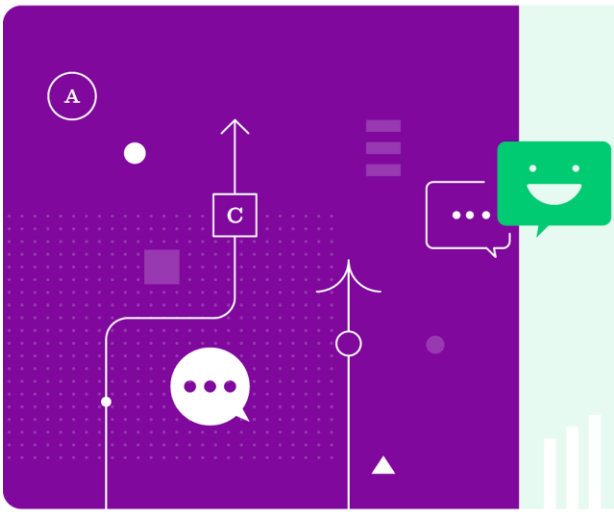
AMIRA
LEARNING

Using Amira to Increase Performance on North Dakota State Assessments English Language Arts Outcomes

Amira Learning | Every Child Deserves the Chance to Become a Reader
5214F Diamond Heights Blvd, #3255, San Francisco, CA 94131
wwww.AmiraLearning.com | 866-883-7323 | info@amiralearning.com



THE INTELLIGENT
Growth Engine

Using Amira to Increase Performance on North Dakota State Assessments English Language Arts Outcomes

PREPARED BY
Matthew Jeans, PhD

DATE
01/16/2025

Executive Summary

Amira is an advanced AI-powered literacy assessment and tutoring platform. By employing voice recognition and machine learning, Amira provides personalized reading practice and immediate feedback, addressing individual skill gaps to improve literacy outcomes.

Amira recommends that students engage with the platform for 30 minutes per week to enhance their reading performance. Previous research with the Amira Tutor has demonstrated increased reading performance. This research evaluates if Amira Tutor improves academic growth on the North Dakota State Assessment (NDSA) in English Language Arts (ELA).

Using data from 27 North Dakota school districts in the 2023–24 school year, hierarchical linear models were used to account for clustering at the school level. Usage was divided into quartiles, with quartile 1 indicating the lowest amount of usage and quartile 4 indicating the highest amount of usage. Results indicated that Amira Tutor usage led to higher NDSA ELA scores in third through fifth grade students:

- Third grade students in usage quartile 4 scored 15 points higher on the NDSA ELA than those in usage quartile 1.
- Fourth grade students in usage quartiles 3 and 4 scored 12-17 points higher on the NDSA ELA than those in usage quartile 1.
- Fifth grade students with higher usage scored up to 10 points higher on the NDSA ELA than those with lower usage, though statistically insignificant.

The results demonstrate that higher engagement with Amira positively influences NDSA ELA performance, particularly in third and fourth grades. These findings align with prior research validating AI-driven literacy interventions as transformative tools in education.

Table of Contents

Introduction

Advances in educational technology are transforming how literacy skills are assessed and developed, offering new opportunities to address longstanding challenges in student achievement. This study focuses on Amira Learning, an AI-powered literacy assessment and tutoring platform designed to enhance reading outcomes through personalized, data-driven instruction and progress monitoring.

Amira Learning utilizes artificial intelligence and voice recognition technology to assess key literacy skills, including phonemic awareness, decoding, and fluency. Through dynamic interactions, Amira diagnoses skill gaps, provides targeted feedback, and recommends practice activities tailored to individual student needs. This approach empowers educators with actionable insights and helps students develop critical reading skills in a supportive environment. Amira also incorporates a tutoring component that actively engages students in reading practice, offering immediate guidance and correction. Studies have shown that Amira's tutoring system is effective in improving literacy outcomes. For example, Mostow et al. (2003) demonstrated that the automated Reading Tutor achieved outcomes comparable to human tutoring and classroom instruction, highlighting its potential to support individualized literacy development. Additionally, Mostow et al. (2013) found that computer-guided oral reading through Amira improved fluency and comprehension more effectively than independent practice.

