Product Evidence Base

Lexia® PowerUp® Literacy Correlational Research

June 2025

Introduction

Lexia Learning has a long history of building digital programs to help students become proficient readers. For more than 40 years the company has focused solely on literacy. Today, Lexia provides a portfolio of literacy solutions for both students and teachers. Included in the portfolio is Lexia® PowerUp Literacy®, a program that provides literacy intervention for students in grades 6–12 using three strands of instruction: Word Study, Grammar, and Comprehension. PowerUp addresses gaps in basic literacy skills and builds higher-order analytical abilities to accelerate learning. As a blended learning program, PowerUp integrates online activities together with teacher-directed instruction. Included in PowerUp is "Assessment Without Testing" technology which – as an application of Artificial Intelligence – provides teachers with program measures to help plan instruction. This document summarizes correlational evidence from specialized customer evaluations which establish the validity of PowerUp's program measures and provide evidence of the program's effectiveness as a classroom assessment of literacy skills.

Effectiveness research focuses on how well a program works in real-world settings. Efficacy research uses experimental designs to control for confounding variables.

Research on the effectiveness of an instructional program considers its benefits in real-world, local settings. The effectiveness of PowerUp is revealed in evaluation reports prepared by the Lexia Research team that are designed to address the specific needs and requests of individual school districts. These evaluation reports demonstrate how well measures obtained from performance in PowerUp correlate with scores on the district's end-of-year reading test. Customized evaluations are negotiated between Lexia and individual districts. The resulting reports analyze data for students that use PowerUp for a full year, demonstrate fidelity of use, and include at least 100 students. Included in many of these reports are correlations between in-program measures and external tests.

Research on *program effectiveness* relates to research on *program efficacy* in that both intend to assess the impact of a program. However, while effectiveness research asks how well a program works in real-world settings, efficacy research utilizes experimental designs and examines how well the program works with controls in place (Rossi et al., 2003). Both types of research are necessary to understand the types of effects a program is likely to have. The Correlation Reports on PowerUp's effectiveness serve to complement the efficacy research summarized in the PowerUp Literacy Efficacy Research document.

Key Findings

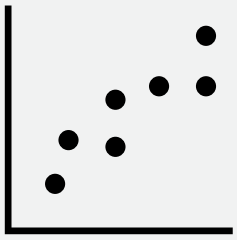
Multiple evaluations of customers' PowerUp implementations show:

- **High to medium correlations between students' progress in PowerUp and common interim and summative tests.**

  Correlations range from 0.43 to 0.63. Data include students who used the online portion of PowerUp for at least 20 weeks and met their weekly usage targets for at least 50% of those weeks.

- **Uniformly positive relationships across districts and states.**

  To date, Lexia researchers have conducted 20 custom evaluations with schools and districts across more than 10 states. These data from over 45,000 students demonstrate consistent positive relationships between PowerUp progress measures and performance on external tests.



Interim Tests

- MAP Growth Reading
- i-Ready Diagnostic
- Star Reading

Accountability Tests

- FAST ELA (FL)
- GMAS ELA (GA)
- M-STEP (MI)
- Smarter Balanced ELA (CA, WA)
- STAAR RLA (TX)
- SOL Reading (VA)

Calculating Correlations

Schools and districts who purchase PowerUp often want to know how measures in PowerUp relate to scores on standardized tests of literacy skills. To answer this question, Lexia researchers typically calculate a correlation, a numerical value which reflects the strength of relationship between PowerUp program measures and student test scores. Positive correlations (ranging from 0 to 1) demonstrate that students who show good progress in PowerUp tend to have high test scores and students with weak progress in PowerUp tend to have low test scores. Correlations closer to 1 indicate a strong relationship; correlations near 0 indicate no relationship and are undesirable.

Across Lexia's research, we categorize correlations into three ranges: high (0.50 to 1.00), medium (0.30 to 0.49), and low (0.10 to 0.29). Correlations of 0.35 and above are considered to be strong evidence of effectiveness when evaluating reading assessments or programs (Hemphill, 2003). When interpreting the magnitude of correlations, it is important to note that many standardized tests used by school districts and states often address broad domains (e.g., English Language Arts) which may include other constructs in addition to literacy skills. For this reason, we may not always expect to observe high correlations between progress in PowerUp and external measures of achievement; instead, medium relationships are likely.

