
**External Review of the Evaluation of
Mondo Publishing's
Building Essential Literacy (BEL) Design and
Bookshop Reading Program**

Jonathan Supovitz
Senior Researcher
Consortium for Policy Research in Education
University of Pennsylvania¹

¹ This report is an independent review conducted by Jonathan Supovitz. The opinions expressed are those of the author and do not reflect the positions of the University of Pennsylvania in any way.

Summary

This report is an independent external review of an evaluation report of the impact of the Mondo Building Essential Literacy (BEL) Design and *Bookshop* Reading Program (hereafter referred to as BEL/*Bookshop*) conducted by Peter Hill of the Center for Applied Educational Research at the University of Melbourne and Angela Jaggar of New York University.

Hill and Jaggar's study shows statistically significant positive effects of BEL/*Bookshop* for students in both Grades K and 1. Although students in control schools had higher pretest scores, on average, than students in BEL/*Bookshop* schools, the students in the BEL/*Bookshop* schools had higher posttest scores, on average, than students in control schools. This suggests that students in BEL/*Bookshop* schools began the school year behind their peers in the control schools, but finished the school year ahead of them. The effect size estimates for the impact of BEL/*Bookshop* on student reading were very large and educationally important—approximately one standard deviation per year in both kindergarten and first grade. However, as the evaluators note, the students in the BEL/*Bookshop* program did not meet the minimal standard set for the program. Thus, while the BEL/*Bookshop* students learned significantly more than did the comparison students, they did not achieve the goals set by the program. This is a common finding in educational research and should not detract from the accomplishments of the program.

This independent examination of Hill and Jaggar's report confirms that it is a high quality quasi-experiment study, using a sound research design and appropriate statistical methods. Hill and Jaggar have reached reasonable conclusions based upon the data. Although there are additional information and clarifications that the evaluators could provide, their analysis methods and conclusions appear to be solid, and the effects of the program that are reported are substantial and educationally important.

Program Overview

The BEL/*Bookshop* program is intended to improve the literacy of students in the early elementary school grades (K–2). The BEL/*Bookshop* design has its roots in literacy projects and research conducted in Australia. The BEL/*Bookshop* design is based on a strategy that includes the belief in the capacity of all students to attain high standards; specific standards for student performance; monitoring and assessment; classroom teaching strategies; programs, professional learning teams, school and class organization structures to facilitate the BEL/*Bookshop* instructional approaches, and special interventions and special assistance. The program also features partnerships with home and community, and leadership and coordination.

Evaluation Design

Hill and Jaggar's evaluation occurred between 1998 and 2000 using data from urban school districts in Illinois, Massachusetts, and New York. The impact evaluation of *BEL/Bookshop* involved a pretest-posttest comparison of the reading scores of 3,051 students in 20 *BEL/Bookshop* schools and 1,395 students in four control schools. Of the *BEL/Bookshop* schools, ten were in Boston, MA five were in the Bronx, NY and six were in Elgin, IL. All four control schools were in The Bronx.

All students were given reading assessments in the fall of 1998, before the implementation of *BEL/Bookshop*. They were retested in the spring of 1999, after *BEL/Bookshop* had been implemented in the trial schools for one year. The reading assessment included reading records, which produced measures of the students' current text level (i.e., 1–28). This was used as the dependent variable for the impact analyses.

Assessments in the control schools were administered by independent evaluators. Assessments in the *BEL/Bookshop* schools were administered by each student's teacher. Each year, a random sample of students in the *BEL/Bookshop* schools was retested by independent evaluators to determine the reliability and validity of the teachers' assessments.

To their credit, the evaluators have taken special care to examine the inter-rater reliability of their instruments, which is very important because it gives us confidence that differences in performance are attributable to the *BEL/Bookshop* program and not due to error in the application of the assessment instruments. As the researchers note, the reliabilities are largely solid, indicating that differences in the program results are not due to differential application of the assessments of childrens' performance.

Demographic and background information were also collected on the students. These variables were gender, language background, free lunch status, special education status, and whether the student experienced a critical event that year (e.g., death of a family member).

The impact estimates for the first two years were obtained by estimating a multilevel model of student posttest scores predicted by participation in *BEL/Bookshop*, after controlling for pretest scores and student background characteristics. These analyses were performed for only kindergarten and first grade students.

