (WRMT-R/NU). Circle Pines, MN: American Guidance Service, Inc.

Wright Group/McGraw-Hill. (2004). *Breakthrough to Literacy*®. Chicago, IL: Author.

Appendix B. Tables and Figures

Not included in page count.

Exhibit 1

Baseline characteristics of the teachers, classrooms, and schools of sample BTL and control teachers.

	BTL		CONTROL	
Teacher baseline characteristics	**Mean**	**SD**	**Mean**	**SD**
Second year of study participation (%)	34.7%		36.2%	
QUEST language development subscale	2.0	0.4	1.9	0.4
QUEST TV subscale	2.9	0.3	2.8	0.3
QUEST average score (minus language and TV subscales)	1.8	0.3	1.8	0.3
Arnett Caregiving Rating Scale	3.1	0.5	3.0	0.4
School baseline characteristics				
Enrollment	1041.3	454.4	801.9	260.1
Mobility rate	22.4	10.2	24.1	8.8
Percentage of students from families with low incomes	91.1%	14.6%	77.7%	27.0%
Percentage of students from minority groups	92.1%	5.2%	87.1%	9.8%
Class baseline characteristics				
Class average Pre-CTOPPP score	23.9	5.8	25.0	5.7
Class average standardized PPVT score	81.5	9.3	82.1	11.6
Percentage of students who spoke only English	50.5%	42.7%	47.6%	40.9%
Percentage of students who were girls	50.0%	10.0%	50.0%	10.0%

[a] **Exhibit reads**: The data from 34.7 percent of the BTL teachers and 36.2 percent of the control teachers came from their second year of study participation.

[b] The sample size was 133 teachers (75 BTL and 58 control teachers). In this table, however, two to four teachers in each of the groups were missing data for the class average Pre-CTOPPP and PPVT scores as well as the proportion of the percentage of students in the class who were girls.

Exhibit 2

Comparison of the dimensions measured in the final BTL implementation index and the original BTL fidelity categories

Original **BTL** *Fidelity Checklist Categories*	*Adapted* **BTL** *Implementation Index Dimensions*
Book of the Week: Reads & Discusses	**Comprehension while reading aloud:** Comprehension activities conducted as part of shared reading experiences when the teacher was reading aloud with more than one child. **Comprehension in non-reading activities:** Comprehension activities conducted outside of shared reading experiences. **Vocabulary:** Activities aimed at fostering vocabulary development. **Oral language:** Activities aimed at developing children's oral language skills other than oral vocabulary.
Daily Writing	**Writing:** Activities and materials intended to support the development of emergent writing skills
Individualized Software Instruction	**Individualized software instruction:** Use of computer for instructional purposes.
Classroom Culture and Management Routines	**Classroom culture and routines:** Organization and management of the classroom
Reads, Writes, and Talks at Home	**Not measured**

Exhibit 3

Items comprising the seven conceptual dimensions of the BTL implementation index

Comprehension while reading aloud

The teacher read a book aloud.*

The teacher read aloud a Book of the Week or a book connected to a class theme.

The teacher used one or more strategies promoting comprehension while reading aloud.*

The teacher made at least one connection between the book being read and children's experiences or class themes.*

The teacher asked book-related questions at the end of the read aloud.*

The average depth of post-read-aloud discussions or activities was moderate to "high" (e.g., they lasted at least 5 minutes and/or they extended or reinforced the comprehension of the book).*

Comprehension in non-reading activities

The teacher provided one or more activities that supported the development of comprehension skills outside of read-alouds.*

The average quality of discussions was moderately high or high (e.g., the content or topics were rich, abstract, open-ended, extended beyond the here-and-now, and/or related activities to children's' experiences).*

Vocabulary

The teacher emphasized vocabulary in at least one read aloud.*

The quality of story-related vocabulary instruction during read alouds was moderate to high (e.g., one or more vocabulary word was discussed and a comprehension support was used).

The teacher provided one or more activities that supported vocabulary development outside of read-alouds (e.g., vocabulary knowledge was provided in the context of explanations, writing, songs, stories, rhymes, language games, discussions, shared reading, emergent writing, and/or child tagging or matching).*

Individualized software instruction

The average child used a computer for the amount of time recommended by BTL or longer (12-15 minutes).

Oral language

The teacher provided at least one literacy activity that afforded students knowledge of oral communication and/or listening skills.

