



Investigating the Efficacy of Learnia in New Jersey: Phase I – Post Hoc Analyses of Effect Sizes and Correlations

Jon S. Twing, Ph.D.
Executive Vice President

Michael J. Young, Ph.D.
Senior Research Scientist

Vernon Shimko
Director, Measurement Services

Dana Schmidek
Learnia Product Manager

January 15, 2010

Abstract

Post hoc data analyses were performed on matched student records containing results for *Learnia* tests used in New Jersey, the New Jersey Assessment of Knowledge and Skills (NJ ASK) and implementation status (full implementation of *Learnia* vs. not fully implemented). Results show that student performance on *Learnia* (pre to post) improved, districts who fully implemented *Learnia* showed more improvement on NJ ASK than their non-fully implemented counterparts, moderate to strong correlations exist between *Learnia* and NJ ASK. The study is limited by the lack of control over possible intervening variables.

Background

The State of New Jersey selected the Pearson product *Learnia* to help meet its goals, obligations, and commitments to its diverse base of constituents. These constituents represent teachers and educators, students and parents, the business community and policy makers. Primarily, *Learnia* was envisioned to be a tool that would assist local educators in their decision making regarding periodic and daily instructional choices—presumably with the end goal of improving instruction and hence, enhanced education. This vision and these specific requirements are outlined in Appendix A that presents the New Jersey Request for Proposal (RFP #07-X-39391, January 26, 2007, pages 28-29). To that end, the effectiveness of *Learnia* is expected to be manifest with other measures of successful improvements in student learning. While it may not be possible to isolate the effects of *Learnia* alone (due to it being intertwined with many of the other variables impacting education—student motivation, resource availability, student and school characteristics, relationships to other key initiatives, etc.) it is expected that if effectively used improved student learning should result and that there should be key metrics or indicators in this regard. As such, this first phase of a research portfolio seeks to understand, quantify and document the indicators of associated effect size and relationship with the New Jersey Assessment of Skills and Knowledge (NJ ASK) program. Such investigations, while not providing proof of “cause and effect” regarding educational improvements, will allow for a more direct understanding of *Learnia* as it is implemented functionally and will provide the opportunity for testimonials and case studies (if not the documentation of best practices) that will be undertaken in other phases of the overall investigation into the effectiveness of *Learnia* as used in New Jersey.

Introduction

This study represents a “post hoc” analysis of existing data. Data from the implementation of *Learnia*, data from the New Jersey Assessment of Skills and Knowledge (NJ ASK), and collateral information collected (primarily from *Learnia* implementation managers) regarding the effectiveness in implementation and ongoing use of *Learnia* was used in this investigation. Because this study uses data available from the assessments and collateral information it does not represent a controlled experiment and, as such, provides efficacy information but does not constitute a complete efficacy study.

Data Matching and Cleanup

The *Test, Measurement and Research Services* group (TMRS) of Pearson was able to match student records from the *Learnia* and NJ ASK files using varying combinations of 1) state student ID, 2) district student ID, and 3) student name (first, middle initial, and last). The result of these matches was a SAS system data file across all of the grades that included NJ ASK information such as district and school identification, demographics, and assessment results for mathematics, reading, and science, together with *Learnia* data that included percent correct scores on the mathematics and reading tests. Given this file, student records with NJ ASK results that were labeled as either “void” or “invalid” were eliminated from consideration. The final numbers of matched cases dropping these records are shown in the following table.

CONFIDENTIAL

Copyright © 2009 Pearson Education, inc. or its affiliates. All rights reserved

Disposition of Analysis Sample Size					
Grade	Original NJ ASK N-Count	Original Learnia N- Count	Percent Learnia Administered	Final Matched N-Count*	Final Match Rate
3	102,761	22,908	22%	18,490	81%
4	103,360	23,728	23%	19,014	80%
5	103,339	24,427	24%	19,797	81%
6	103,584	26,744	26%	21,167	79%
7	104,297	24,417	23%	19,281	79%
8	105,808	23,951	23%	18,922	79%

*This represents the final sample sized used in the analyses presented pending other attrition.

