January, 2008
International Association of School Libraries Annual Conference Proceedings
DOI: 10.29173/iasl7962

Is The Library Important? Multivariate Studies at the National and International Level

Stephen Krashen
Professor Emeritus
University of Southern California
USA

Syying Lee
Professor
National Taipei University
Taiwan

Jeff. McQuillan
Senior Research Associate
Center for Educational Development
USA

Three multivariate analyses, all controlling for the effects of poverty, confirm the importance of the library. Replicating McQuillan's analysis of 1992 NAEP scores, access to books in school and public libraries was a significant predictor of 2007 fourth grade NAEP reading scores, as well as the difference between grade 4 and grade 8 2007 NAEP reading scores, suggesting that access is important for improvement after grade 4. Access (school/classroom libraries) was a significant predictor of scores on the PIRLS test, a reading test given to fourth graders in 40 countries.

It has been firmly established that more reading leads to better reading (and writing, spelling, vocabulary and grammar), and that more access to books results in more reading (Krashen, 2004).

It is thus reasonable to hypothesize that more access means better reading. This prediction has been confirmed by a number of studies showing a positive relationship between library quality and reading achievement (McQuillan, 1998; Lance, 2004, and studies reviewed in Krashen, 2004).

In a multivariate study, McQuillan (1998) examined the relationship between access to reading material and scores on the 1992 NAEP reading test given to samples of fourth graders in 42 states in the US. His measure of access was a combination of three measures of access to reading material at home, two of access to reading in school, and two of access to reading in the community. Table 1, a multiple regression analysis from McQuillan (1998), tells us that even after controlling for the effect of poverty, access to print was a significant and strong predictor of performance on the NAEP: Those with more access did better.

The combination of poverty and print access accounted for 72% (r2 = .72) of the variability on the NAEP, that is, if we know the level of poverty of families in a state, and how much reading material is available to children in that state, we have 72% of the information we need to predict how well fourth graders in that state scored on the NAEP.

Table 1: Predictors of NAEP grade 4, 1992, 42 states

	beta	t	p
Poverty	-0.45	-5.07	0
Print Access	1.12	4.3	0

r2 = .72
From: McQuillan, 1998

The goal of this paper is to report some recent progress in this area, using multivariate analysis.

A Replication

Table 2 presents a replication of McQuillan's findings using the 2007 fourth grade NAEP and more recent measures of poverty and access to books (a combination of books per student in school libraries and per capita total circulation in public libraries in each state). (Means, standard deviations, and inter-correlations among the variables are presented in the Appendix, tables A1 and A2.) This analysis also controls for the presence of English learners by only including scores for fluent English proficient children.[1] Once again poverty is a strong predictor of scores, and once again access to books makes an independent contribution to reading achievement.

Table 2: Predictors of NAEP grade 4, 2007, 51 states

	b	beta	t	p
Poverty	-0.919	0.72	7.42	0
Access	0.658	0.53	1.62	0.055
r2 = .6468				

r2 = .65 adjusted r2 = .63
Fluent English proficient students only

The Grade 4 to 8 Difference

A separate analysis was performed to try to determine what factors are responsible for improvement after grade 4, or, more accurately in this case, the difference between grade 4 and grade 8 scores. This multiple regression analysis is presented in table 3. This analysis indicates that, not surprisingly, that grade 4 scores are a strong predictor of grade 8 scores. It is surprising, however, that poverty is a weak predictor of the difference between grade 4 and grade 8. Recall that the impact of poverty is strong, however, on the grade 4 test.

Of interest to us is that access to books, again a combination of school library holdings and public library circulation, is a significant predictor of the difference in NAEP reading scores between grade 4 and grade 8.

The r2 of .89 means that knowing the fourth grade NAEP scores for a state, the level of poverty, school library holdings and public library circulation is 89% of what we need to predict a state's grade 8 NAEP reading score.

Table 3: Predictors of NAEP grade 8, 2007, 51 states

	b	beta	t	p
NAEP4	-0.848	-0.857	10.68	0
Poverty	0.0958	0.076	0.96	0.17
Access	1.05	0.126	4.59	0

R2 = .89, adjusted r 2 = .89
Fluent English proficient only

Late intervention

The effect of poverty on fourth grade reading is enormous, but access to books can contribute to fourth grade reading, regardless of poverty. The analysis also indicates that those who read better in grade four also read better in grade eight, but access to books can help here as well. This agrees with data showing that "late intervention" in the form of recreational reading is possible and effective (Krashen and McQuillan, 2007).