Table 1 provides average estimates of correlations among measures in PowerUp and scores on standardized tests of literacy skills. The correlations in Table 1 were derived from detailed evaluation reports created by the Lexia Research Team. Each row in the table summarizes results obtained for one test, which may have been administered in one or more locations. External tests are broadly divided into two categories – interim and summative tests (Perie, Marion, and Gong, 2009). Interim tests are often marketed as benchmark, formative, or predictive assessments, and they are typically administered during instruction to evaluate students' skills relative to a specific set of academic goals to inform educator decisions or district policy. Summative (or accountability) tests are designed to measure students' attainment of state standards at specific points in time, and they are often accompanied by criterion-referenced levels of proficiency or performance. The correlations presented here are based on both interim and summative tests administered at the end of the school year. Links are provided to either the vendor website or the state website that describes the nature of the test in greater detail.

Table 1.

Correlations among PowerUp Progress Measures and Common Standardized Tests.

Test	n Reports	Recent Year	Cor.	Sample Size
Interim Tests				
MAP Growth Reading	9	23-24	0.52	25,145
i-Ready Diagnostic	1	23-24	0.63	1,742
Star Reading	1	20-21	0.56	1,127
Summative State Accountability Tests				
Florida Assessment of Student Thinking (FAST) ELA	1	22-23	0.43	3,965
Georgia Milestones Assessment System (GMAS) ELA	1	20-21	0.60	960
Michigan Student Test of Educational Progress (M-STEP)	1	23-24	0.60	902
Smarter Balanced ELA	2	23-24	0.59	3,058
State of Texas Assessments of Academic Readiness (STAAR) ELA	2	21-22	0.48	4,179
Virginia Standards of Learning (SOL) Reading Test	2	22-23	0.62	4,315

The correlations presented in Table 1 are weighted averages of the Pearson correlations between students' percent of units completed across the three strands of PowerUp and their scores on the external test. At the request of customers, the Lexia Research team calculates correlations by grade and often by demographic subgroup. When there are multiple evaluations that involve a single test, we compute one grade-specific weighted correlation for each test using two steps. First, we multiply each grade-specific correlation in a report by the number of students in that grade that contributed to that correlation estimate. Second, we divide by the total number of students in each grade across all evaluations with the same test. The resulting grade-specific correlations for each test are presented in the Appendix. To arrive at one estimate of the average correlation between progress in PowerUp and scores on an external test, we follow a similar procedure to calculate weighted average correlations for each test across grades. The correlations in Table 1 fall in the high range (0.50 to 1.00) or medium range (0.30 to 0.49) and provide evidence for the local effectiveness of PowerUp, since these data are collected from various states and school districts across the United States.

Evidence of Correlations as a Measure of Validity

PowerUp is often used as part of teachers' classroom assessment practices. Classroom assessment consists of formative and summative evaluations of students that occur during the process of instruction. The notion of validity developed for large-scale testing programs is often misaligned with the goal of classroom assessment, which is to understand the role of assessment information in the ongoing classroom learning environment (Brookhart, 2003). According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), validity refers to "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). Although large-scale tests and classroom assessments have different goals, some aspects of validity theory can be useful in developing a validity argument (Kane, 2013) for measures obtained from instructional tools like PowerUp that are designed to provide assessment information. Specifically, correlations between percent of PowerUp completed and scores on external tests of literacy skills presented in Table 1 provide evidence of *concurrent validity* – that is, how well PowerUp measures relate to standardized test scores collected at the same time.

The program measures in PowerUp constitute elements of Assessment Without Testing (AWT). As one component of classroom assessment, these measures provide ongoing information about students' literacy skills without the need to administer a test. Reliance on interim or summative tests often requires a separate testing event to collect information from students which may then be used to inform instruction. AWT, on the other hand, provides teachers with information about students' literacy skills via the online component of PowerUp without the need to interrupt instruction to administer a test. This information can be used by teachers to tailor students' classroom experiences using offline lessons, create small instructional groupings, or identify additional scaffolds to support student learning.