This table clearly shows that *Learnia* was only administered to roughly 22 to 26% of students in New Jersey in 2009. The final column of the table shows the percent of cases at each grade level that were successfully matched as a function of the original number of *Learnia* cases which is the limiting factor. This match rate was between 79 and 81% and was seen as quite good given the post hoc nature of the study.

While the table above represents the final numbers of matched cases, it should be noted that not every case was complete. That is, the number of *Learnia* tests taken varied from student to student. In some cases, students took either a single mathematics or a single reading test, while in other cases the students took all of the A and B forms of mathematics and reading for a total of four tests. What this meant for subsequent analyses was that the number of cases that could be used was less than the total numbers of matched cases shown in the table above.

Descriptive Statistics and Effect Sizes for Learnia

Table 1 shows descriptive statistics for *Learnia* proportion correct scores across grades, content areas, and forms. The proportion correct metric was used since the number of points on the *Learnia* test forms varied across content areas and grades. Generally, more mathematics tests were given than reading tests, and the A forms were given more often than the B forms. At a given grade level, scores tended to be higher on the B forms than the A forms. One could conclude from Table 1 that the A forms were used as the initial tests in classrooms while the B forms were used for a follow-up administration later on. This will be investigated as part of the user survey planned in Phase II of the research portfolio. With this assumption, the form B proportion correct score minus the form A proportion correct score was used as the difference or “growth measure” between different the *Learnia* “pre/post” administrations.

Table 1. Descriptive Statistics for Learnia Percent Correct Scores.

Grade	N Obs	Learnia Test	N	Mean	Std Dev
3	18,490	Math A	15,176	0.523	0.187
		Math B	11,933	0.637	0.178
		Reading A	14,816	0.549	0.191
		Reading B	11,832	0.661	0.191
4	19,014	Math A	15,928	0.623	0.168
		Math B	12,126	0.668	0.156
		Reading A	15,152	0.584	0.183
		Reading B	11,798	0.583	0.175
5	19,797	Math A	16,868	0.612	0.190
		Math B	12,668	0.651	0.181
		Reading A	15,943	0.550	0.185
		Reading B	12,134	0.612	0.196
6	21,167	Math A	17,443	0.506	0.179
		Math B	11,782	0.563	0.196
		Reading A	16,314	0.618	0.176
		Reading B	11,560	0.632	0.183
7	19,281	Math A	15,268	0.509	0.205
		Math B	9,914	0.534	0.202
		Reading A	14,698	0.611	0.184
		Reading B	10,007	0.562	0.187
8	18,922	Math A	15,593	0.518	0.212
		Math B	9,939	0.578	0.193
		Reading A	13,705	0.595	0.162
		Reading B	10,083	0.656	0.184

Table 2 shows descriptive statistics for the differences in Learnia proportion correct scores and the associated effect sizes, where the effect size d is given by $d = Mean_{Difference} / SD_{Difference}$. Note

that only students that had taken both the A and B forms for a content area were used for this analysis

One way to interpret these effect sizes is to use Cohen’s criteria of $d = 0.20$ being a small effect size, 0.50 being medium-sized, and 0.80 being large. Using these criteria, grade 3 showed effect sizes of over one standard deviation unit for both mathematics and reading. The effect sizes for other grades tended to be in the small to medium range. An interesting exception was in grade 7 reading where the effect size is negative but very small. All in all, most grades and content areas showed some growth across administrations of Learnia.

Table 2. Effect Sizes of Test Differences Between Learnia “Pre and Post Test” Administrations in Reading and Mathematics using the Proportion Correct Metric

Grade	N	Learnia Test	N Complete	Mean	Std Dev	Effect Size
3	18,490	Reading	9,243	0.154	0.139	1.108*
		Mathematics	9,588	0.151	0.137	1.102*
4	19,014	Reading	9,401	0.031	0.137	0.226
		Mathematics	10,117	0.066	0.118	0.559*
5	19,797	Reading	9,968	0.089	0.139	0.640*
		Mathematics	10,820	0.057	0.119	0.479
6	21,167	Reading	9,331	0.054	0.126	0.429
		Mathematics	9,765	0.076	0.122	0.623*
7	19,281	Reading	8,117	-0.016	0.130	-0.123
		Mathematics	8,087	0.051	0.120	0.425
8	18,922	Reading	8,021	0.090	0.123	0.732*
		Mathematics	7,956	0.067	0.126	0.532*

