Using Automated Feedback to Develop Writing Proficiency

Yue Huang, M.S.Ed., Joshua Wilson, Ph.D.
University of Delaware

Context

Implementing an ongoing and rapid practice-feedback cycle is critical for developing writing proficiency, however, it is quite challenging for teachers to enact such a cycle. To help address the challenges, formative writing assessment systems called automated writing evaluation (AWE) have been developed to provide students with immediate automated scoring and automated feedback on their writing.

AWE = web-based formative writing assessment software that uses computer algorithms to analyze the quality of students' writing and provide students with automated scoring and automated feedback to help students calibrate and improve their writing performance.

Findings of prior research show promise (e.g., Stevenson & Phakiti, 2014), but there is a dearth of research on the effectiveness of AWE for supporting the teaching and learning of writing in the elementary grades.

Thus, the current study examined whether the use of an AWE system called MI Write helped students in Grades 4–5 develop their writing proficiency across a school year. We specifically looked at students' growth in writing quality when revising with the aid of automated feedback from MI Write, and whether the use of MI Write was associated with transfer to improved first-draft writing performance or more efficient and effective revising at the end of the school year.

Research Questions

RQ1: Do students using AWE improve their writing performance across successive drafts of an initial essay (administered in the Fall)? What is the shape and rate of growth across those drafts? Is the rate of improvement equal for all groups of students?

RQ2: Do students using AWE improve their writing performance across successive drafts of a subsequent essay (administered in the Spring)? What is the shape and rate of growth across those drafts? Is the rate of improvement equal for all groups of students?

RQ3: After implementing AWE as part of writing instruction, do students display gains in their independent first-draft writing performance (i.e., unaided transfer) and their efficiency and effectiveness of revising with AWE (i.e., aided transfer)?

RQ4: Do students who use AWE more exhibit greater evidence of transfer?

Sample

Sample A - for RQs 1 and 2, including 431 students and 44 teachers in Grades 4–5 from 9 elementary schools for Fall 2018 and Spring 2019.
Sample B - for RQs 3 and 4, which selected students who had complete information for all the covariates and predictors across both time points.

Sample Demographics

Variable	Sample A (N = 431)		Sample B (N = 429)	
	Number	Percentage	Number	Percentage
Grade				
4	213	49.4	211	49.2
5	218	50.6	218	50.8
Gender				
Male	207	48.0	206	48.0
Female	224	52.0	223	52.0
Race				
African American	42	9.7	41	9.6
Asian	64	14.8	64	14.9
Hispanic/Latino	86	20.0	86	20.0
White	325	75.4	324	75.5
SPED	28	6.5	27	6.3
ELL	50	11.6	50	11.7

Note. Racial categories were not mutually exclusive, therefore, the percentages total to more than 100 percent.

Measures

Teacher Survey

The survey was based on national surveys used by Graham and colleagues (Cutler & Graham, 2008; Gilbert & Graham, 2010; Graham et al., 2014) and incorporated 32 Likert-like items probing the frequency with which teachers implemented various writing instruction practices.

Writing Quality

Prompts scored for holistic quality via PEG Overall Score (range = 6-30).
The PEG Score is formed as the sum of six traits, each measured on a 1-5 scale: development of ideas, organization, style, sentence structure, conventions, and word choice. PEG is highly reliable; quadratic weighted kappa of machine-human agreement average in the low .80s.

MI Write Usage Data

Student Level
Total number of unique essays students completed within MI Write across the school year;
Average number of drafts completed per essay (total number of drafts for the school year divided by the total number of unique essays for the school year) - the degree to which students, on average, revised their writing using automated feedback from MI Write, measured as the;
Interaction between the number of essays and the average number of drafts;
Total number of MI Write lesson minutes students completed for the school year.

Teacher Level
Total number of essays assigned by the teacher within MI Write;
Classroom average number of drafts per essay, which describes between-classroom variability in the extent that teachers utilized MI Write to facilitate drafting (total essays) and revising (drafts/essay).

Covariates

Covariates include students' demographics information (e.g., grade, race, special education status, and ELL status), measures of their reading ability (i.e., *HMH Reading Inventory* ranging from <100 to 1500+) and writing ability (i.e., *Smarter Balanced ELA and Writing* scale scores, ranging from 2000-3000).

Data Analysis

The present study adopted Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002) to examine growth in students' MI Write Overall Scores across successive drafts of an initial essay administered in Fall 2018 and a later essay administered in Spring 2019. We designed two sets of longitudinal growth curve models for the Fall 2018 and Spring 2019 essays and then used HLM gain score analyses to compare whether (a) students demonstrated gains in their average initial draft essay score (i.e., unaided transfer) and their rate of growth when revising (i.e., aided transfer); and (b) whether, after controlling for relevant covariates, those gains were predicted by the degree to which students and teachers utilized MI Write.