The teacher provided one or more activities that included language games, rhymes, songs, storytelling, or discussion; that afforded students knowledge in the comprehension of a text; that included an average depth of the discussion that was moderately high or high (e.g., they included turn-taking by students and the teacher elaborated or asked students to elaborate on comments).*

The teacher used one or more instructional strategies that promoted higher-order thinking during or after a read-aloud session (e.g., they asked book-related open-ended questions that required speculation, and/or expanded responses, thinking, or analyses).*

The teacher provided at least a few opportunities for oral language.

The average quality of the oral language activities provided by the teacher was moderate to high (e.g., the teacher provided integrated, higher quality oral language opportunities of varying types for most students individually or in small groups in which the teacher scaffolded or extended discussions with multiple turns that focused on non-management topics).

Writing

Examples of children's writing were on display in the classroom (other than just names).

Exhibit 3

Items comprising the seven conceptual dimensions of the BTL implementation index

Children spent time in journal-writing activities.
The teacher used at least one high quality activity aimed at developing knowledge of the functions and features of print (e.g., using authentic print materials, graphic organizers, or word webs).
The teacher provided a separate writing area and/or writing materials accessible to children.
In at least some writing activities, the writing was done by the children themselves rather than by adults.
The teacher sometimes, often, or consistently encouraged children to write on their own rather than insisting on conventional letter formation or spelling.
The teacher provided one or more literacy activities that involved emergent writing, copying, or tracing.
Children were engaged in one or more writing activity (including emergent writing, copying, and tracing) that was connected to a class theme during the day.
The teacher worked with most or all of the students in writing activities during the day.
Writing activities and opportunities were sometimes, often, or consistently conducted with individuals or small groups.
The teacher provided two or more opportunities to engage in writing.
The teacher provided three or more types of writing activities.
The teacher provided children with writing opportunities that were sometimes, often, or consistently of higher quality (e.g., emergent writing, captioning, dictating, writing names on work, book-making, and/or writing in journals.
The teacher provided writing opportunities that were sometimes or often integrated into activities with goals other than literacy.
Classroom culture and routines
The classroom was well-organized and had the space and layout to afford children a variety of independent activity choices, including learning centers (e.g. it had at least one distinct activity center, some materials that were marked, sorted, and stored; the layout allowed for at least two group sizes, some independent choices for students, sufficient space, adequate light, and/or no odors).
The average child spent less than 33% of the time in transitions, routines, and management or being uninvolved in activities.
The average child spent at least 50% of their time engaged in educationally "high value" activities (including: reading, alphabet, oral language, sounds, writing, science/nature, math concepts, dramatic play, creative play, block play, and fine motor play).
The average child spent at least 20% of their time in groups of 5 or fewer (not including meals, routines, transitions, or management).

[a] Item Scores: For most items, a teacher scored one point if they engaged in the described instructional activity. Items with an "*" were scored as follows: "0" if the teacher never engaged in the instructional activity, "1" if they engaged in the activity with a small or a large group, and "2" if they engaged in the activity with *both* a small and a large group.

[b] Items from the OMLIT (Goodson et al., 2004).

Exhibit 4

Parameter estimates from the model predicting BTL teachers' BTL implementation rating as a function of baseline teacher, school, and classroom characteristics

	Beta	se(Beta)	Pr > |t|
Intercept	-29.22	14.76	0.053
Teacher baseline characteristics			
Second year of study participation (%)	2.99	1.15	0.012
QUEST language development subscale	-0.42	1.45	0.773
QUEST TV subscale	1.96	1.64	0.237
QUEST average score (minus language and TV subscales)	2.93	2.33	0.213
Arnett Caregiving Rating Scale	-0.61	1.42	0.671
School baseline characteristics			
Enrollment	0.00	0.00	0.171
Mobility rate	-0.03	0.06	0.636
Percentage of students from families with low incomes	0.05	0.06	0.335
Percentage of students from minority groups	0.07	0.15	0.629
Administrative area	-1.87	1.91	0.331
Primary school language	-1.61	2.66	0.547
*Administrative area * School language*	1.65	2.30	0.476
Class baseline characteristics			
Class average Pre-CTOPP score	0.05	0.13	0.708
Class average standardized PPVT score	0.10	0.08	0.246
Percentage of students who spoke only English	-0.04	0.02	0.063

[a] The sample includes 75 BTL teachers.