*Medium and large effect sizes (using Cohen’s criteria)

Descriptive Statistics for NJ ASK Scale Scores and NJ ASK Performance in Schools That Implemented Learnia

Table 3 shows the descriptive statistics that were calculated overall for the mathematics, reading, and science scale scores for the NJ ASK tests. For this table, all students across the state in the final matched sample were used.

Table 3. Descriptive Statistics for NJ ASK using the Scale Score Metric.

Grade	NJ ASK Test	N	Mean	Std Dev
3	Language Arts Literacy	18,490	206.61	26.05
	Mathematics	18,490	228.77	40.95
	Science	-	-	-
4	Language Arts Literacy	19,014	206.79	27.56
	Mathematics	19,014	225.63	38.93
	Science	19,014	240.03	26.49
5	Language Arts Literacy	19,797	209.95	26.38
	Mathematics	19,797	229.47	37.46
	Science	-	-	-
6	Language Arts Literacy	21,167	212.15	24.63
	Mathematics	21,167	224.08	36.60
	Science	-	-	-
7	Language Arts Literacy	19,281	216.11	31.00
	Mathematics	19,281	218.12	37.92
	Science	-	-	-
8	Language Arts Literacy	18,922	221.19	22.22
	Mathematics	18,922	224.25	41.95
	Science	18,922	232.63	29.84

Table 4 shows a breakout of the descriptive statistics for the language arts literacy (LAL), mathematics, and science NJ ASK tests based on whether or not a district had fully implemented *Learnia*. This categorization of “fully implemented” was gathered from the Pearson Implementation Managers for *Learnia* who identified districts¹ as having fully implemented *Learnia* if they had:

- Used *Learnia* for over one year;

¹ The districts identified as fully implementing *Learnia* were the Camden City, Cherry Hill, Franklin Township, Hackettstown, Hamburg, Hawthorne, Moonachie, Ridgewood, Saddle Brook, Sparta Township, and Springfield.

- Used the *Learnia* item bank to develop and use tests in addition to the *Learnia* ClassViews tests;
- Used *Learnia* in a formative way to differentiate student instruction through empowered local decision making.

Table 4 also shows for each grade and NJ ASK test the mean scale score difference between those students who were in districts that had fully implemented *Learnia* versus those districts that had not. The effect sizes for these NJ ASK scale score differences are calculated as

$$g = \frac{Mean_1 - Mean_2}{s_p}$$

where s_p is the standard deviation of the pooled within-groups variance

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Using the same criteria as above, these differences show small effect size across most grades (except for grade 5 mathematics and grade 6 LAL and mathematics) indicating somewhat higher NJ ASK performance for students in those districts that fully implemented *Learnia* versus those districts that didn't.

Correlations among NJ ASK and Learnia Tests

Tables 5 through 10 present the Pearson correlations among the NJ ASK scale scores and the *Learnia* proportion correct scores. Correlations between NJ ASK reading and mathematics ran from 0.688 at grade 3 to 0.756 at grade 7. The correlations of science (grades 4 and 8 only) with reading or mathematics were between 0.712 and 0.756. These correlations suggest a moderate to high association between performance on *Learnia* and performance on NJ ASK.

The correlations among *Learnia* proportion correct scores were higher within content area (e.g., Math A with Math B) than across content areas (e.g., mathematics tests with reading tests). A similar result held with the correlations between NJ ASK and *Learnia* test scores: the correlations of NJ ASK scores in a content area with *Learnia* scores in the same content area were higher than the correlations across different content areas. The correlations of NJ ASK mathematics with either of the *Learnia* mathematics scores ran from 0.713 to 0.810, the correlations of NJ ASK Language Arts with the *Learnia* reading scores were from 0.649 to 0.748.