To get a more precise idea of the impact of access to books, we can analyze the increase in r2 achieved by adding access to the effect of poverty. In grade 4, after controlling for poverty, access adds .02 to the r2, increasing our ability to predict reading scores by 2%. Access increases our ability to predict the grade 4 to 8 difference by nearly 5%. As indicated in table 4, both public library circulation and school library holdings contributed to these increases.

Table 4: Gains in r2

	access	PL	SL
grade 4	*2%	1.60%	1%
diff 4-8	*4.8%	*2.7%	*3%

* = statistically significant, $p < .10$.

This investigation used states of the USA as units. Our second study expands the investigation of the relationship of access to reading to the international level, with countries as units.

The PIRLS Study

PIRLS (Progress in International Reading Literacy Study) administered a reading test to fourth graders in 40 countries. PIRLS provides not only test scores, but also the results of an extensive questionnaire given to teachers and students, including attitudes, reading behavior outside of school, and classroom practices. PIRLS also supplies data on socio-economic class. The items on the questionnaire relevant to this study and SES statistics are presented in the Appendix (table A3).

We present here two analyses of the PIRLS data, designed to further test the impact of access to books (school libraries, classroom libraries) on scores on the PIRLS reading test. The first is a complex or full analysis that included as much of the information provided by PIRLS as possible, and the second is a simpler analysis, using only selected variables. We only included countries for which complete data was available for all factors (for a list of the countries included, see Appendix table A4).

The full (complex) analysis

In order to deal with the vast amount of information supplied by the PIRLS questionnaire, the data was factor analyzed, a statistical technique that assigns predictors into groups that behave similarly, as one factor.

Factor analysis revealed four factors: SES/home (Socio-economic status and home resources, including books in the home), Literacy (free reading of fiction, sustained silent reading in school, parental reading, parental education), Libraries (school and classroom), and Instructional Factors. (Inter-correlations are in table A5 of the Appendix and details of the factor analysis are presented in table A6 of the Appendix.)

The Library factor was the strongest predictor in the multiple regression analysis (table 4). The Literacy (free reading) factor was positively related to reading scores but did not reach statistical significance. Strangely, the SES/home factor was not a significant predictor of reading scores. The amount of formal reading instruction students received was negatively associated with reading proficiency. All factors combined accounted for 72% of the variation of PIRLS reading scores, with is very high.

Table 4: Multiple Regression: Complex (Full) Analysis

predictors	beta	t	p
SESHome	0.02	0.122	0.9
Literacy	0.164	1.343	0.19
Library	0.493	4.801	0
Instruction	-0.483	-3.454	0.002

r2 = .72

The simple analysis

In the simple analysis, one predictor was chosen to represent each factor, one that was felt to be most representative of the factor we were interested in investigating. For SES/Home, only one measure of socio-economic status was used, the Human Development Index (HDI) developed by the United Nations. The measure of literacy used was SSR (sustained silent reading), the percentage of students who read independently in school every day or almost every day in each country. The library factor was represented by the percentage of school libraries in each country with over 500 books. Instruction was represented by the average hours per week devoted to reading instruction in each country. Inter-correlations among these variables are in the Appendix, table A7).

Table 5: Multiple Regression: Simple analysis

predictor	beta	t	p
SESHOME	0.41	2.74	0.005
LITERACY	0.161	1.343	0.143
LIBRARY	0.346	2.75	.005
INSTRUCT	-0.186	1.4	0.085
r2 = .63			

The results are quite similar to the complex solution, except that SES, as measured by the HDI, is now a significant predictor.

As we did in the previous analysis of NAEP scores, we now examine the increase in r2 as a means of judging the impact of access to books. In the full, or complex analysis, SES alone accounted for 40% of the variability in reading scores. Adding access increased the r2 to 61%. In the simple analysis, poverty alone accounted for 50% of the variability in reading scores. Adding access increased the r2 to .60. The combination of poverty, literacy (SSR) and instruction produced an r2 of .55; adding access increased this to 63%. All of these increases are substantial.

Conclusion

In all of the multivariate studies considered here the library emerges as a consistent predictor of reading scores. This is remarkable, especially when we consider that the measures used are crude: library holdings, and even general circulation, in the case of public libraries.

