

---

An Evaluation Report on the Impact of  
Mondo Publishing's  
Building Essential Literacy (BEL) Design and  
*Bookshop* Reading Program  
1998–1999 and 1999–2000 School Years

Peter W. Hill, Ph.D.

*Centre for Applied Educational Research, The University of Melbourne*

Angela M. Jaggar, Ph.D.

*School of Education, New York University*



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

Jonathan Supovitz

*Consortium for Policy Research in Education  
University of Pennsylvania*

---

All rights reserved.

No part of this publication may be reproduced, except in the case of quotation for articles or reviews, or stored in any retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher.

For information contact:

MONDO Publishing

980 Avenue of the Americas

New York, NY 10018

Visit our website at [www.mondopub.com](http://www.mondopub.com)

Printed in the United States of America

04 05 06 07 08 09      9 8 7 6 5 4 3 2 1

ISBN 1-59336-353-2

# Contents

**Executive Summary** ..... 2

**Evaluation Report on the Impact of  
Mondo Publishing’s Building Essential Literacy (BEL) Design  
and *Bookshop* Reading Program  
Peter W. Hill, Ph.D. and Angela M. Jaggar, Ph.D.** ..... 3

The Project ..... 4  
The Intervention ..... 4  
The BEL Design for Literacy Improvement ..... 5  
Sample of Schools .....9  
Data Collection ..... 11  
Data Analysis ..... 13  
Reliability of Measures ..... 13  
Performance Against Targets and Standards .....14  
Factors Affecting Student Achievement ..... 17  
Measuring the Impact of BEL/*Bookshop* ..... 18  
The Impact of Reading Recovery ..... 20  
Conclusion ..... 20  
References ..... 21  
Appendix and Author Biographies .....22

**External Review of the Evaluation of  
Mondo Publishing’s Building Essential Literacy (BEL) Design  
and *Bookshop* Reading Program  
Jonathan Supovitz** ..... 25

External Review (August 2003) ..... 26  
Re-review (October 2003) ..... 30  
Revision Summary ..... 31

# Executive Summary

The Building Essential Literacy (BEL) and *Bookshop* Reading Program is intended to improve the literacy of students in the first three years of schooling. BEL is a whole-school design for improving literacy outcomes, and *Bookshop* is a comprehensive reading program that formed the core literacy resource in all classrooms. The BEL/*Bookshop* Program is an initiative of Mondo Publishing Inc. in conjunction with The University of Melbourne (UMelb) and the Faculty of Education, New York University (NYU).

An evaluation study, involving both trial and control schools, was conducted over the 1998–99 and 1999–2000 school years by Dr. Angela Jaggard (NYU) and Professor Peter Hill (UMelb). The report of this study is provided in the first part of this document.

This report was independently reviewed by Senior Researcher Jon Supovitz, Center for Policy Research in Education, University of Pennsylvania (CPRE). Supovitz’s comments on the evaluation study are contained in the second part of this document, together with a brief list of responses by Jaggard and Hill.

The BEL/*Bookshop* evaluation study began in September 1998, with ten trial schools in inner-city Boston, five trial schools and four control schools in The Bronx, and six trial schools in Elgin, Illinois. Complete data were obtained from 3,051 students in the initial trial group cohort K–1. The control group consisted of 1,395 students K–1. The control group schools did not receive the Mondo-BEL treatment incorporating the *Bookshop* Reading Program. In September 1999, a further eight schools—five in Boston and three in The Bronx—joined the project, and complete data were obtained from a total of 4,989 students. No Grade 2 data were collected in the control schools due to administration issues and therefore there are no Grade 2 data available for comparison to the trial schools.

To evaluate the impact of the BEL/*Bookshop* treatment during its first full year of implementation, a series of multi-level regression models was fitted to the data of students in both the control and trial schools participating in the first year of the project.

Analyses of the BEL/*Bookshop* data for the first year of implementation yielded effect size estimates of 1.19 for the effect of the program on Kindergarten students, and 1.00 for the effect of the program on Grade 1 students. At the end of the second year of implementation, further analyses were undertaken to compare the progress made by Kindergarten students in intake 1 and intake 2 schools in both the 1998–1999 and 1999–2000 school years. Effect sizes ranged from between 0.78 and 0.98. Jaggard and Hill note that these effect sizes represent extremely large effects relative to those typically encountered in the research literature.

In his independent external review of the Jaggard and Hill report, Supovitz agrees that “...the effect size estimates for the impact of BEL/*Bookshop* on reading achievement were very large and educationally important.” He also confirms that the evaluation study was “... a high quality quasi-experiment study, using a sound research design and appropriate statistical methods.”

---

An Evaluation Report on the Impact of  
**MONDO Publishing's**  
**Building Essential Literacy (BEL) Design**  
and ***BOOKSHOP*** Reading Program  
1998–1999 and 1999–2000 School Years

---

**Peter W. Hill, Ph.D.**

Centre for Applied Educational Research  
The University of Melbourne

and

**Angela M. Jaggard, Ph.D.**

School of Education  
New York University

JANUARY 2001

REVISED OCTOBER 2003

---

## BUILDING ESSENTIAL LITERACY (BEL) DESIGN AND Bookshop READING PROGRAM

### The Project

The Building Essential Literacy (BEL) and Bookshop Reading Program is intended to improve the literacy of students in the first three years of schooling. The BEL/Bookshop Program is an initiative of Mondo Publishing, Inc. in conjunction with the Centre for Applied Educational Research at the University of Melbourne, and Dr. Angela Jaggar of the School of Education, New York University. An initial study, involving both trial and control schools, was conducted over the 1998–99 and 1999–2000 school years to measure the impact of the Mondo-BEL Design, incorporating the *Bookshop* Reading Program, on student outcomes.

The project was funded by Boston Public Schools; by Community School District 11, Bronx, New York; and by Elgin District U46, Illinois, largely from federal grants. The project began in September 1998 with ten trial schools in inner-city Boston, five trial schools and four control schools in the Bronx, and five trial schools in Elgin. Complete data were obtained from 3,051 students in the initial trial group cohort K–1. The control group consisted of 1,395 students K–1. The control group schools did not receive the Mondo-BEL treatment incorporating the *Bookshop* Reading Program. In September 1999, a further eight schools – five in Boston and three in the Bronx – joined the project, and complete data were obtained from a total of 4,989 students.

### The Intervention

BEL seeks to provide a coherent, whole-school approach to early literacy that minimizes the internal variation in the performance of students so often evident in schools. Research into school and classroom effectiveness indicates that while there is significant variation among schools in the progress made by students, this variability is not as great as the differences among classes within schools.<sup>1</sup> If this internal variation can be minimized and all classes brought up to the level of the most effective classes, dramatic improvements in outcomes can be achieved.

The BEL design for school improvement was developed as the result of two earlier projects initiated in Victoria, Australia. The first of these, the Early Literacy Research Project (ELRP), began in 1995 and was a joint initiative of the State Department of Education and the Center for Applied Educational Research at The University of Melbourne. The intent was to develop a system-wide approach to maximizing the literacy achievement of students at risk of failure in the early years of schooling (ages 5–8). The project ran for three years (1996–1998) in 27 trial schools and 25 reference schools. Intensive professional development took place in each of the trial schools. The impact was so positive that the design<sup>2</sup> was adopted statewide in Victoria, Australia.