Exhibit 5

Estimated Impacts for High & Low Implementers

	Impact estimate	Standard Error	P-Value	Effect Size
Panel A: Impact Estimates (Full Sample)				
Expressive One-Word Picture Vocabulary Test	0.69	1.53	0.654	0.04
Word Attack	1.28	0.87	0.149	0.11
Word ID	-0.09	1.44	0.950	-0.01
Letter ID	1.37	1.19	0.257	0.10
Panel B: Cut-off Approach (High & Low Implementers)				
High Implementers				
Expressive One-Word Picture Vocabulary Test	4.35	2.38	0.080	0.26
Word Attack	3.11	1.58	0.061	0.27
Word ID	2.24	2.69	0.413	0.14
Letter ID	3.06	2.03	0.145	0.23
Low Implementers				
Expressive One-Word Picture Vocabulary Test	-0.86	1.96	0.663	-0.05
Word Attack	0.09	1.43	0.951	0.01
Word ID	-2.00	2.22	0.375	-0.13
Letter ID	0.55	1.79	0.760	0.04
Panel C: Matching Approach (High & Low Implementers)				
High Implementers				
Expressive One-Word Picture Vocabulary Test	4.37	2.54	0.098	0.26
Word Attack	2.60	1.56	0.108	0.23
Word ID	1.44	2.77	0.606	0.09
Letter ID	2.80	2.10	0.194	0.21
Low Implementers				
Expressive One-Word Picture Vocabulary Test	-3.50	2.72	0.211	-0.21
Word Attack	0.26	1.55	0.870	0.02
Word ID	-2.94	2.58	0.267	-0.19
Letter ID	-0.08	2.00	0.968	-0.01

[a] "*"Denotes statistical significance at the p= 0.05 level.

Exhibit 6

Comparison of Impacts Estimated Using the Three Analytic Approaches

	Difference in Impact	Difference in Effect Size	P-value
Panel A: Cut-off Approach			
Expressive One-Word Picture Vocabulary Test	5.21	0.32	0.099
Word Attack	3.02	0.27	0.166
Word ID	4.24	0.27	0.231
Letter ID	2.51	0.19	0.360
Panel B: Matching Approach			
Expressive One-Word Picture Vocabulary Test	7.87	0.48	0.041*
Word Attack	2.35	0.21	0.292
Word ID	4.38	0.28	0.254
Letter ID	2.88	0.22	0.326

[a] "*"Denotes statistical significance at the p= 0.05 level.

Exhibit 7

Impacts Estimated Using the Interaction Method

Outcome	Variable	Impact estimate	Standard Error	P-Value	Effect Size
Panel A: Regular Impact Model					
Expressive One-Word Picture Vocabulary Test	BTL	0.69	1.53	0.654	0.04
Word Attack	BTL	1.28	0.87	0.149	0.11
Word ID	BTL	-0.09	1.44	0.950	-0.01
Letter ID	BTL	1.37	1.19	0.257	0.10
Panel B: Interaction Model					
Expressive One-Word Picture Vocabulary Test	BTL [a]	-0.55	1.63	0.736	-0.03
	Predicted imp. Rating (PRI)	0.86	0.46	0.058	-
	BTL * PRI [b]	0.97	0.54	0.071	0.06
	BTL Effect at Low PRI [c]	-1.42	-	-	-0.09
	BTL Effect at High PRI. [d]	1.82	-	-	0.11
Word Attack	BTL	0.99	1.06	0.357	0.09
	Predicted imp. Rating (PRI)	0.22	0.30	0.460	-
	BTL * PRI	0.36	0.35	0.301	0.03
	BTL Effect at Low PRI	0.67	-	-	0.06
	BTL Effect at High PRI	1.88	-	-	0.17
Word ID	BTL	-1.36	1.74	0.439	-0.09
	Predicted imp. Rating (PRI)	0.56	0.48	0.240	-
	BTL * PRI	0.97	0.56	0.086	0.06
	BTL Effect at Low PRI	-2.22	-	-	-0.14
	BTL Effect at High PRI	1.00	-	-	0.06
Letter ID	BTL	0.29	1.38	0.832	0.02
	Predicted imp. Rating (PRI)	0.61	0.37	0.097	-
	BTL * Pred. Rat.	0.64	0.43	0.134	0.05
	BTL Effect at Low PRI	-0.28	-	-	-0.02
	BTL Effect at High PRI	1.87	-	-	0.14

[a] The estimate for BTL is the estimated impact when predicted rating is at the mean (0.4).

[b] Predicted ratings range from -6.4 to 6.6 with mean 0.4. The 25th and 75th percentiles of the predicted ratings in the whole sample are -1.6 and 2.1. The 25th and 75th percentiles of the predicted ratings in the BTL sample are -0.9 and 2.5.

[c] The estimated BTL effect when predicted rating is at the 25 percentile in the BTL sample (-0.9)

[d] The estimated BTL effect when predicted rating is at the 75 percentile in the BTL sample (2.4).