Of course, providing access is only the first step: Even with access, some children (but surprisingly few) will not read. The research literature consistently indicates that rewards for reading are not effective (Krashen, 2003), but that read-alouds and conferencing do help. But in order for these approaches to work, the books need to be there.

But what is clear is that libraries definitely matter and they matter a lot.

Note

1. This was not possible for previous years' NAEP scores because separate scores for English learners and fluent English speakers were not available. Even though English learners who have recently arrived in the US are not required to take the NAEP, criteria for including English learners vary from state to state, and it is likely that many English learners who take the NAEP cannot show their full proficiency in reading on the test. The means for all students and for fluent English speakers only were similar for NAEP 2007: For all students, mean = 220.4, for fluent English only, mean = 222.4), but English learners are concentrated in a few states, and in these cases the scores with and without English learners are quite different: For California,

the difference was 11 points, for Nevada, 9 points, for Arizona, Oregon, New Mexico and Alaska, 6 points.

References

Krashen, S. (2004). *The power of reading.* Portsmouth: Heinemann and Westport: Libraries Unlimited.

Krashen, S. & McQuillan, J. (2007). Late intervention. *Educational Leadership* 65(2), 68-73.

Lance, K. (2004). The impact of school library media centers on academic achievement. In C. Kuhlthau (Ed.), *School Library Media Annual*. (pp. 188-197). Westport, CT: Libraries Unlimited.

Lee, J., Grigg, W. & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007–496). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C

McQuillan, J. (1998). *The literacy crisis: False claims and real solutions*. Portsmouth, NH: Heinemann Publishing Company.

Mullis, I, Martin, M., Kennedy, A. and Foy, P. (2006). *PIRLS 2006 international report.* Boston: International Study Center, Boston University.

APPENDIX

Table A1: NAEP 2007 analysis: Means and standard deviations, 51 states

	mean	sd
NAEP 8	263.4	6.69
NAEP 4	222.4	6.74
Poverty	17.75	5.28
Public library circulation.	7.52	2.82
School library Holdings	19.57	6.21

The measure of poverty used was the percentage of families with children in each state at the poverty level or below for 2005, available at hppt://www.kidscount.org., from the from the U.S. Census Bureau, American Community Survey.
Access consisted of a combination of two variables: (1) Per capita public library circulation for each state, from Chutem A. and Kroe, P. 2007. Public Libraries in the United States: Fiscal Year 2005 (NCES 2008-301). National Center for Educational Statistics, Institute of Education Science, U.S. Department of Education, Washington D.C. (2) School library holdings for each state (books per student), from Holton, B., Boe, Y., Baldridge, S., Brown, M., and Heffron, D. 2004. The Status of Public and Private School Library Media Centers in the United States. Washington D.C.: U.S. Department of Education, National Center for Educational Statistics.

Table A2: NAEP 4, 2007 analysis: Inter-correlations

	NAEP 4	Poverty	Access
NAEP8	0.92	0.72	0.64
NAEP4		0.79	0.49
Poverty			0.47

Table A3. PIRLS: Variables and means

Predictor	n	mean	sd
Gross National Income per capita	42	18458.7	14387
Gross Nat. Income: Purchasing power	40	20242.8	12081.8
Score on PIRLS reading test	45	505.9	67.91
Socio-economic status: HDI index	45	0.8803	0.089
Percent children with high early home literacy activities	43	55.98	15.37
Percent of homes with high educational resources	43	11.86	6.72
Percent of homes with 100 books or more	43	15.14	11.55
Percent with university education or higher	42	27.48	12.88
Percent of parents reading more than five hours per week	43	37.67	9.78
Percent students reading fiction outside of school everyday or nearly every day	45	34	10.55
Percent students reading nonfiction outside of school everyday or nearly every day	45	15.33	7.45
Percent students reading for fun outside of school everyday or nearly every day.	45	40.69	8.57
Teacher reads aloud to entire class daily.	45	59.5	22.24
Students read independently in school every day or almost every day	45	67.4	12.44
Students answer questions in workbooks about reading (almost) every day	45	36.33	14.15
Teacher Reports Giving Written Quiz or Test After Students Read – At Least Weekly	45	24.53	17.4
Percent of schools with school libraries	44	89.84	16.35
Percent of schools with school libraries containing more than 500 books.	44	73.64	27.4
School library has more than ten magazines.	44	25.67	22.07
Percent of students with access to classroom libraries.	45	71.49	21.76
Average number of books in classroom library	45	66.13	58.13
Average number of magazine titles in classroom library	45	3.36	1.84
Percent of students who can borrow books from classroom library to take home.	45	57.78	20.15
Percent Students Using Instructional Software to Develop Reading Skills	45	30.93	18.97
Percent Students Reading Stories or Other Texts on Computer	45	41.67	23.05
Hours per week on reading instruction	45	2.54	0.938