The second project, the Children’s Literacy Success Strategy (CLaSS) is a joint initiative of the Catholic Education Commission of Victoria, Australia, and the Center for Applied Educational Research at The University of Melbourne. The project began in January 1998 with an intake of 39 schools. The project, now in its fourth year, involves close to 300 trial schools.

<sup>1</sup> See especially: Hill, P.W. & Rowe, K.J. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement*, 7 (1), 1-34; Monk, D.H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Education Evaluation and Policy Analysis*, 14, 307-332, and; Scheerens, J., Vermeulen, C.J.A.J., & Pelgrum, W.J. (1989). Generalizability of instructional and school effectiveness indicators across nations. *International Journal of Educational Research*, 13 (7), 789-799.

<sup>2</sup> Crévola, C.A., & Hill, P.W. (1998). Evaluation of a whole-school approach to prevention and intervention in early literacy. *Journal of Education for Students Placed at Risk*, 3, 133-157; Hill, P.W., & Crévola, C.A. (1997). The literacy challenge in Australian primary schools. *LARTV Seminar Series, No. 69*, November, 1997, and; Hill, P.W., & Crévola, C.A. (1999a). The role of standards in education reform for the 21<sup>st</sup> century. In David D. Marsh (Ed.), *ASCD Yearbook 1999: Preparing our Schools for the 21<sup>st</sup> century*. (117-142) Alexandria, VA.: Association for Supervision and Curriculum Development.

---

## The BEL Design for Literacy Improvement

BEL is based on the same design as the one developed and implemented in the ELRP and CLaSS initiatives. The BEL design and a brief description of each of the design elements follow.

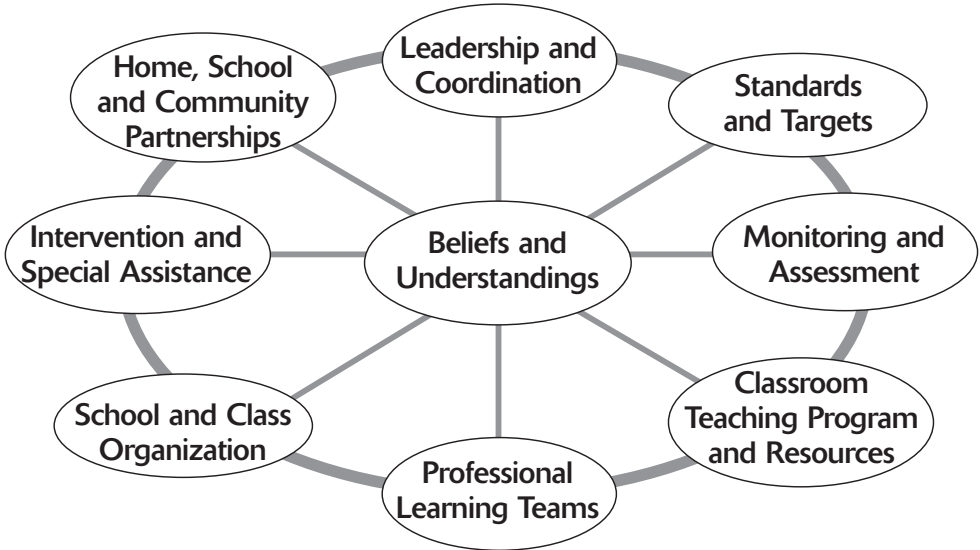


Figure 1: BEL Design for Literacy Improvement

### *Beliefs and Understandings*

Beliefs and understandings form the central element of the design template as represented in Figure 1. The BEL design is based on the notion that ongoing school improvement requires professionals who have both a deep understanding of teaching and learning and a belief in the capacity of all students to attain high standards, given the right support and sufficient time.

### *Standards and Targets*

High expectations for student achievement should be reflected in explicit standards that have been benchmarked against those of other school systems to ensure that they reflect best practice. Standards and associated targets drive the model summarized in Figure 1. They constitute the starting point for refocusing the mission of schools and for redesigning how they operate, so that meeting the standards becomes a priority. Performance standards attempt to define ‘to what level’ students will progress and by which grade. They form the basis of both long-term goals for school systems and short-term targets for individual schools and students.

BEL uses a set of performance standards and targets derived from Clay and Tuck in their *“three waves of teaching.”*<sup>3</sup> Clay and Tuck state that with good teaching in the first year of schooling, one can expect 80 percent of students to be under way with reading and writing. During the second year of schooling, with good teaching and Reading Recovery as a one-to-one intervention, one can expect to have another 18 percent under way. This leaves approximately two percent for whom further referral and special support will be necessary during the remainder of their schooling.

Standards are set in terms of student text levels, which are established by taking reading records using a set of 28 unseen graded leveled texts. Two standards are identified – a minimum standard and a target standard. These targets and standards are summarized in Table 1. Research strongly indicates that the target standards are achievable, although not in the initial years of implementation. Identification of both a minimum standard and a target standard ensures that the targets embody longer-term challenges for all schools and all students, and do not focus solely on the low end of the performance continuum.

**Table 1. BEL Standards and Targets**

<b>Stage of schooling</b>	<b>Minimum percent required at standard</b>	<b>Minimum standard</b>	<b>Target standard</b>
<b>End of Kindergarten</b>	80%	Text Level 1 or above	Text Level 5 or above
<b>End of Grade 1</b>	98%	Text Level 15 or above	Text Level 20 or above
<b>End of Grade 2</b>	98%	Text Level 26 or above	Text Level 28 or above

<sup>3</sup> Clay, M.M. & Tuck, B. (1991). *A study of Reading Recovery subgroups: Including outcomes for children who did not satisfy discontinuing criteria*. Auckland, New Zealand: University of Auckland.



---

### *Monitoring and Assessment*

Assessment is critical in monitoring student progress towards defined targets, but it is even more important as a means of establishing starting points for teaching and learning. In the BEL project, extensive assessment is conducted at the beginning and end of each school year using a series of teacher observation measures, notably the six measures comprising Clay's *An Observation Survey of Early Literacy Achievement*.<sup>4</sup> These measures both provide diagnostic information that reveals students strengths and weaknesses, and suggest focuses for instruction.

### *Classroom Teaching Strategies and Programs*

Effective teaching is both structured and focused on the learning needs of each student. This constitutes one of the greatest challenges faced by teachers, particularly given the wide range of abilities within the typical classroom. Focused literacy teaching requires expert organization and management techniques related to the teaching of small groups. It also requires teachers to be adept in implementing a range of classroom practices and strategies in response to individual needs.

In BEL schools, the emphasis is on raising the professional competence of teachers so that they are better able to implement a range of powerful teaching strategies that are both structured and focused on the learning needs of all students. It is within this three-part (whole-class/small-group/whole-class) structure that the range of instructional strategies is integrated into effective classroom practice (see Figure 2). The Mondo *Bookshop* Reading Program formed the core literacy resource in all classrooms. Each classroom set of *Bookshop* included a library of systematically leveled texts, comprehensive lesson plans for each text selection, and an array of additional materials.