Table 4. NJ ASK Mean Scale Scores by Grade, Content Area, and Learnia Implementation Status.

Grade	NJ ASK Test	Implemented Learnia?	N Obs	ASK SS				Pooled Std Dev	Effect Size
				Mean	Std Dev	Mean Diff.	Var		
3	LAL	No	15,944	205.6	26.0		676.0		0.29
		Yes	2,546	213.2	25.5	7.6	650.3	25.9	
	Mathematics	No	15,944	226.8	41.1		1689.2		
		Yes	2,546	241.3	37.8	14.5	1428.8	40.7	
4	LAL	No	16,419	205.5	27.3		745.3		0.34
		Yes	2,595	214.8	27.6	9.3	761.8	27.3	
	Mathematics	No	16,419	224.0	38.8		1505.4		
		Yes	2,595	235.9	38.4	11.9	1474.6	38.7	
	Science	No	16,419	239.1	26.8		718.2		
		Yes	2,595	246.2	23.3	7.1	542.9	26.4	
5	LAL	No	17,192	209.1	26.2		686.4		0.24
		Yes	2,605	215.5	26.6	6.4	707.6	26.3	
	Mathematics	No	17,192	228.7	37.6		1413.8		
		Yes	2,605	234.7	36.2	6.0	1310.4	37.4	
6	LAL	No	18,605	211.7	24.7		610.1		0.15
		Yes	2,562	215.5	23.8	3.8	566.4	24.6	
	Mathematics	No	18,605	223.3	36.8		1354.2		
		Yes	2,562	229.4	34.4	6.1	1183.4	36.5	
7	LAL	No	16,740	215.1	30.9		954.8		0.25
		Yes	2,541	222.7	30.7	7.6	942.5	30.9	
	Mathematics	No	16,740	216.8	37.9		1436.4		
		Yes	2,541	226.7	36.9	9.9	1361.6	37.8	
8	LAL	No	16,443	220.4	22.2		492.8		0.29
		Yes	2,479	226.8	21.7	6.4	470.9	22.1	
	Mathematics	No	16,443	222.6	41.9		1755.6		
		Yes	2,479	235.4	40.5	12.8	1640.3	41.7	
	Science	No	16,443	231.6	29.7		882.1		
		Yes	2,479	239.5	29.8	7.9	888.0	29.7	

Table 5. NJ ASK and Learnia Test Correlations: Grade 3.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.688	.	0.587	0.593	0.655	0.649
ASK Math	0.688	1.000	.	0.715	0.744	0.616	0.602
ASK Science
Math A	0.587	0.715	.	1.000	0.688	0.667	0.547
Math B	0.593	0.744	.	0.688	1.000	0.584	0.599
Reading A	0.655	0.616	.	0.667	0.584	1.000	0.638
Reading B	0.649	0.602	.	0.547	0.599	0.638	1.000

Table 6. NJ ASK and Learnia Test Correlations: Grade 4.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.696	0.724	0.614	0.590	0.682	0.694
ASK Math	0.696	1.000	0.712	0.719	0.713	0.599	0.630
ASK Science	0.724	0.712	1.000	0.594	0.586	0.608	0.625
Math A	0.614	0.719	0.594	1.000	0.684	0.650	0.607
Math B	0.590	0.713	0.586	0.684	1.000	0.565	0.617
Reading A	0.682	0.599	0.608	0.650	0.565	1.000	0.680
Reading B	0.694	0.630	0.625	0.607	0.617	0.680	1.000

Table 7. NJ ASK and Learnia Test Correlations: Grade 5.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.724	.	0.609	0.645	0.687	0.662
ASK Math	0.724	1.000	.	0.717	0.757	0.609	0.576
ASK Science
Math A	0.609	0.717	.	1.000	0.714	0.642	0.552
Math B	0.645	0.757	.	0.714	1.000	0.635	0.601
Reading A	0.687	0.609	.	0.642	0.635	1.000	0.647
Reading B	0.662	0.576	.	0.552	0.601	0.647	1.000