Suggestions for Additional Information and Clarification

There are several additions and clarifications that would refine Hill and Jaggar's report. These would make the report more thorough and provide more information, but they would be highly unlikely to affect the conclusions that Hill and Jaggar have reached.

Design

- Given that the intervention focuses on students in Grades K–2 and that data were collected for students in Grade 2 (at least pretest scores in 1999–2000 are reported for second graders in Table 2), it is curious that second grade results are not reported. What is the rationale for their omission?
- The evaluation begins with the precept that BEL/*Bookshop* seeks to minimize “the internal variation so often evident in schools in the performance of students.” (p.2). By stating this up front, the evaluators create the expectation that they will report on the possible reduction of variation in student performance within schools using BEL/*Bookshop*. This does not seem to be followed through in the analysis. Although this would require an even more complicated log-linear variance component in their multilevel model, it would be good to know whether there are differences in the variation in achievement within BEL/*Bookshop* schools compared to the non-BEL/*Bookshop* schools.

Sampling

- How was the control group chosen? It seems clear they were chosen in part because of their proximity to Jaggar at New York University, who conducted all of the assessments of students. But within geographic proximity, was the sample chosen so that they were similar students in schools similar to those who were to receive the BEL/*Bookshop* treatment?
- How do the control group students compare demographically to the treatment group prior to the implementation of BEL/*Bookshop*? We know something about them in that their pretest mean and standard deviation scores are reported in Table 7. The Kindergarten sample in the control group was substantially higher performing than the BEL/*Bookshop* group on the pretest, while the Grade 1 samples for treatment and control were similar. But it would be good to provide more complete demographic information about the two samples (both treatment and control).

Analysis

- Missing data. The authors do not clearly state the extent to which missing data plagued their study and how they handled missing data. The Table 4 results, for example, are reported for only those with complete pre-and posttest data, which makes perfect sense, but this raises the question of the magnitude of the missing data problem. Any additional information on missing data and how the evaluators handled it would be helpful.
- Absence of school-level control variables. The control variables used in these analyses are all measured at the student level and included as level-1 covariates. This produces an adjustment for student characteristics; however, if the schools in this analysis are dissimilar

in terms of poverty rates and student body composition, additional school-level controls may improve the accuracy of the results.

- Calculation of effect sizes. Traditional effect size estimates are calculated by dividing the difference in group means by an unadjusted standard deviation of treatment and control groups (Cohen, 1977). For this evaluation the authors report that they calculated final effect size estimates using an approach developed by Tymms, Merrell, and Henderson (1997). This involves dividing the raw coefficient representing the effect of BEL/*Bookshop* by the residual (i.e. adjusted) standard deviation of the posttest scores after controlling for student background characteristics and school and classroom membership. Thus, the final effect-size estimates represent the change in student performance relative to the scores of similar students in the same school and class, as opposed to all other students in the sample. This may cause some overestimation of the effect sizes because the adjusted standard deviation is always going to be smaller than the unadjusted standard deviation. Unfortunately, the Tymms et al. article is published in a journal with relatively limited distribution and was unavailable for this review. The authors could include more information about why the Tymms et al. method is appropriate here.

- Finally, a minor point about effect sizes. A table on page 16 describes the magnitude of small, medium and large effect sizes. The authors are providing very helpful information by giving the reader a guide with which to determine the size of the effect sizes they report. Cohen (1977) reports effect sizes of .2 as small, .5 as medium, and .8 as large. The effect sizes in the table are smaller than this. It would be helpful if the authors can cite a reference for their guide for interpreting the magnitude of coefficients. *It should be noted that even using Cohen's guide for effect sizes, the magnitude of the effects for the BEL/Bookshop program remain large.*

Conclusion

The points raised above should not detract from the overall finding that the evaluation report of the BEL/*Bookshop* program is methodologically sound and that the conclusions that the evaluators reach are in line with the results that they report. Hill and Jaggar's evaluation report provides a solid piece of quasi-experimental evidence in support of the effectiveness of the Building Essential Literacy/*Bookshop* program in improving early grade childrens' literacy skills.

References

- Cohen, J. (1977). *Statistical Power Analysis for the Social Sciences*. New York: Academic Press.
- Tymms, P., Merrell, C., Henderson, B. (1997) "The First Year at School: A quantitative investigation of the attainment and progress of pupils". *Educational Research and Evaluation*, vol. 3, no.2.