### *Professional Learning Teams*

Effective professional development has an impact on teachers' beliefs and understandings and establishes a process for institutionalizing a whole-school design approach to early literacy. Within BEL, teachers in the early grades (K–2) are formed into teams that attend off-site professional development sessions. The main purpose of these sessions is to provide impetus for further thought and discussion. On-site professional development takes place daily as well. A literacy coordinator is appointed to lead the team at each school. The coordinator acts as a mentor and lead learner, and organizes demonstration teaching and classroom observation. The coordinator also chairs weekly meetings of the professional learning team. This position requires a significant time commitment.

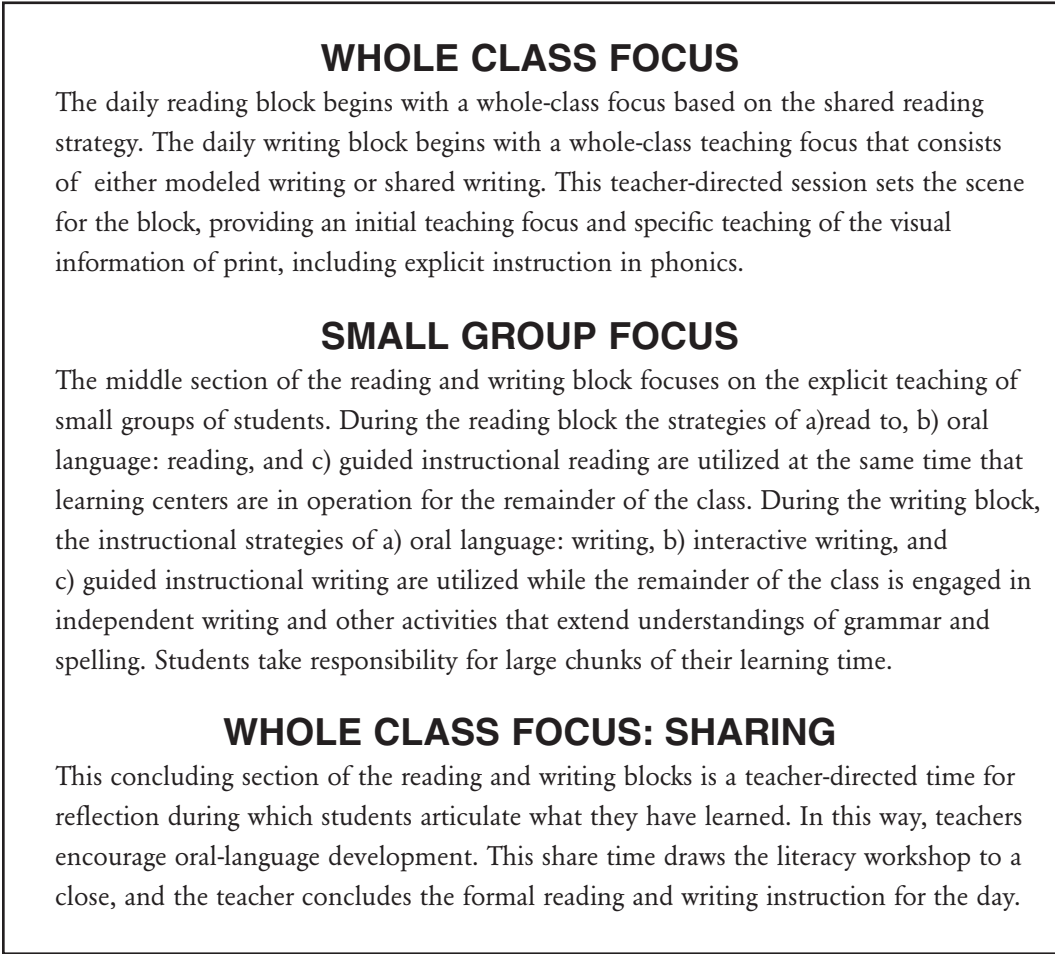
### *School and Class Organization*

For classroom teaching to be effective, key aspects of how schools operate require careful examination. Thus, attention is given within BEL to issues such as allocation of time; staffing; the provision of resources, such as the availability of multiple copies of student texts; class sizes; and so on. Schools are encouraged to develop policies to eliminate unnecessary interruptions and to establish classroom routines that minimize disruption.

<sup>4</sup> Clay, M.M. (1993). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann Education.

---

The latter involves daily use of mixed-ability groups and of instructional blocks organized according to a whole class/small group/whole class structure.



**Figure 2. Structure of the Daily Two-Hour Literacy Teaching Block**

---

### *Intervention and Special Assistance*

Even with the best teaching, experience indicates that many students need extra time and support to reach minimum standards. Research indicates that for students who are most at-risk, one-to-one intervention is the most effective.<sup>5</sup> Within BEL, Reading Recovery<sup>6</sup> is used as a mandatory element in each of the project schools to provide intervention for the lowest-achieving students in Grade 1. For students who continue to experience difficulty individual learning plans are developed, and ongoing specialist support is provided.

### *Home, School, and Community Partnerships*

There is an enormous amount of evidence that effective schools are proactive and systematic about linking with the home, prior schools, other service providers, and the wider community.<sup>7</sup> Problems such as poor attendance are more likely to be tackled successfully when schools develop genuine partnerships with the home.<sup>8</sup>

A critical design feature of BEL is the implementation programs for parents and other classroom personnel to enable them to provide appropriate support in the classroom. This support is directed at students in learning centers on task, thus enabling teachers to work fruitfully with small groups.

### *Leadership and Coordination*

Whole-school design approaches depend to a high degree on the leadership provided by the principal and senior school administration. They are the ones in a position to make sure that each of the design elements is properly attended to. On a day-to-day basis, the literacy coordinator undertakes a key leadership role. Teachers selected for this role need professional development that goes beyond gaining a better understanding of literacy and classroom teaching, and includes training as a coach, mentor, and lead learner. Separate professional development sessions are provided for school principals, who face the challenge of ensuring the initiation and implementation of the design in their schools.

## **Sample of Schools**

A group of 20 volunteer schools, designated as trial schools, agreed to participate in the BEL/*Bookshop* treatment. A further group of four schools, designated as reference/control schools, did not participate in the BEL/*Bookshop* treatment but were involved in the pre- and post-testing program. The control schools were selected partly on the basis of the Bronx, NY districts proximity to NYU and for access to the testing team. They were also self-selected to some extent, being those schools in the NYC district that responded positively to the request to become control schools. (Many schools that were approached declined, not seeing any benefit in being a control school.) Finally, they were chosen because they had student demographic profiles that matched the profiles of the trial schools. The schools and the number of students in each school are summarized in Table 2.

<sup>5</sup> See especially: Wasik, B.A. & Slavin, R.E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 179–200.

<sup>6</sup> Clay, M.M. (1993a). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann Education.