Table 8. NJ ASK and Learnia Test Correlations: Grade 6.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.745	.	0.666	0.665	0.669	0.687
ASK Math	0.745	1.000	.	0.783	0.788	0.560	0.586
ASK Science	.	.	1.000
Math A	0.666	0.783	.	1.000	0.748	0.584	0.559
Math B	0.665	0.788	.	0.748	1.000	0.568	0.581
Reading A	0.669	0.560	.	0.584	0.568	1.000	0.666
Reading B	0.687	0.586	.	0.559	0.581	0.666	1.000

Table 9. NJ ASK and Learnia Test Correlations: Grade 7.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.756	.	0.697	0.691	0.724	0.748
ASK Math	0.756	1.000	.	0.810	0.810	0.633	0.641
ASK Science	.	.	1.000
Math A	0.697	0.810	.	1.000	0.811	0.641	0.633
Math B	0.691	0.810	.	0.811	1.000	0.619	0.641
Reading A	0.724	0.633	.	0.641	0.619	1.000	0.731
Reading B	0.748	0.641	.	0.633	0.641	0.731	1.000

Table 10. NJ ASK and Learnia Test Correlations: Grade 8.

	ASK LAL	ASK Math	ASK Science	Math A	Math B	Reading A	Reading B
ASK LAL	1.000	0.737	0.746	0.647	0.653	0.689	0.736
ASK Math	0.737	1.000	0.756	0.797	0.802	0.599	0.651
ASK Science	0.746	0.756	1.000	0.668	0.638	0.583	0.628
Math A	0.647	0.797	0.668	1.000	0.762	0.595	0.608
Math B	0.653	0.802	0.638	0.762	1.000	0.576	0.612
Reading A	0.689	0.599	0.583	0.595	0.576	1.000	0.690
Reading B	0.736	0.651	0.628	0.608	0.612	0.690	1.000

Conclusions and Discussion

Some summary conclusions that might result from inspection of the data and analyses presented previously are the following:

1. With the exception grade 7 reading, there is evidence to support the assertion that students improved in their *Learnia* mathematics and reading proportion correct scores between successive *Learnia* administrations. This improvement was most dramatic at grade 3 in both mathematics and reading.
2. Districts that fully implemented *Learnia* tended to score higher on NJ ASK tests than districts that had not fully implemented *Learnia*. Note that the analysis supporting this result does not control for pre-existing achievement differences in these two groups of school districts.
3. There were moderate to strong correlations of *Learnia* proportion correct scores with NJ ASK scale scores in their corresponding content areas. The correlations tended to be higher in mathematics than in reading.

Again since the current study was not a controlled experiment caution should be taken when speculating about any “cause and effect” regarding the performance on *Learnia* and resulting performance on NJ ASK. That said, however, for whatever reasons the districts that fully implemented *Learnia* (as defined in this study) did indeed show higher scores on NJ ASK than their less than fully implemented counter parts. This and the strong correlation between performance on NJ ASK and *Learnia* would suggest that further investigation is warranted to understand why and how districts are using *Learnia* to improve learning at least as it is measured by NJ ASK. Phase II of the *Learnia* Research Portfolio will involve district surveys that might shed some light regarding this important aspect of the investigation.

Limitations

Interpretation of the results should be considered in light of some of the limitations of this study. As such, these limitations are explicitly listed as the following:

1. This study used existing “post hoc” data and does not represent a controlled experiment
2. Implementation of *Learnia* was essentially a district voluntary choice and as such, other background factors have not been controlled in its use
3. Only 22 to 26 percent of the New Jersey student population uses *Learnia*.
4. Of those using *Learnia*, only 79 to 81 percent of the cases survived the match with NJ ASK (i.e., had valid NJ ASK scores that could be matched)
5. Of those matching, not all had complete records so not all cases could be used in all analyses

6. No background factors or potentially intervening or contaminating variables other than the operationally defined “full implementation” of *Learnia* were controlled.