Since its inception, PowerUp has contained elements of Artificial Intelligence, which can be defined as "automation based on associations" (U.S. Department of Education, Office of Educational Technology, 2023). The auto placement and branching features in PowerUp are automatic, personalized components of the program based on students' performance. Auto placement serves as a screening tool which is used to assign students to an appropriate

PowerUp's auto placement and branching features are early applications of Artificial Intelligence in literacy instruction.

starting level in each strand of PowerUp. When students experience difficulties in PowerUp, a branching algorithm automatically moves them to guided practice or, if necessary, a direct instruction mode in the program.

The correlations presented in Table 1 provide evidence that progress within PowerUp – as measured by percent of program completed – is associated with external tests of literacy skills. Although these correlations do not provide direct insights into how teachers are using this information to guide instruction, they do suggest that there is a positive relationship between instruction that incorporates PowerUp and students' literacy skills. Future research will continue to advance a classroom assessment validity argument for PowerUp measures by providing further evidence of concurrent validity and gaining greater insight on how educators are using PowerUp to inform instruction.

Appendix: Grade-Specific Correlations

Tables A1 and A2 present grade-specific correlations for each assessment. In cases where multiple evaluations were conducted for customers (i.e., the number of reports is greater than 1), we take the weighted average of the reported correlation for each grade. In Table A3 we provide correlations for historical state summative accountability tests, which are no longer in use.

Table A1.

Correlations with Common Interim Tests.

Test	n Reports	Recent Year	Grade-Specific Correlation (Total Sample)		
			6	7	8
MAP Growth Reading	9	23-24	0.56 (9962)	0.50 (8545)	0.50 (6638)
i-Ready Diagnostic	1	23-24	0.60 (818)	0.67 (494)	0.63 (430)
Star Reading	1	20-21	0.56 (370)	0.52 (429)	0.61 (328)

Table A2.

Correlations with Common Summative State Accountability Tests.

Test	States	n Reports	Recent Year	Grade-Specific Correlation (Total Sample)				
				6	7	8	9	10
FAST ELA	FL	1	22-23	0.52 (1904)	0.42 (473)	0.43 (519)	0.36 (546)	0.22 (523)
GMAS ELA	GA	1	20-21	0.60 (364)	0.55 (396)	0.70 (200)	-	-
M-Step	MI	1	23-24	0.54 (444)	0.65 (458)	-	-	-
Smarter Balanced ELA	CA, WA	2	23-24	0.60 (1413)	0.59 (960)	0.59 (685)	-	-
STAAR ELA	TX	2	21-22	0.50 (1877)	0.45 (1592)	0.52 (710)	-	-
SOL Reading	VA	2	22-23	0.64 (2013)	0.55 (1480)	0.60 (822)	-	-

Table A3.

Correlations with Historical Summative State Accountability Tests.

Test	State	n Reports	Recent Year	Grade-Specific Correlation (Total Sample)		
				6	7	8
ACT Aspire English	AR	1	20-21	0.72 (1,078)	- -	- -

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*. https://www.testingstandards.net/

Brookhart, S. M. (2005). Developing measurement theory for classroom assessment purpose and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5-12. https://doi.org/10.1111/j.1745-3992.2003.tb00139.x

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*(1), 78–79. https://doi.org/10.1037/0003-066X.58.1.78

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448-457. https://doi.org/10.1080/02796015.2013.12087465

Lexia Research. (2025). Lexia® PowerUp® Literacy Efficacy Research. https://www.lexialearning.com/resources/research/lexia-powerup-literacy-efficacy-research

Perie, M., Marion, S., & Gong, B. (2009). Moving towards a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, *28*(3), 5-13. https://doi.org/10.1111/j.1745-3992.2009.00149.x

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach* (7th ed.) Thousand Oaks, CA: Sage Publications, Inc.

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations*. Washington, DC. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Lexia®

a **cambium** company

Lexia®, a Cambium Learning Group company, is the Structured Literacy expert. For more than 30 years, the company has focused solely on literacy, and today provides science of reading-based solutions for both students and educators. With robust offerings for differentiated instruction, personalized learning, assessment, and professional learning, Lexia helps more learners read, write, and speak with confidence.

lexialearning.com