<sup>7</sup> See especially: Cairney, T.H., Ruge, J., Buchanan, J., Lowe, K. & Munsie, L. (1995) *Developing partnerships: The home, school and community interface*, Canberra: Department of Employment, Education and Training, and; Epstein, J.L. (1995). *School/family/community partnerships: Caring for children we share*. *Phi Delta Kappan*, 76, 701–712.

<sup>8</sup> Chrispeels, J. (1996). Effective schools and home-school-community partnership roles: A framework for parent involvement. *School Effectiveness and School Improvement*, 7, 297–323.

**Table 2. Schools and Number of Students in BEL Trial and Reference Schools, 1998–1999**

<b>SCHOOL</b>	<b>GRADE K</b>	<b>%</b>	<b>GRADE 1</b>	<b>%</b>	<b>TOTAL</b>	<b>%</b>
<b>NEW YORK</b>						
<i>Trial</i>						
PS41	121	5.98	173	7.14	294	6.61
PS68	130	6.42	188	7.76	318	7.15
PS78	134	6.62	187	7.72	321	7.22
PS96	143	7.07	166	6.85	309	6.95
PS112	105	5.19	122	5.04	227	5.11
<b>Total</b>	633	31.27	836	34.52	1469	33.04
<i>Reference</i>						
PS21	128	6.32	190	7.84	318	7.15
PS76	150	7.41	178	7.35	328	7.38
PS103	180	8.89	218	9.00	398	8.95
PS121	135	6.67	216	8.92	351	7.89
<b>Total</b>	593	29.3	802	33.11	1395	31.38
<b>BOSTON</b>						
Dante Alighieri	38	1.88	24	0.99	62	1.39
James F. Condon	79	3.90	103	4.25	182	4.09
John Eliot	30	1.48	28	1.16	58	1.30
Mary Lyon	10	0.49	16	0.66	26	0.58
Wolfgang A. Mozart	51	2.52	30	1.24	81	1.82
Michael J. Perkins	65	3.21	23	0.95	88	1.98
Franklin D. Roosevelt	81	4.00	48	1.98	129	2.90
Pauline A. Shaw	54	2.67	67	2.77	121	2.72
Lyman F. Winship	36	1.78	26	1.07	62	1.39
John P. Holland	80	3.95	127	5.24	207	4.66
<b>Total Boston</b>	524	25.89	492	20.31	1016	22.85
<b>ELGIN</b>						
Garfield	50	2.47	57	2.35	107	2.41
Highland	50	2.47	67	2.77	117	2.63
Huff	66	3.26	68	2.81	134	3.01
Illinois Park	49	2.42	55	2.27	104	2.34
Sheridan	59	2.92	45	1.86	104	2.34
<b>Total Elgin</b>	274	13.54	292	12.06	566	12.73
<b>TOTAL</b>	2024	100.00	2422	100.00	4446	100.00

While the control schools were selected because they had demographic profiles that were similar to the trial schools, there were some minor differences in the demographic profiles of the achieved sample of students in the control and trial schools. The demographic characteristics of the two groups of schools are summarized in Table 3.

**Table 3. Demographic Characteristics of the Achieved Sample of Control and Trial School Students**

<b>Kindergarten</b>	<b>Control</b>	<b>Trial</b>
Non-English Speakers	9.5	14.1
Free and Reduced Lunch	88.1	81.2

<b>Grade 1</b>	<b>Control</b>	<b>Trial</b>
Non-English Speakers	5.9	15.0
Free and Reduced Lunch	85.3	83.7

Table 3 indicates a higher percentage of students from non-English speaking backgrounds in the trial schools and a slightly higher percentage of students in receipt of free or reduced lunch in the control schools.

**Data Collection**

A total of 4,446 students initially participated in the first year of the evaluation study. Of these, 3,051 students were in the 20 trial schools. In four control schools, 1,395 of these students did not receive the BEL/*Bookshop* treatment. In the second year of the study, a further eight schools – five schools in Boston and three in the Bronx – joined the project involving Grades K and 1. In the second year, testing was carried out in the trial schools only. For the first cohort of trial schools (those that had participated in the first year of the project), testing and treatment was extended to involve Grades K–2.

In the first year of the study, administration of the pretest measures to both control and treatment (trial) students was conducted in the fall of 1998. Posttesting occurred in late May and early June of 1999. Staff from New York University supervised all testing activities and managed the assessment activities of a team of field evaluators for the pre- and posttesting during the first year of the study.

In the control schools, teams of external testers administered all assessments. The assessments used were identical to those used in the trial schools. New York University researchers closely monitored the procedures. Teachers in the treatment (trial) schools administered the assessments to their own students, and these assessments were then audited using a randomized retesting process (see p.9).

---

Students in Grades K and 1 in BEL schools were administered all six of the measures comprising *An Observation Survey of Early Literacy Achievement* and *The Record of Oral Language: Biks and Gutches*<sup>9</sup>. These measures are listed below:

*An Observation Survey of Early Literacy*

1. Reading Record Text Level
2. Letter Identification
3. Concepts About Print
4. Word Test
5. Writing Vocabulary
6. Dictation Task

*The Record of Oral Language: Biks and Gutches*  
Oral Language Leveled Sentences

Students in Grade 2 with a Reading Record text levels greater than 15 were administered the *Gentry Grade Spelling Test*<sup>10</sup> and the *Peters Spelling in Context*<sup>11</sup> in addition to ongoing measurement of their text levels, but did not take the other components of the Observation Survey or the Record of Oral Language.

This report focuses on certain elements of the data collection: Reading Records, Letter Identification, Concepts About Print, Word Test, Writing Vocabulary, and the Dictation Task. The other three measures, Record of Oral Language, Gentry, and Peters in Context, were all used to inform instruction, but for the purposes of this report, results for these measures are not reported.

To supplement measures of students' literacy progress, information on the following student background characteristics were collected:

- **GENDER** (coded 0 = male; 1 = female)
- **LBOTE** (whether the student was from a language background other than English)
- **FREE LUNCH** (whether the student was in receipt of free lunches)
- **SPECIAL EDUCATION** (whether the student was designated as a special education student)
- **CRITICAL EVENT** (whether the student is identified by the teacher as having experienced a critical event such as the death of a close relative or an extended illness, considered to have adversely affected the student's academic progress during the year)

All data were forwarded to New York University for checking and collating prior to being sent to The University of Melbourne for keying and analysis.

<sup>9</sup> Clay, Marie M. et. al. *Record of Oral Language and Biks and Gutches*. Portsmouth, NH: Heinemann, 2001.

<sup>10</sup> Gentry, R.J. *My Kid Can't Spell*. Portsmouth, NH: 1997.

<sup>11</sup> Peters, Margaret and Smith, Bridig. *Spelling in Context*. Windsor, UK: NFER-Nelson, 1993.

---

## Data Analysis

As with all studies of this kind, it was necessary to make decisions about the handling of missing data. The approach taken was to exclude all cases with missing data, rather than seek to impute values. This resulted in considerable attrition, especially in the control schools, as indicated in Table 4.

The high attrition rate in the control schools was largely due to failure to gain teacher cooperation or failure to obtain parental consent forms for students to be involved in the testing program. This was a significant problem because it tended to affect whole classes of students. In one of the control schools, there were two classes with no data but many students without parental consent forms. In another of the control schools there were four classes with almost no data. In the trial schools, there was an attrition rate of around 20%, some of which reflected changes in enrollment over the year and some of which represented missing demographic data.