Re-review (October 2003)

In August, 2003 I submitted an independent review of the evaluation of the Mondo Building Essential Literacy (BEL) Design and *Bookshop* Reading Program, conducted by Peter Hill of the Center for Applied Educational Research at the University of Melbourne and Angela Jaggar of New York University. In that review (*see attached*) I concluded that the Hill/Jaggar report was a high quality quasi-experimental study employing a sound research design and appropriate statistical methods. I concluded then, and continue to believe, that the results that they presented represent an important piece of empirical research evidence as to the effectiveness of the Mondo-BEL/*Bookshop* treatment. In that report I did, however, suggest several revisions and additions that would improve Hill and Jaggar's evaluation report.

Based on these suggestions, Hill and Jaggar have revised the report and Mondo Publishing has asked me to re-review it. With their modifications, Hill and Jaggar have addressed virtually every suggestion that was made in my original review that could feasibly have been done. Hill and Jaggar's alterations only serve to strengthen a report that was already of high quality. Their revisions have rounded out the report, making it more thorough and readable. They have strengthened an already sound evaluation effort.

More particularly, Hill and Jaggar have taken care to describe in greater detail how they have chosen their sample and how their control group was selected. They have provided more pertinent information about the demographics of the control group in comparison to the group that received the BEL/*Bookshop* treatment, which allows the reader to better understand the initial differences between the two groups. Hill and Jaggar have also included more information about the extent of missing data in their sample and have thoroughly addressed how they have handled the crucial problem of missing data. The candor with which they confront the potential implications of missing data to their report is admirable.

The authors have also provided more clarity about their decision to use a more contemporary method to calculate effect sizes. I support the authors' decision to maintain the use of the Tymms method for calculating their effect sizes. Although there are other methods that differ from the Tymms approach, this is a developing area of statistical methodology, and there is no reason to assume that one method is superior to another.

In sum, the evaluation of the Mondo Building Essential Literacy (BEL) Design and *Bookshop* Reading Program is a solid piece of research that provides important evidence of the effectiveness of the program on student reading performance.

Revision Summary

Supovitz, in his initial independent review of the original Hill & Jagggar evaluation, suggested several revisions and additions that would further substantiate the findings. Based on these suggestions, Hill and Jagggar revised their report and addressed the suggestions where possible and appropriate. Hill & Jagggar's alterations have been validated by Supovitz as "...sound research design (that used) appropriate statistical methods... ."

Suppovitz's suggestions and Hill & Jagggar's revisions follow.

Design

- What is the rationale for not reporting on the Grade 2 results included in Table 2? *The authors have removed the trial school Grade 2 pretest scores from this table in the revised report as the external testers were unable to collect Grade 2 control school data for comparison.*
- The evaluation begins with the precept that BEL/Bookshop seeks to minimize "the internal variation so often evident in the performance of students in schools." Why is this not followed through in the analysis? *To provide these data would require additional complex modeling which the authors elected to address in a future report on the longitudinal studies of the BEL/Bookshop treatment.*

Sampling

- Was the sample chosen so that there were students in schools similar to those students who were to receive the BEL/Bookshop treatment? *The authors have provided additional information in the revised report. See the "Sample of Schools" section on pages 8–10; Tables 2 and 3.*
- It would be good if the authors were to provide more complete demographic information about the two samples (both treatment and control). *The authors have provided additional information in the revised report. See the "Sample of Schools" section on pages 8–10 and Tables 2 and 3.*

Analysis

- Any additional information on missing data and how the evaluators handled it would be helpful. *The authors have provided additional information in the revised report. See the "Data Analysis" section on page 10 and Table 4.*
- Would additional school-level controls improve the accuracy of results? *The authors have provided additional information in the revised report. See the "Measuring the Impact of BEL/Bookshop on page 15.*
- Could the authors include more information about the Tymms et al. method that was used for the calculation of effect size in this report? *The authors have provided additional information in the revised report. See the "Measuring the Impact of BEL/Bookshop" on page 16.*



For More Information contact Mondo Publishing

at (888) 88-MONDO

www.mondopub.com

980 Avenue of the Americas

New York, NY 10018