Appendix A. New Jersey Request for Proposal Language**RFP #07-X-39391, January 26, 2007 (pages 28-29)****3.10 TIER I TEST DESIGN AND ADMINISTRATION****3.10.1 TIER I LOCAL DIAGNOSTIC ASSESSMENT AND PROFESSIONAL DEVELOPMENT SERVICES**

The chief goals of the Tier I local assessment program are as follows:

- To foster assessment literacy among local district staff so as to promote the infusion of assessment methods at the local level into daily instructional practices and educational decision making; such assessment literacy would help teachers integrate assessment concepts and methods (e.g., benchmarking, interim assessments, rubric development) into their classroom routines, and help them use test results from both local and statewide assessments to shape instructional practices.
- To provide specific tools and resources that foster dialogue among local district staff about student performance and the ways in which assessments, both formative and summative, can be used to identify appropriate instructional strategies;
- To provide resources for local test authoring and scoring, using non-secure item pools and testlets for use by local district staff in diagnostic activities such as interim benchmarking;
- To provide formative assessments and assessment options that can help both teachers and students prepare for the grade level summative testing event, in such a manner as to integrate, at the cluster level, formative assessment into the classroom learning environment.

The contractor shall prepare and deliver a core, modular diagnostic/formative testing program or set of diagnostic testing resources for delivery and administration at the local level. Such local diagnostics 29 instruments and resources shall be scheduled, administered, scored, and interpreted locally, at the discretion of each district. In consultation with NJDOE content staff, the contractor shall develop written guidelines to assist districts in using best practices to administer and interpret Tier I assessment results.

The purpose of these deliverables is to provide each district with a core set of instruments and resources around which to construct a fuller, holistic local diagnostic assessment program. The districts may supplement this core set of instruments with other diagnostic materials developed or acquired separately by the district.

CONFIDENTIAL

Copyright © 2009 Pearson Education, inc. or its affiliates. All rights reserved

3.10.2 TIER I DIAGNOSTIC/FORMATIVE ASSESSMENT DELIVERABLES

The NJDOE envisions that Tier I shall include a core complement of diagnostic and formative assessment services. Those services and resources shall be designed to equip teachers to implement an ongoing classroom-based assessment program that provides a profile of student progress by means of periodic feedback on student performance throughout the year, developing evidence of student progress in mastering content standards and helping to identify standards not yet mastered. The purpose of this requirement is to assure that the formative assessments play a substantive role in allowing districts to monitor student achievement of the CCCS, in advance of the administration of the secure operational tests, and in such a manner as to shape instructional decisions and strategies.

Therefore, the contractor shall include detailed description of the processes by which the content of its formative assessment system is developed. The contractor shall provide the following:

- A secure, password protected electronic delivery system for accessing a non-secure item bank and/or bank of pre-formatted tests in mathematics, science, and language arts literacy, aligned to New Jersey standards and reflecting the breadth and depth of those standards, that allows teachers to record and track student performance on such tests. All are subject to approval by NJDOE.
- Multiple choice and constructed response items and pre-formatted tests (or testlets) representing multiple levels of difficulty and lend themselves to interim and benchmark testing within the classroom;
- Self-contained training materials, supported by the contractor.

Items and testlets included in this formative assessment system shall be of equivalent rigor to the secure items used on the operational assessments, and include constructed response as well as multiple choice items.

3.10.3 TIER I DIAGNOSTIC ASSESSMENTS – TECHNICAL FEATURES

- The contractor shall provide test authoring and item banking capacities, and the capacity to mix teacher-developed with existing test items.
- The contractor shall provide pre-formatted tests, or “testlets.”
- The contractor shall provide Items and testlets that align with New Jersey’s CCCS.
- The contractor shall provide a system that is scalable and modular, allowing for expansions and changes during the term of the contract.
- The contractor shall provide for varying levels of users to have the capacity to create, administer, and share formative assessments.
- The contractor shall provide a diagnostic assessment system that allows for random and selected subsets of items within or across content clusters to be chosen for the formative tests.

- The contractor shall provide a system that allows for monitoring of formative assessment results over time, on a per student basis.
- The contractor shall provide a system that allows for automatic scoring of formative assessments.