**Table 4. Sample Attrition in Control and Trial Schools**

<b>Kindergarten</b>	<b>Enrollment</b>	<b>Complete Cases</b>	<b>% Enrollment</b>
Control	593	243	41.1
Trial	1431	1148	80.2

<b>Grade 1</b>	<b>Enrollment</b>	<b>Complete Cases</b>	<b>% Enrollment</b>
Control	802	341	42.5
Trial	1620	1284	79.2

In proceeding to analyze the data for students with complete records, the first step was to analyze the percent of students meeting the various targets described earlier and summarized in Table 1. This involved analyzing pre- and posttest scores on students' text level. The second step was to make use of composite scores comprising all the outcome measures to analyze the impact of participation in BEL/*Bookshop* treatment adjusting for prior achievement and student background characteristics.

## Reliability of Measures

Each of the measures of literacy used in BEL was an individually administered teacher observation of student performance. The assessments require a considerable amount of time to administer, but in the process, the observer can learn a great deal. In the interest of objectivity, it is better that testing be done by a third party and not by the classroom teacher. But in the interest of improving teacher knowledge of students, it is preferable that classroom teachers undertake the assessment. In initiating BEL, a decision was made not to involve external persons in testing

students except in the control schools that were included in the first year of the project (1998–1999). However, it was also decided that each year there would be retesting by independent assessors of a small sample of student results from each class in the trail schools in order to check the quality of assessment data provided by classroom teachers.

Table 5 summarizes the results of retesting undertaken at the end of 1998–1999 and the beginning of 1999–2000. The table indicates inter-rater reliability of 0.8 or above for all measures except Concepts About Print. The mean scores of classroom teachers and independent assessors are also very comparable, with differences largely accounted for by the fact that retesting typically took place some three weeks after the initial testing, explaining the slightly higher scores of the independent assessors.

**Table 5. Results of Retesting, Including Inter-rater Reliability**

TEST	INTER-RATER RELIABILITY		MEAN SCORES (PRE 1999–2000)					
	POST 98–99	PRE 99–00	GRADE K		GRADE 1		GRADE 2	
			SCHOOL	AUDIT	SCHOOL	AUDIT	SCHOOL	AUDIT
ORAL	—	0.89	19.0	17.1	20.7	21.2	22.9	24.0
TEXT	0.93	0.91	0.1	0.2	2.0	3.5	10.0	13.0
LETT	0.93	0.87	30.8	36.9	45.7	48.8	52.4	53.1
CAP	0.75	0.74	9.0	10.8	12.5	14.8	18.1	20.2
WORD	—	0.91	0.5	1.4	4.0	6.0	14.4	16.5
WRIT	—	0.81	4.1	5.8	10.5	16.7	27.6	38.2
DICT	0.71	0.80	5.0	8.4	13.9	21.0	28.8	31.9
GENT	—	0.87	—	—	—	—	44.3	50.2
PETER	—	0.90	—	—	—	—	63.3	68.0

## Performance Against Standards and Targets

As mentioned earlier, at the commencement of BEL, a set of standards and associated targets was established with respect to the reading achievements of students in the first three years of school. These are summarized in Table 1.

The progress of BEL schools towards meeting these targets is summarized in Table 6. This table shows the percent of students at or above text levels 1, 5, 15 and 20. The percentages relate to the proportions of students able to read unseen texts with 90% or better accuracy. A set of 28 texts was used that had been carefully leveled using Reading Recovery text<sup>12</sup> levels for the first 26 levels. Levels 27 and 28 were devised using the Noun Frequency Method<sup>13</sup> as a guide to giving a readability measure to assign more sophisticated text gradients for the purposes of tracking change

<sup>12</sup> Peterson, Barbara (1991). "Selecting Books for Beginning Readers." In D.E. Deford, C.A. Lyons, and G.S. Pinnell (eds.) *Bridges to Literacy: Learning from Reading Recovery* (pp. 119–147). Portsmouth, NH: Heinemann.

<sup>13</sup> Elley, W., & Croft, C. (1991). *Assessing the difficulty of reading materials: The noun frequency method*. Wellington, NZ: New Zealand Council for Educational Research.



over time. The percentages reported in Table 6 are not for all students assessed, but only for those students with complete pre- and posttest data. Results are shown for all schools and for schools in each of the three districts participating in BEL in 1998–2000 (Intake 1). In the second year of BEL, a second intake of schools (Intake 2) in Boston and New York joined the project. Results for this second intake are recorded separately.

In the first year of implementation, Kindergarten students moved from just under 5 percent of students underway in reading (minimum standard of text level 1+) to just under 66 percent by the end of the first year, and over 73 percent by the end of the second year of implementation. Grade 1 students moved from half a percent at minimum standard of text level 15 and above at the beginning of the first year of implementation to almost 48 percent at the end of the first year and to over 59 percent by the end of the second year of implementation. By the end of the second year, 87 percent of Grade 2 students were reading at a text level of 15+, and 82 percent at text level 20 and above.

**Table 6. Percent of Students At or Above Nominated Text Levels: BEL 1998–2000**

GRADE/ INTAKE	YEAR	SITE	N	1+		5+		15+		20+	
				PRE	POST	PRE	POST	PRE	POST	PRE	POST
GRADE K Intake 1	98–99	All	1148	4.9	65.5	0.9	26.2	0.2	5.4	0.1	3.4
		New York	505	2.8	58.0	0.6	23.2	0.0	4.2	0.0	2.8
		Boston	427	5.6	69.8	1.4	29.0	0.2	7.7	0.2	4.9
		Elgin	216	8.3	74.5	0.5	27.8	0.5	3.7	0.0	1.9
	99–00	All	1145	5.5	73.4	1.6	31.1	0.5	7.9	0.3	5.7
		New York	570	6.7	75.1	2.3	34.6	0.5	9.8	0.2	7.4
		Boston	360	5.3	71.7	1.1	32.2	0.3	8.1	0.3	5.6
		Elgin	215	2.8	72.1	0.5	20.0	0.5	2.3	0.5	1.4
GRADE K Intake 2	99–00	All	500	7.4	70.6	2.8	37.0	1.2	13.4	1.0	10.2
		New York	353	9.6	75.4	3.7	45.9	1.4	18.1	1.1	13.9
		Boston	147	2.0	59.2	0.7	15.6	0.7	2.0	0.7	1.4
GRADE 1 Intake 1	98–99	All	1284	39.9	95.9	8.9	84.6	0.5	47.7	0.5	34.3
		New York	637	36.1	94.5	10.8	81.5	0.8	43.6	0.8	31.9
		Boston	416	35.3	95.7	5.0	82.0	0.0	48.6	0.0	34.6
		Elgin	231	58.4	99.6	10.4	97.8	0.4	57.1	0.4	40.7
	99–00	All	1359	55.6	97.4	24.4	90.3	7.2	59.3	5.3	44.7
		New York	709	51.6	97.0	23.0	89.6	6.3	56.3	4.8	43.0
		Boston	440	59.3	96.8	26.4	88.4	9.8	63.0	7.0	46.6
		Elgin	210	61.4	99.5	24.8	96.7	4.8	62.9	3.3	46.7
GRADE 1 Intake 2	99–00	All	624	48.7	88.9	20.2	25.2	6.4	43.8	5.0	33.8
		New York	466	50.2	86.5	22.7	73.2	8.4	44.8	6.4	35.0
		Boston	158	44.3	96.2	12.7	81.0	0.6	40.5	0.6	30.4
GRADE 2 Intake 1	99–00	All	1361	94.5	98.2	88.0	95.9	55.5	87.1	43.8	82.0
		New York	735	91.3	97.0	83.5	93.9	51.3	84.6	42.9	80.5
		Boston	414	97.6	99.5	92.5	98.1	61.1	88.9	45.2	81.6
		Elgin	212	99.5	99.5	94.3	98.6	59.0	93.4	44.3	87.7

