

NoRedInk and the NWEA MAP Language Usage Test

A Case Study at Shadow Ridge Middle School in Colorado

Christian F. Perez, Ph.D.

Abstract

NoRedInk is a web-based adaptive learning environment that helps millions of students improve their grammar and writing skills. The Measures of Academic Progress (MAP) test is used nationally to measure student progress through 12th grade in four subject areas. This study compared the NoRedInk usage and performance of 265 seventh-grade students to the improvement in MAP Language Usage score and subscores from Fall to Spring over the same period. Results from linear models suggest that, on average, every 10-percentage-point increase in the percent of NoRedInk questions answered correctly correlates with more than one grade level improvement on the MAP.

Introduction

NoRedInk's online software helps students learn grammar and writing skills with curriculum generated from their interests and an adaptive learning engine that caters to their strengths and weaknesses. The company was founded in 2012 by Jeff Scheur, a National Board Certified Teacher who had graded over 15,000 English papers and wanted to leverage technology to help his students become strong, confident writers.

NoRedInk (NRI) has seen rapid adoption and enthusiasm from English teachers nationwide, and millions of students have answered hundreds of millions of questions on the site. Teachers' reported experience has been overwhelmingly positive, but until now, evidence of NRI's effect on student learning has been anecdotal.

Here we present a regression model that compares student usage of NRI and their performance on the Measures of Academic Progress (MAP) Language Usage test, administered by the NWEA to 10 million 2nd-12th grade students every year. We analyze the correlation between NRI usage and performance and improvements on the MAP to ask two questions relevant to educators:

1. Does NRI activity correlate with MAP improvements?
2. Does NRI performance correlate with MAP improvements?

Design

This study considers 265 seventh-grade students from one middle school taking English for 33 weeks of instruction between a pre-test in August and a post-test in May. Seventh-grade students used the premium version of NRI. Almost all students in the seventh grade used NRI as part of their grammar instruction. Teachers were free to implement NRI however they saw fit,

and there was no control group to establish efficacy. Therefore, this analysis explores the correlations between NRI usage and MAP performance over the course of one school year.

NRI was typically used to supplement students' grammar instruction in English class by completing voluntary NRI assignments and compulsory in-class graded quizzes on teacher-assigned grammar topics. All three teachers used NRI for at least one 40-minute class per week.

NWEA MAP Language Usage test scores and subscores use the RIT scale, a measurement of student learning scaled using Item Response Theory, which has the same interpretation independent of a student's grade or age.¹ Learning gains are measured as the difference in RIT score between a student's Spring and Fall MAP Language Usage assessments.

Table 1: Student Characteristics

Male	52.3%	Female	47.7%
ESL	14.7%		
Special needs	5%		
Teacher 1 & 2	~48%	Teacher 3 (ELL)	~4%

Table 2: MAP Language Usage Test Performance

	Pre-test (Fall)		Post-test (Spring)	
	Mean	SD	Mean	SD
Overall	213.5	11.7	218.3	11.6
Grammar	212.6	12.4	216.8	11.1
CPS	214	12.4	219.3	12.7
PODRR	213.9	13.1	219.1	13.6

Every NRI question attempted and answered was recorded for every student, as well as the date, time, difficulty, number of attempts, grammar category, and grammar rule tested. Students were matched to the NRI database using a combination of their name and student ID number. Some students had multiple accounts: their usage over the treatment period was aggregated, and in one case where a single student had multiple teachers, she was assigned to the teacher with whom she had answered the most questions on NRI.

¹ Northwest Evaluation Association. *The RIT Scale*. Retrieved from <http://legacysupport.nwea.org/node/4344>

Analysis

Ordinary least squares (OLS) regression was used to model improvement on the MAP overall score and subscores between Fall and Spring, using student attributes from Table 1, including pre-test scores, teacher fixed effects, and NRI usage (total # of questions answered) and performance (measured by overall percent correct on answers over the treatment period, typically between 50 and 90%). All real-valued model inputs were median-subtracted to reduce adverse effects from multicollinearity. Students with language proficiencies below FEP5 (See Appendix B) were categorized as ESL, and students with missing values were assumed to have English fluency.