Amira's efficacy is further supported by studies conducted in diverse educational contexts. In Utah, a 2023 evaluation found that Amira significantly outperformed a comparison group in raising literacy achievement in a randomized controlled trial (Evaluation and Training Institute, 2023). Other research has confirmed Amira's ability to improve literacy outcomes in bilingual students and students from under-resourced communities (Poulsen & Wiemer-Hastings, 2007; Korsah et al., 2010). These findings reinforce the program's versatility and alignment with evidence-based practices for literacy instruction.

This study explores the relationship between time spent on Amira's tutoring platform and student literacy growth. Specifically, it aims to answer the following questions:

1. How does usage time on Amira's tutoring platform improve scores on the NDSA ELA assessment?
2. Does Amira usage vary between schools?

Methodology

Analytical Sample

The data are from students across 27 districts in North Dakota. This study focused on third through fifth grade students (n=3,046). Table 1 provides demographic characteristics of the sample. The largest racial/ethnic group was White (75.5%), followed by Hispanic (8.8%), American Indian (6.4%), and Pacific Islander (<1%). Approximately 5.0% of the sample were classified as English language learners.

Table 1. *Student Characteristics for Analytical Sample*

Group	Percent (%)
White	75.48
Hispanic	8.83
American Indian	6.43
Black	5.95
Asian	2.64
Pacific Islander	0.59
Male	50.26
English Learner	5.04
N	3,046

Measures

NDSA ELA

The NDSA ELA is a comprehensive summative assessment administered annually to students in grades 3 through 8 and grade 10. This assessment measures students' proficiency in ELA, encompassing reading, writing, speaking, listening, and research/inquiry skills, aligning with the North Dakota ELA content standards.

The NDSA is designed to evaluate students' mastery of state-adopted content standards, providing valuable information on their readiness for the next grade level. As a summative assessment, it offers a single measure at the conclusion of the

academic year. However, integrating classroom formative assessments alongside the NDSA can offer supplementary insights, enabling educators to tailor instruction more effectively.

In this study, the NDSA ELA assessment scores serve as the primary outcome measure, providing a standardized benchmark against which the impact of instructional practices, such as Amira Tutor, on student achievement in English Language Arts can be evaluated.

Amira Tutor

Amira Tutor is a research-based, AI-driven instructional program designed to improve literacy outcomes by providing personalized reading practice and support. Unlike traditional assessments, Amira Tutor serves as a one-on-one digital tutor, using advanced artificial intelligence to deliver targeted instruction in real-time. The system engages students in oral reading activities, providing immediate feedback on fluency, pronunciation, and comprehension while tracking progress over time.

The tutor incorporates tasks aligned with foundational literacy skills, including phonological awareness, phonics, decoding, oral reading fluency, and vocabulary acquisition (Amira Technical Guide, 2024). By leveraging voice recognition and machine learning, Amira Tutor listens to students read aloud, identifies errors, and provides corrective feedback tailored to their specific needs. These interactions are informed by principles from the science of reading and are further enriched with diagnostic reports that offer actionable insights for educators and parents.

Key features of Amira Tutor include:

- **Immediate Feedback**: Amira analyzes students' reading in real time, helping them self-correct errors and improve fluency.
- **Adaptive Challenges**: The platform adjusts the difficulty of reading passages and exercises based on student performance.
- **Progress Monitoring**: Detailed, automatically generated reports highlight student strengths and areas for improvement, empowering educators to make data-driven instructional decisions.
- **Evidence-Based Design**: Developed in collaboration with leading institutions such as Carnegie Mellon University and Johns Hopkins University, Amira is rooted in validated literacy research and predictive accuracy for learning outcomes.

Analytical Approach

Due to the sample having students nested in schools, a hierarchical linear model (HLM) was used to examine the efficacy of Amira Tutor usage on NDSA ELA assessment scores from the 2023-24 school year.

Two-level HLM models control for the effects at the student level (Level 1) and the school level (Level 2). Three nested models were tested. Model 1 is the baseline model that consists of only the random effect for the intercept. Model 2 is an extension of model 1 that includes fixed effects at Level 1, which includes usage quartiles. Model 3 is an extension of model 2 that includes random slopes for Level 1 (total usage).

Results

Table 2 shows the total minutes by observed usage quartiles associated with each grade level. Amira recommends spending 30 minutes each week using Amira Tutor, which would be approximately 960 minutes total for the school year. Those meeting the recommended usage are represented in quartile 4.