Table 7 presents comparable information according to certain groups of students. It is evident that the levels of students from language backgrounds other than English generally lag behind levels of “All” students in the Kindergarten grade, but that they quickly catch up in their second year of schooling. By Grade 2, the differences are negligible.

In the case of students who receive of free lunches, Table 7 indicates that students in the free lunch program perform at a slightly lower level than “All” students, particularly with respect to the higher-level standards (15+ and 20+). Special education students understandably perform significantly less well than all students. They continue to make progress, however, as is reflected in the results for the Grade 2 Intake 1 special education students, 70 percent of whom are performing at Text Level 15 or above by the end of the year.

**Table 7. Percent of Students At or Above Nominated Text Levels: BEL 1998–2000, Students from Language Backgrounds Other Than English, Students in Receipt of Free Lunches, and Special Education Students**

GRADE / INTAKE	YEAR	SITE	N	1+		5+		15+		20+	
				PRE	POST	PRE	POST	PRE	POST	PRE	POST
GRADE K Intake 1	98–99	All	1148	4.9	65.5	0.9	26.2	0.2	5.4	0.1	3.4
		LBOTE	162	2.5	59.9	0.6	19.1	0.0	4.3	0.0	1.9
		Free Lunch	672	2.1	60.7	0.3	22.0	0.0	3.6	0.0	2.4
		Special	74	2.7	43.2	0.0	13.5	0.0	2.7	0.0	1.4
	99–00	All	1145	5.5	73.4	1.6	31.1	0.5	7.9	0.3	5.7
		LBOTE	223	0.4	58.7	0.4	18.8	0.0	2.7	0.0	1.8
		Free Lunch	404	3.2	65.3	1.0	22.5	0.0	4.5	0.0	3.5
		Special	107	6.5	66.4	2.8	19.6	0.9	4.7	0.9	1.9
Grade 1 Intake 1	98–99	All	1284	39.9	95.9	8.9	84.6	0.5	47.7	0.5	34.3
		LBOTE	193	34.7	96.4	6.7	82.4	0.5	48.7	0.0	34.2
		Free Lunch	783	34.1	95.7	7.5	82.0	0.4	42.1	0.3	28.7
		Special	142	27.5	87.3	4.2	59.2	0.0	21.8	0.0	22.7
	99–00	All	1359	55.6	97.4	24.4	90.3	7.2	59.3	5.3	44.7
		LBOTE	149	49.7	88.9	20.2	25.2	6.4	43.8	5.0	33.8
		Free Lunch	500	50.2	86.5	22.7	73.2	8.4	44.8	6.4	35.0
		Special	203	44.3	96.2	12.7	81.0	0.6	40.5	0.6	30.4
Grade 2 Intake 1	99–00	All	1361	94.5	98.2	88.0	95.9	55.5	87.1	43.8	82.0
		LBOTE	156	92.3	98.1	82.7	96.8	60.3	85.9	46.8	80.8
		Free Lunch	580	92.4	97.2	82.8	94.0	46.6	79.5	34.0	72.4
		Special	206	84.0	97.6	72.3	89.3	36.9	70.4	26.7	63.6

---

## Factors Affecting Student Achievement

Analyses were undertaken to ascertain the effects of different factors on student progress during the first year of implementation.<sup>14</sup> The approach was to estimate the proportion of variation in posttest literacy scores due to the following factors:

- Prior achievement (a composite score made up of student scores on the seven pretest measures)
- Student characteristics (gender, language background, free lunch, special education, critical event)
- School (the school attended by the student)
- Class (the class to which the student had been assigned within the school)
- Residual variation (unmeasured factors, such as student attendance, whether the student had poor hearing or vision, how the student was feeling on the day of testing, measurement error in the tests used to assess achievement, etc.)

The results are summarized in Table 8 for both the Kindergarten and Grade 1 students. From the table, it is clear that it was possible to account for 62% of the variation in Kindergarten student performance and 68.3% of Grade 1 student performance.

The most important factor explaining student performance at the end of Kindergarten and Grade 1 was prior student achievement. In other words, the best predictor of student progress was their starting points. Measures of prior achievement, of course, embody the impact of student characteristics. The proportion of variation in outcome measures accounted for by student characteristics over and above that already influencing scores on the prior achievement measure was relatively small (4.8% at Kindergarten), although this proportion was significantly higher for Grade 1 students (14.7%). The results indicates that it is more helpful to focus on the starting points of individual students than to focus on group characteristics.

**Table 8. Proportion of Variance in BEL/Bookshop Student Posttest Results Accounted For by Various Factors**

<b>FACTOR</b>	<b>GRADE K</b>	<b>GRADE 1</b>
Prior achievement (pretest)	41.2	37.8
Student characteristics	4.8	14.7
School	6.6	7.1
Class	9.3	8.7
<b>Total explained variation</b>	<b>61.9</b>	<b>68.3</b>
Unexplained (residual) variation	38.1	31.7

<sup>14</sup> This involved fitting a series of three-level (student/class/school) multi-level regression models and partitioning the variation accounted for by the fixed and random parts of each model.

---

Student characteristics and prior achievement are factors that represent the educational givens, or unalterable factors. Differences due to the school attended by the student and the class to which the student was assigned represent the alterable factors, or the set of decisions made with respect to each student's education. Taken together, they account for a substantial proportion of the total variation – certainly more important than that accounted for by student characteristics. They indicate that educators are by no means powerless in affecting student progress. Schools can make a difference when effective teaching takes place in each classroom.

## **Measuring the Impact of BEL/*Bookshop***

To evaluate the impact of BEL/*Bookshop* treatment during its first full year of implementation, a series of multi-level regression models was fitted to the data of students in both the control and trial schools participating in the first year of the project.

In some exploratory analyses, a number of school-level variables were computed and included in the models for predicting student outcomes, including average levels of performance of students and percent of students in various demographic groups, such as free and reduced lunch. In none of these analyses were the school-level variables significant, so they were excluded from further modeling of the data.