The strongest overall predictors of growth are fall pre-treatment scores, consistent with results from other studies.² Therefore, the regression model of fall-spring improvement on the overall MAP score includes the overall fall MAP score (Table 3, column 1). The MAP Language Usage test consists of three subgoals:³ Grammar; Capitalization, Punctuation, and Spelling (CPS); and Plan, Organize, Develop, Revise, and Research (PODRR). In order to explore the effect of NRI on the individual subgoals, a linear model of MAP subscore improvements was regressed against each of the fall pre-treatment subscores and NRI activity associated with either Grammar or CPS subgoals (Table 3, column 2; see Appendix A). Reported correlations represent statistically meaningful relationships between academic performance measured by the MAP test and NRI metrics.

Seven students were classified as outliers according to Cook's D and influence analysis and excluded from the regression: 3 had abnormally large positive MAP improvements, 3 had abnormally large or inexplicable negative MAP declines, and 1 answered too few NRI questions. The OLS results were also consistent with a robust linear model on data that excluded only the two largest outliers, with the exception of the PODRR subscore regression (see Table 4).

Results

Overall performance

The relationship between NRI usage and the improvement in MAP scores was analyzed in two ways. The simplest model (Table 3, column 1) compares the overall MAP improvement to the pre-test score and the aggregated total number of questions answered and the overall percent correct on NRI questions over the entire school year. Every 10-percentage-point increase in percent correct (i.e., 60% to 70%) is correlated with an average of 3.9 RIT points improvement

² See, e.g., Northwest Evaluation Association (2011). "Table C.17: Means (M) and Standard Deviations (sd) of Academic Growth by Starting RIT Score and Starting Instructional Week for Various Ending Instructional Weeks - Language Usage, Grade 7." *RIT Scale Norms Study*. Retrieved from http://legacysupport.nwea.org/sites/www.nwea.org/files/resources/NWEA%202011%20Norms%20Report_01.17.2012_2.pdf

³ Northwest Evaluation Association (2014). *RIT Reference Charts for Common Core and Science*. Retrieved from <https://www.nwea.org/content/uploads/2014/07/NWEA-RIT-Reference-Brochure-Digital.pdf>

on the MAP, which corresponds to more than one grade level.⁴ This suggests that NRI assessments are well aligned to MAP topic areas in general, although the low correlation coefficient indicates the model is not highly predictive of any individual student’s performance.

The second model compares the overall improvement in MAP score to the individual subscores on the pre-test (Table 3, column 2) and the same NRI metrics. The similarity between both models suggests that either approach can capture the same relationships, lending validity to the subscore approach. This second model can then be used as a point of comparison to the subscore analysis that follows. As expected, the initial MAP subscores are also significant predictors of overall performance gains over the course of the year (Table 3, column 2). Other model coefficients agree within error, suggesting either approach is equally descriptive of the data. In both models, higher NRI performance is correlated with higher MAP performance, and negative pre-test coefficients indicate that students with lower initial fall scores have more potential for learning gains as they are caught up to their grade level, while those students with higher scores typically realize smaller gains.

Although answering more questions is generally positively correlated with MAP improvement, in this sample, there was not a statistically significant correlation with the total number of questions answered. That is, for students with the same attributes, pre-test score, and percent correct on NRI, simply answering more questions (and not improving in percent correct) did not result in a significant increase in MAP performance. Since only 12% of students answered more than 2000 questions over the course of the year, there is not enough data at this high usage rate for the model to detect an effect, if one exists.

Table 3: Overall Performance

	Overall improvement	Overall improvement vs subscores
Pre-test	-0.47*** (0.05)	
Grammar		-0.24*** (0.05)
CPS		-0.20*** (0.05)
PODRR		-0.05 (0.05)
Intercept	6.9*** (0.7)	7.2*** (0.7)
Percent correct	0.39*** (0.06)	0.4*** (0.06)
Total answered	0.85	0.79

⁴ Northwest Evaluation Association (2015). “2015 Language Usage Student Growth Norms.” *2015 NWEA Measures of Academic Progress: Normative Data*. Retrieved from <https://www.nwea.org/content/uploads/2015/07/2015-NWEA-MAP-Normative-Data-JUN15.pdf>

(in thousands)	(0.48)	(0.47)
Teacher A	-3.3*** (0.8)	-3.5*** (0.8)
Teacher B	-7.9*** (2.2)	-7.4** (2.2)
Special needs	-4.9** (1.8)	-5.0** (1.8)
ESL	-0.08 (1.2)	0.06 (1.2)
Female	-0.22 (0.75)	0.9 (0.76)
R ² adj.	0.278	0.288
N	258	258
*** p < 0.001 ** p < 0.01 * p < 0.05		

Standard errors are in parentheses.