Table 2. *Usage Quartiles and Total Minutes for the School Year*

Usage Quartile	Usage Percentile Rank	Grade 3	Grade 4	Grade 5
1	≤25	≤39	≤46	≤57
2	26–50	40–113	47–127	58–163
3	51–75	114–295	128–242	164–298
4	>75	>296	>243	>299

Table 3 shows the proportion of students within each NDSA ELA performance band by grade level. The largest group of students across grade levels were in the *Novice* performance band, followed by *Proficient*, *Partially Proficient*, and *Advanced*.

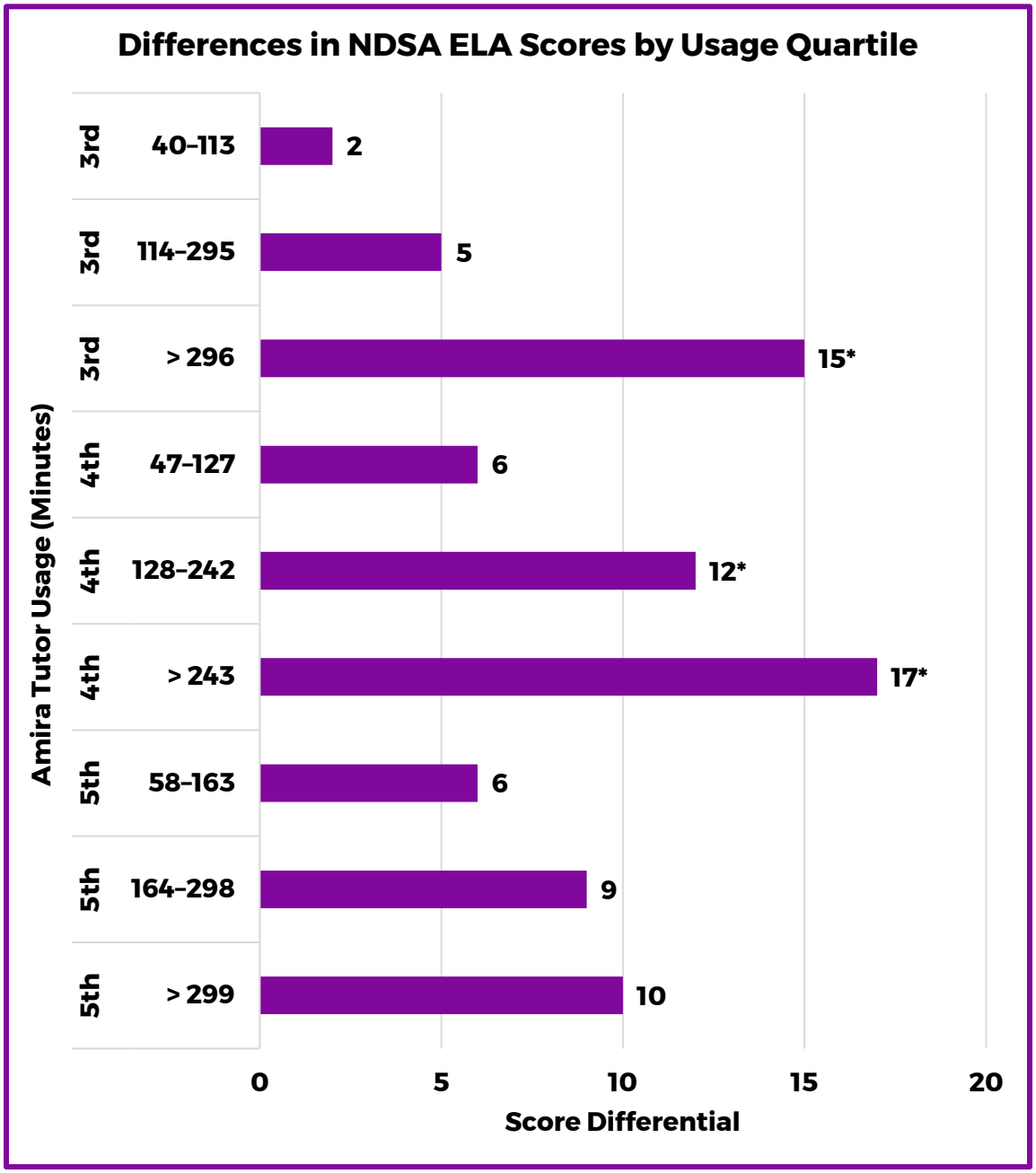
Table 3. *NDSA ELA Performance Bands by Grade Level*

Performance Band	Grade 3	Grade 4	Grade 5
Novice	33.3%	30.2%	36.6%
Partially Proficient	23.8%	25.7%	24.8%
Proficient	31.4%	29.8%	27.9%
Advanced	11.5%	14.2%	10.7%

Amira Usage on NDSA ELA Outcomes

Figure 1 shows the graphical representation of increases in NDSA ELA scores by total minutes of usage per school year for each grade level.