Table 9 presents unadjusted and adjusted means and standard deviations for total literacy scores. The term *adjusted means* is used to refer to average scores that have been adjusted statistically for the effects of the measured student intake characteristics and the prior achievement of each student. The final column thus provides the best indication of the impact of BEL, with Mondo *Bookshop* Reading Program, on student progress. It will be noted that at the beginning of the 1998–1999 school year, both the Kindergarten and Grade 1 students in the control schools started with slightly higher scores than did students in the trial schools. By the end of the year, however, students in the treatment (trial) schools were performing at more than half a standard deviation higher than were students in the control schools.

The superior rates of progress of students in the trial schools is captured in the adjusted posttest means, which represent the progress made by students, adjusting for initial differences in background characteristics and prior achievement. The adjusted posttest mean of Kindergarten students in the treatment (trial) schools was 14.0 compared to 8.6 in the control schools, while the posttest mean of Grade 1 students in the trial schools was 26.4 compared to 22.9 in the control schools. These results are summarized in Figure 3.

Table 9. BEL/Bookshop: Pre- and Posttest Literacy Scores

Measure	UNADJUSTED PRETEST		UNADJUSTED POSTTEST		ADJUSTED POSTTEST	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
<b>KINDERGARTEN</b> Control	5.55	3.63	11.60	5.79	8.59	3.12
	3.95	3.06	14.50	7.41	14.04	3.32
<b>GRADE 1</b> Control	10.70	5.09	23.30	6.84	22.91	4.13
	9.97	5.55	25.70	7.75	26.38	3.88

Note: *Adjusted posttest means were calculated by fitting a three-level model to the raw posttest scores (as opposed to normalized scores for all students), using a dummy variable to estimate the effect of being a trial school and adjusting for the effects of gender, non-English speaking background, free lunch, special education, critical event, and prior achievement. Student-level standard deviations were estimated by modeling level 1 variance using dummy variables to provide separate estimates for control and trial schools after the method described in section 1.13.1 of Woodhouse, Rasbash, Goldstein and Yang (1996).<sup>15</sup>*

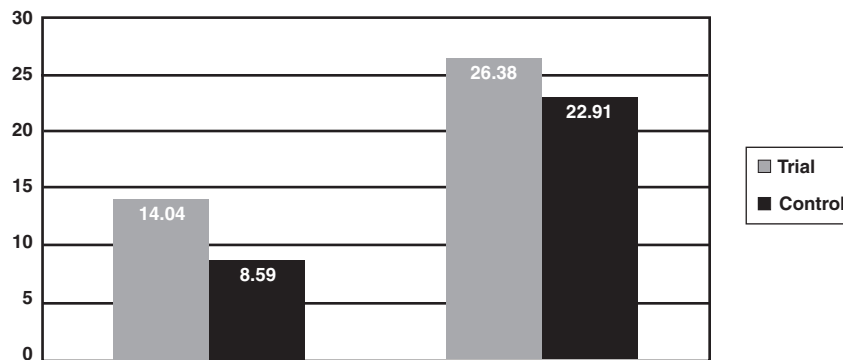


Figure 3. Adjusted Posttest Scores for Students in BEL Trial and Control Schools, 1998–1999

Analyses of the BEL data for the first year of implementation yielded effect size estimates of 1.19 for the effect of the program on Kindergarten students, and 1.00 for the effect of the program on Grade 1 students.<sup>16</sup> Parameter estimates for the models fitted to the Kindergarten and Grade 1 data are reported in the Appendix, and Tables A1 and A2. These represent very substantial effects relative to those typically encountered in the research literature.

<sup>15</sup> Woodhouse, G., Rasbash, J., Goldstein, H. and Yang, M. (1996). "Introduction to multilevel modeling." In G. Woodhouse (ed.) *Multilevel modeling applications: A guide for users of MLN* (pp. 9–57). London: University of London Institute of Education.

<sup>16</sup> Effect sizes were calculated by dividing the dummy coefficient representing the effect of BEL by the square root of the residual variance at the student level of a three-level regression model in which the fixed factors were gender, language background, free lunches, special education, and critical events, and in which the random factors were school, class, and student. For further details of the logic of this approach, see Tymms, P., Merrell, C., & Henderson, B. (1995). *The first year at school: A quantitative investigation of the attainment and progress of pupils. Educational Research and Evaluation*, 3, 101–118.

---

At the end of the second year of implementation, further analyses were undertaken to compare the progress made by Kindergarten students in Intake 1 and Intake 2 schools in both the 1998–1999 and 1999–2000 school years.<sup>12</sup> The following effect sizes were obtained (See Table A3 of the Appendix).

Intake 1 treatment/trial schools, 1998–1999	0.89
Intake 1 treatment/trial schools, 1999–2000	0.98
Intake 2 treatment/trial schools, 1999–2000	0.78

The effect size estimate for Intake 1 treatment (trial) schools in 1998–1999 differs from that reported earlier because the analysis was carried out on a different sample of schools. Nonetheless, it can be seen that in all three cases, effect sizes were large. Furthermore, Intake 1 treatment (trial) schools were more effective in their second year, indicating that it is likely that schools need several years before they are performing at their full potential.

## The Impact of Reading Recovery

A significant cost component of BEL is the implementation of Reading Recovery, a one-on-one intervention program for the lowest achieving students in Grade 1. An analysis was undertaken to estimate the effect of Reading Recovery on the overall performance of Grade 1 students in the first year of implementation.

The analysis indicated that when Reading Recovery was ignored, BEL had a 1.00 standard deviation effect on the progress of students in Grade 1. When Reading Recovery was taken into account, the estimate of the effect size for BEL reduced slightly to 0.91, but the estimate for Reading Recovery was 0.27, which is a medium-sized effect. In 1998–1999, a total of 313 students were provided with Reading Recovery. The data thus suggest that the intervention provided to these students accounted for about 30 percent of the total impact of BEL. Furthermore, the impact of Reading Recovery was to supplement the remainder of the program.

## Conclusion

This initial evaluation study of the impact of BEL and the Mondo *Bookshop* Reading Program indicates stronger positive effects than were achieved in earlier implementations in the two Australian studies quoted at the beginning of this report, indicating that the design was highly effective in American inner-city urban contexts. The recorded effect sizes at the end of the first year were very large indeed, as were effect sizes for the two intakes implementing the project in the second year of the study.

Sustaining the momentum for large-scale reform requires high levels of commitment and effort on the behalf of administrators and teachers. Achieving early and substantial gains is important in providing encouragement for them to persevere and institutionalize change. The results at the end of the first two years of this study provide strong vindication of their efforts and incentive to continue.

<sup>12</sup> No analysis was undertaken comparing the performance of Grade 1 students in the second year of the study with that of Grade 1 students in the control schools in the first year of the study because the former group had had the benefit of participation in the program in Kindergarten, making any comparison with those in the control schools unfair.