Subscore performance

The improvement in subscores was modeled against NRI engagement, measured by the total number of questions attempted, and NRI performance on topics aligned with the grammar and CPS subscores, measured by the percent correct.

- Subscore improvements all have significant negative correlations with their corresponding pre-test value, consistent with analysis on overall MAP improvements
- Every 10-percentage-point increase in percent correct on grammar usage questions in NRI was correlated with 3 RIT points higher on the MAP grammar usage subscore on average.
- Similarly, every 10-percentage-point increase on CPS percent correct was correlated with 2.2 RIT higher on CPS subscore on average
- NRI does not have any topics specifically aligned to the PODRR section of the MAP, but there were statistically significant positive correlations for NRI performance on both grammar and CPS-related questions
- MAP improvement is positively correlated with total questions answered, but the relationship is not statistically significant, consistent with the analysis of the overall improvement
- NRI performance is most predictive of grammar improvements on the MAP, as expected from an adaptive grammar learning tool

Table 4: Subscore Performance

Subscore Improvement	Grammar	CPS	PODRR ⁵
Pre-test			
Grammar	-0.97*** (0.06)	0.18* (0.07)	0.08 (0.08)
CPS	0.10 (0.06)	-0.78*** (0.07)	0.06 (0.07)
PODRR	0.28*** (0.06)	0.15 (0.07)	-0.61*** (0.08)
Intercept	6.7*** (0.8)	6.6*** (1.0)	8.1*** (1.0)
Grammar percent correct	0.30*** (0.05)	0.15* (0.06)	0.22** (0.07)
CPS percent correct	0.02 (0.05)	0.22** (0.06)	0.21** (0.07)
Total answered (in thousands)	0.68 (0.52)	1.34 (0.63)	0.41 (0.67)
Teacher A	-3.45*** (0.85)	-2.89* (1.0)	-3.54** (1.08)
Teacher B	-7.0* (2.4)	-4.71 (2.86)	-7.0 (3.1)
Special needs	-4.0 (1.9)	-5.3 (2.3)	-3.2 (2.2)
ESL	1.6 (1.4)	-0.5 (1.6)	-0.63 (1.8)
Female	0.06 (0.9)	0.8 (1.0)	-0.53 (1.1)
R ² adj.	0.552	0.336	0.256
*** p < 3.3e-4 ** p < 3.3e-3 * p < 0.017 (Sidak corrected ⁶)			

⁵ Due to residual errors, ordinary linear regression was replaced with a robust linear model for PODRR only, resulting in smaller effects and larger p-values but more reliable parameters

⁶ Corrected p-value of 0.05, 0.01, and 0.001 for n = 3 multiple comparisons

Discussion

Results suggest NRI usage is correlated with gains on the MAP Language Usage test, and it is recommended to pursue further study in a controlled trial. As with any educational tool, ongoing evaluation of its impact on student learning outcomes in different settings will help educators understand NRI's potential in their classrooms.

Appendix A

MAP subscores and corresponding NoRedInk topics, as assigned by NRI designers.

Grammar and Usage:

Subject-Verb Agreement
Pronoun Case
Adjectives and Articles
Adverbs
Sentences and Fragments
Pronoun Agreement
Clauses and Phrases
Nouns
Verb Tense

Writing Conventions:

Capitalization
Quote Transitions
Apostrophes
Commas
Colons and Semicolons
Conjunctions

Appendix B

ENGLISH LANGUAGE LEARNERS AND LANGUAGE CODING

ELL: English Language Learner; a general term for a student whose dominant language is not English and who is learning English

NEP: Non-English Proficient; an English Language Learner whose overall proficiency in listening, speaking, reading, and writing is at the Beginning or Early Intermediate stages; this level includes students who are just beginning to understand and respond to simple routine communication through those who can respond with more ease to a variety of social communication tasks

LEP: Limited English Proficient; an English Language Learner whose overall English proficiency in listening, speaking, reading, and writing is at the Intermediate or Advanced stages; students at this level are able to understand and be understood in many to most social communication situations, are gaining increased competence in the more cognitively demanding requirements of content areas, but are not yet ready to fully participate in academic content areas without linguistic support

FEP: Fluent English Proficient; a student who was formerly ELL but who has acquired fluent English proficiency in listening, speaking, reading, and writing

Language Proficiency field codes in Infinite Campus:

0 – English

1 – NEP 1

2 – NEP 2

3 – LEP 3

4 – LEP 4

5 – FEP 5