Table A4. PIRLS: Countries included in the analysis presented here:

Austria
Belgium (French)
Belgium (Flemmish)
Bulgaria
Canada-Alberta
Canada-British Columbia
Canada-Nova Scotia
Canada-Ontario
Canada -Quebec
Taiwan
Denmark
France
Georgia
Germany
Hong Kong SAR
Hungary
Iceland
Indonesia
Iran
Israel
Italy
Kuwait
Latvia
Lithuania
Macedonia, Rep. of
Moldova, Rep. of
Morocco
Netherlands
New Zealand
Norway
Poland
Romania
Russian Federation
Singapore
Slovak Republic
Slovenia
South Africa
Spain
Sweden
Trinidad and Tobago

PIRLS treated five provinces as separate countries, for some reason. Also, Hong Kong was included but China was not, and Flemish and French sections of Belgium were treated separately.

Table A5: PIRLS: Complex (full) factor analysis: Inter-correlations

	Read Prof	SESHome	Literacy	Library
SESHome	0.64			
Literacy	0.47	0.51		
Library	0.57	0.35	0.51	
Instruction	-0.64	-0.72	-0.18	-0.09

Table A6: PIRLS: Factor Analysis

	I. SES & Home	**II. Library***	**III. Literacy activities**	**IV. Instruction****	**Factor loadings**
1	Income				.85
2	Purchase				.88
3	HDI				.87
4	Home resource				.70
5	Home book 100				.81
6	Computer-skill				.88
7	Computer-text				.84
8		School lib exist			.94
9		School lib over 500			.92
10		School lib over 10 magazines			.62
11		Classroom lib exist			.89
12		Class-book			.74
13		Classroom lib magazines			.78
14		Classroom lib can borrow			.89
16			Readiness		.67
17			Parent education		.64
18			Parent read		.44 (.64 on Factor I)
19			FVR-Fiction		.64
20			FVR-Fun		.38 (.71 on Factor IV)
21			SSR		.65
22				Read aloud	.56 (.57 on Factor III; .30 on Factor I)
23				Instruction-test	.60 (.57 on Factor I)
24				Instruction-skill	.32 (-.64 on Factor I)
25				Instruction -reading	.09 (-.68 on Factor I)
26				Reading non-fiction	.36 (-.66 on Factor I)
α	.94	.84	.81	.79	

Some variables were not included in the multivariate analyses. For example, PIRLS reported data on hours spent on reading and writing instruction, but because of the vague description and the fact it is did not correlate with any of the other variables, it was not included. Also, among the library variables, PIRLS reported the percentage of students who reported borrowing books. This variable was omitted because it loaded on a single factor and reduced reliability.

A Principle Components Analysis extracted six factors and a Varimax Rotation produced three clear factors: SES/home, school library and classroom library.

The literacy and instruction factors were determined based on the inter-correlations among the variables and the concept each variable represented. We thus arrived at a four-factor solution, presented in table A6. Table A6 also presents the results of the reliability test of the four factors, and the alpha for each factor was satisfactorily high.

Note that read-alouds were in Factor IV and correlated highly with other instructional variables, suggesting that read-alouds were used primarily as instruction, and not for enjoyment.

All raw scores of the variables selected were then converted to z scores and were added up and averaged to arrive at composite score for the hierarchical regression analyses, presented in the text.

Table A7: PIRLS: simple analysis: Inter-correlations

	Read Prof	Poverty (HDI)	SSR	Sch. Lib.
Poverty (HDI)	0.71			
SSR	0.5	0.43		
Sch Lib	0.56	0.37	0.51	
Instruction	-0.26	-0.4	0.04	0.17

The Human Development Index is an average of three factors: education (adult literacy rates, school enrollment), life expectancy and wealth (logarithm of income); See http://hdr.undp.org/en/statistics/indices/hdi/. The UN considers high HDI to be between .8 and .95, mid to be between .5 and .79 and low to be between .34 and .49.