---

## References

- Cairney, T. H., J. Ruge, J. Buchanan, K. Lowe, & L. Munsie (1995). *Developing Partnerships: The home, school and community interface*. Canberra: Department of Employment, Education and Training.
- Chrispeels, J. (1996). Effective Schools and Home-School-Community Partnership Roles: A framework for parent involvement. *School Effectiveness and School Improvement*, 7, 297–323.
- Clay, M. M. & V. Tuck (1991). *A study of Reading Recovery subgroups: Including outcomes for children who did not satisfy discontinuing criteria*. Auckland, New Zealand: Ministry of Education.
- Clay, M. M. (1993b). *Reading Recovery: A Guidebook for Teachers in Training*. Portsmouth, NH: Heinemann.
- Clay, M. M. (1993a). *An Observation Survey of Early Literacy Achievement*. Portsmouth, NH: Heinemann.
- Crévola, C. A. & P. W. Hill (1998). Evaluation of a whole-school approach to prevention and intervention in early literacy. *Journal of Education for Students Placed at Risk*, 3, 133-157;
- Elley, W. & C. Croft (1991). *Assessing the difficulty of reading materials: The noun frequency method*. Wellington, NZ: New Zealand Council for Educational Research.
- Epstein, J. L. (1995). School/Family/Community Partnerships: Caring for children we share. *Phi Delta Kappan*, 76, 701–712.
- Hill, P. W., & C. A. Crévola (1997). The literacy challenge in Australian primary schools. *IARTV Seminar Series*, No. 69, November, 1997.
- Hill, P. W., & C. A. Crévola (1999a). The role of standards in education reform for the 21st century. In David D. Marsh (Ed.), *ASCD Yearbook 1999: Preparing our Schools for the 21st Century*. (117–142) Alexandria, VA: Association for Supervision and Curriculum Development.
- Hill, P. W. & K. J. Rowe (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement*, 7 (1), 1–34.
- Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Education Evaluation and Policy Analysis*, 14, 307–332.
- Scheerens, J., C. J. A. J. Vermeulen, & W. J. Pelgrum (1989). Generalizability of instructional and school effectiveness indicators across nations. *International Journal of Educational Research*, 13 (7), 789–799.
- Tymms, P., C. Merrell, & B. Henderson (1995). The first year at school: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3, 101–118.
- Wasik, B. A. & R. E. Slavin (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 179–200.

## APPENDIX

**Table A1. Parameter Estimates for Three Multilevel Regression Models of Grade K Students' Literacy Progress, 1998–1999 (1,391 students, 79 classes, 24 schools)**

PARAMETER	MODEL 1		MODEL 2		MODEL 3	
	Est.	s.e	Est.	s.e	Est.	s.e
<b>Fixed</b>						
Intercept	-0.080	0.113	-0.100	0.104	-0.706	0.162
Gender (female)	0.188	0.041	0.061	0.029	0.061	0.029
NESB	-0.143	0.075	0.199	0.054	0.202	0.054
Free Lunch	-0.222	0.066	0.031	0.047	0.039	0.047
Special Education	-0.674	0.125	-0.311	0.091	-0.316	0.091
Critical Event	-0.424	0.123	-0.282	0.092	-0.269	0.091
Prior Achievement	—	—	0.689	0.018	0.690	0.018
BEL	—	—	—	—	0.718	0.164
<b>Random</b>						
School	0.125	0.051	0.157	0.057	0.070	0.032
Class	0.101	0.028	0.087	0.021	0.087	0.021
Student	0.742	0.029	0.361	0.014	0.361	0.014
-2*log(lh)	3658		2697		2683	

**Table A2. Parameter Estimates for Three Multilevel Regression Models of Grade 1 Students' Literacy Progress, 1998–99(1,625 students, 97 classes, 24 schools)**

PARAMETER	MODEL 1		MODEL 2		MODEL 3	
	Est.	s.e	Est.	s.e	Est.	s.e
<b>Fixed</b>						
Intercept	0.181	0.106	0.124	0.088	-0.358	0.146
Gender	0.162	0.039	0.088	0.030	0.088	0.030
NESB	-0.032	0.066	0.029	0.051	0.026	0.051
Free Lunch	-0.207	0.056	-0.102	0.044	-0.099	0.044
Special Education	-0.715	0.086	-0.375	0.067	-0.376	0.067
Critical Event	-0.511	0.116	-0.334	0.090	-0.327	0.090
Prior Achievement	—	—	0.587	0.026	0.590	0.026
BEL	—	—	—	—	0.587	0.158
<b>Random</b>						
School	0.137	0.065	0.117	0.046	0.057	0.028
Class						
Intercept	0.243	0.047	0.096	0.021	0.096	0.021
Slope/Intercept	—	—	0.000	0.010	-0.003	0.009
Slope	—	—	0.028	0.008	0.029	0.008
Student	0.588	0.021	0.342	0.013	0.342	0.013
-2*log(lh)	3963		3131		3120	



**Table A3. Parameter Estimates for Three Multilevel Regression Models of Grade K Students' Literacy Progress 1998–2000 (2,971 students, 147 classes, 30 schools)**

PARAMETER	MODEL 1		MODEL 2		MODEL 3	
	Est.	s.e	Est.	s.e	Est.	s.e
<b>Fixed</b>						
Intercept	0.165	<i>0.104</i>	-0.160	<i>0.077</i>	-0.707	<i>0.114</i>
Gender (Female)	0.165	<i>0.029</i>	0.072	<i>0.022</i>	0.063	<i>0.021</i>
NESB	-0.108	<i>0.058</i>	0.181	<i>0.044</i>	0.195	<i>0.044</i>
Free Lunch	-0.127	<i>0.041</i>	0.015	<i>0.031</i>	0.048	<i>0.033</i>
Special Education	-0.908	<i>0.083</i>	-0.431	<i>0.063</i>	-0.452	<i>0.063</i>
Critical Event	-0.382	<i>0.129</i>	-0.283	<i>0.096</i>	-0.291	<i>0.095</i>
Prior Achievement	—	—	0.653	<i>0.014</i>	0.658	<i>0.022</i>
T9899	—	—	—	—	0.608	<i>0.108</i>
T19900	—	—	—	—	0.671	<i>0.111</i>
T29900	—	—	—	—	0.529	<i>0.120</i>
<b>Random</b>						
School	0.220	<i>0.060</i>	0.121	<i>0.033</i>	0.072	<i>0.020</i>
Slope/Intercept	—	—	—	—	-0.012	<i>0.006</i>
Slope	—	—	—	—	0.008	<i>0.003</i>
Student	0.844	<i>0.022</i>	0.474	<i>0.012</i>	0.342	<i>0.013</i>
-2*log(lh)	8021		6307		6248	

**PETER W. HILL** is Deputy Dean, Head of the Department of Education Policy and Management and Professor of Education (Leadership and Management) in the Faculty of Education of The University of Melbourne, Director of the Centre for Applied Educational Research, and Chair of the Board of Directors of the Australian Principals Centre. He was formerly Chair of the Victorian Curriculum and Assessment Board and Chief General Manager of the Department of School Education in Victoria.

**ANGELA M. JAGGAR** is Professor of Education (ret.) in the School of Education at New York University, where she taught courses and conducted research on children's language and literacy development for over 30 years. She was co-founder of the NYU Reading Recovery Project and served as co-director for 10 years. She also served as president of National Council on Research in Language and Literacy.

---

**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 Pennsylvania<sup>1</sup>

---

<sup>1</sup> 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](http://www.mondopub.com)

980 Avenue of the Americas

New York, NY 10018