Figure 1. *Differences in NDSA ELA Gain Scores for Third Through Fifth Grade Students by Amira Tutor Usage*



* $p < 0.05$

Usage Results in Third Grade Students

Table 4 presents the results of the two-level HLM for third grade students (n = 1,259). Those in the fourth Amira usage quartile (>296 minutes) scored approximately **15 points higher** on the NDSA ELA assessment compared to students in the lowest usage quartile ($p < 0.05$). Sixteen percent (16%) of the score variability was attributable to differences between schools, while 84% was due to individual student differences. The significant Level-2 error variance highlights that schools play a role in score variability even after controlling for total usage. Model 3, which includes random slopes for total usage, reveals that higher Amira usage is significantly associated with improved scores, with this relationship varying across schools.

Table 4. *Two-Level HLM for Third Grade, Coefficients and Standard Errors (SE)*

Fixed Effects	Model 1	Model 2	Model 3
Intercept	572.84* (2.37)	568.89* (3.21)	568.28* (3.25)
Usage 2 (26-50)		1.40 (3.19)	2.25 (3.19)
Usage 3 (51-75)		3.13 (3.56)	4.53 (3.73)
Usage 4 (>75)		10.04* (4.10)	15.28* (4.94)
Error Variance			
Level-1	1227.66* (142.67)	1224.98* (50.06)	1186.65* (49.55)
Level-2 Intercept	238.79* (62.76)	233.92* (61.03)	234.16* (70.84)
Total Usage			0.002* (0.001)
Model Fit			
AIC	12614.7	12601.7	12591.4
BIC	12630.1	12632.5	12627.4

* $p < 0.05$; ICC = .16

Values based on Stata 18.5 Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = REML; Satterthwaite degrees of freedom.

Usage Results in Fourth Grade Students

Table 5 presents the results of the two-level HLM for fourth grade students (n = 1,243). Since Model 3 did not significantly improve model fit compared to Model 2, the latter was interpreted. In Model 2, students in the highest Amira usage quartile (>243 minutes) scored approximately **17 points higher** on the NDSA ELA than students in the lowest usage quartile ($p < 0.05$). Similarly, students in the third usage quartile (128–242 minutes) scored approximately **12 points higher** than those in the lowest quartile ($p < 0.05$).

The intra-class correlation (ICC) of 0.10 indicates that 10% of the score variability was attributable to differences between schools, while the remaining 90% was due to

differences between students within schools. The significance of the Level-2 error variance suggests that schools influence the variability in students' scores after accounting for total usage.

Table 5. *Two-Level HLM for Fourth Grade, Coefficients and Standard Errors (SE)*

Fixed Effects	Model 1	Model 2
Intercept	590.51* (2.28)	581.40* (3.27)
Usage 2 (26-50)		6.20 (3.44)
Usage 3 (51-75)		12.39* (3.83)
Usage 4 (>75)		17.43* (4.07)
Error Variance		
Level-1	1509.58* (61.95)	1490.10* (61.19)
Level-2 Intercept	168.10* (55.59)	165.59* (54.45)
Model Fit		
AIC	12685.6	12659.9
BIC	12701.0	12690.6

* $p < 0.05$; ICC = .10

Values based on Stata 18.5 Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = REML; Satterthwaite degrees of freedom.

Usage Results in Fifth Grade Students

Table 6 displays the results of the two-level HLM for fifth-grade students (n = 544). In this analysis, Model 3 was interpreted as it included random slopes for total usage. Students in the highest Amira usage quartile (>299 minutes) scored approximately **10 points higher** on the NDSA ELA compared to students in the lowest usage quartile, although this difference was not statistically significant ($p = 0.16$). Students in the third usage quartile (164–298 minutes) scored about **9 points higher**, and those in the second quartile (58–163 minutes) scored about **6 points higher** than the lowest quartile, but these differences were also not statistically significant ($p > 0.15$ and $p = 0.26$, respectively).

The intra-class correlation (ICC) was 0.15, indicating that 15% of the score variability was attributable to differences between schools, with the remaining 85% due to differences between students within schools. Despite the inclusion of random slopes for total usage, the non-significant effects for usage quartiles may be partially due to the smaller sample size for this grade level, which limited the statistical power of the model.

Table 6. *Two-Level HLM for Fifth Grade, Coefficients and Standard Errors (SE)*

Fixed Effects	Model 1	Model 2
Intercept	608.51* (3.52)	602.45* (5.34)
Usage 2 (26-50)		6.35 (5.57)
Usage 3 (51-75)		8.85 (6.24)
Usage 4 (>75)		9.59 (6.83)
Error Variance		
Level-1	1440.09* (89.90)	1436.88* (89.95)
Level-2 Intercept	260.86* (98.80)	284.18* (106.62)
Model Fit		
AIC	5539.3	5527.6
BIC	5552.2	5553.4

* $p < 0.05$; ICC = .10

Values based on Stata 18.5 Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = REML; Satterthwaite degrees of freedom.

Conclusion

This study provides compelling evidence for the effectiveness of Amira Tutor in improving literacy outcomes among elementary students, while also highlighting key factors that influence its impact. The analysis demonstrated that students in higher usage quartiles consistently outperformed their peers in lower quartiles on the NDSA ELA, with statistically significant gains observed for third and fourth graders. These findings align with previous research demonstrating the benefits of increased engagement with AI-driven instructional tools like Amira.

The majority of students in this study scored within the Novice and Proficient performance bands, with relatively fewer students achieving Advanced proficiency. This distribution underscores the critical need for targeted interventions to support students at lower performance levels. Amira Tutor's adaptive approach, which aligns with evidence-based literacy practices, positions it as a valuable tool to address these gaps by providing personalized reading practice and real-time feedback.

The study also highlights the importance of adhering to recommended usage thresholds for maximizing Amira's impact. Students in the highest quartile of usage, who approached or exceeded the recommended 960 minutes per year, consistently demonstrated better outcomes than those with lower engagement. This emphasizes the need for consistent usage to realize the full benefits of the platform. Additionally, the significant intra-class correlations observed across grade levels indicate that school-level factors contribute meaningfully to student outcomes. Variability in

implementation fidelity, teacher support, and access to resources may influence how effectively Amira is used in different educational contexts.

Typically, analyses of this nature would include a school-level variable such as socioeconomic status (SES) to account for additional school-level effects that could influence student outcomes. However, this data was not available for the present study. Including such a variable could have provided a more nuanced understanding of how contextual factors shape the relationship between Amira usage and literacy achievement. Despite this limitation, the significant results observed across usage quartiles underscore the robustness of Amira's impact on literacy growth, even in the absence of a school-level SES control.

Amira Tutor's design, which incorporates principles of the science of reading and leverages advanced AI technology, highlights its potential to support diverse student populations. This includes students from under-resourced schools and English learners, who face unique challenges in achieving literacy proficiency. These findings reinforce Amira's role as a scalable solution that can complement traditional classroom instruction and provide equitable access to high-quality literacy support.

References

Amira Learning. (2024). Amira Technical Guide.

Evaluation and Training Institute. (2023). Utah's Early Intervention Reading Software Program Report.

Korsah, G. A., Mostow, J., Dias, M. B., Sweet, T., Belousov, S. M., Dias, M. F., & Gong, H. (2010). Improving literacy in Africa: Experiments with an automated reading tutor. *Information Technologies & International Development, 6*(2), 1–19. Retrieved from https://www.cs.cmu.edu/~listen/pdfs/2010%20ITID%20Korsah%20517-1388-1-PB.pdf

Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., Huang, C., Junker, B., Sklar, M. B., & Tobin, B. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29*(1), 61–117. https://doi.org/10.2190/06AX-QW99-EQ5G-RDCF

Mostow, J., Nelson, J., & Beck, J. E. (2013). Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research, 49*(2), 249–276. https://doi.org/10.2190/EC.49.2.g

Poulsen, R., Wiemer-Hastings, P., & Allbritton, D. (2007). Tutoring bilingual students with an automated reading tutor that listens. *Journal of Educational Computing Research, 36*(2), 191–221. https://doi.org/10.2190/A007-367T-5474-8383