Teacher Survey Results

Fall 2018 Teacher Survey Report on Teachers' Writing Instruction (N = 17)

Item	*M* (*SD*)	Mdn	Mode
Time Devoted to Teaching Basic Writing Skills			
1. Time for teaching spelling	0.94 (0.97)	1	0
2. Time for teaching capitalization	1.88 (0.70)	2	2
3. Time for teaching punctuation	2.35 (0.49)	2	2
4. Time for teaching handwriting	1.12 (0.99)	1	1
5. Time for teaching keyboarding	0.47 (0.87)	0	0
6. Time for teaching sentence construction	2.53 (0.62)	2	2
Average Time for Teaching Basic Writing Skills	1.55 (0.55)	1.67	1.00
Time Devoted to Teaching Writing Strategies			
7. Time for modeling writing strategies using think-alouds	2.76 (0.66)	3	3
8. Time for teaching ways that different types of writing are organized	2.76 (0.66)	3	3
9. Time for teaching strategies for planning	2.88 (0.60)	3	3
10. Time for teaching strategies for revising	2.41 (0.62)	2	2
11. Time for using anchor (mentor) texts to teach writing	2.53 (0.87)	3	3
12. Time for teaching strategies for peer review	1.94 (0.83)	2	2
13. Time for teaching evaluation criteria	2.18 (0.64)	2	2
Average Time for Teaching Writing Strategies Scale	2.50 (0.48)	2.43	2.43
Frequency for Teaching Writing Processes			
14. Frequency for teaching planning	3.47 (0.87)	4	4
15. Frequency for teaching revising	3.82 (0.53)	4	4
16. Frequency for teaching giving or receiving peer review	2.53 (1.23)	3	3
17. Frequency for teaching conferencing with the teacher	3.18 (0.81)	3	4
18. Frequency for teaching using computers during writing	4.12 (0.78)	4	4
19. Frequency for teaching using graphic organizers	3.59 (0.71)	4	4
20. Frequency for teaching publishing their writing	2.24 (1.52)	2	1
21. Frequency for teaching assessing their own writing using a rubric	2.29 (1.10)	2	3
Average Frequency for Teaching Writing Processes	3.15 (0.59)	3.25	3.38
Frequency for Giving Writing Feedback			
22. Frequency for giving the students a grade	2.94 (1.20)	3	2
23. Frequency for writing general comments at the end of their paper	3.18 (1.42)	3	2
24. Frequency for providing ongoing comments directly within the body of students' writing	2.76 (1.39)	3	3
25. Frequency for correcting their spelling and grammar mistakes	2.59 (1.23)	2	2
26. Frequency for providing specific suggestions on reorganizing their ideas	3.12 (0.93)	3	3
27. Frequency for providing specific suggestions on idea development or elaboration/detail	3.06 (1.03)	3	2
28. Frequency for conferencing with students to review their writing	3.00 (1.17)	3	2
Average Frequency for Giving Writing Feedback	2.95 (0.91)	2.57	2.00
Frequency for Giving Certain Assignments			
29. Frequency for giving writing assignments that students have to complete within one class period	2.59 (1.12)	3	2
30. Frequency for giving writing assignments that are designed for students to work on over several class periods	2.71 (0.69)	3	3
31. Frequency for giving writing assignments that students have to complete with other students	1.82 (1.24)	2	3
32. Frequency for giving assignments of at least 4 or more paragraphs	2.06 (0.83)	2	2

Notes:
Scale for item 1- 13: No time: 0; Very little time: 1; Some time: 2; Quite a bit of time: 3; A lot of time: 4
Scale for item 14 - 21: Never: 0; Rarely: 1; Monthly: 2; More than once a month: 3; Weekly: 4; More than once a week: 5
Scale for item 22 - 28: Never or hardly ever: 0; Less than half of the time: 1; About half of the time: 2; More than half of the time: 3; Almost always: 4; Always: 5
Scale for item 29 - 32: Never: 0; About once or twice a year: 1; About once or twice a month: 2; About once or twice a week: 3; Every day or almost every day: 4

Quantitative Results

RQ1 - Shape and Rate of Growth in Writing Quality for Fall Essay

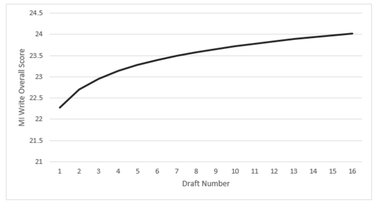

Figure 1. Unconditional Logarithmic Growth Model Predicting MI Write Overall Score for Fall Essay. Range of Y-axis range is restricted. Full range of the MI Write Overall Score is 6-30.

RQ2 - Shape and Rate of Growth in Writing Quality for Spring Essay

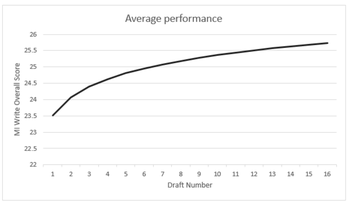

Figure 2. Unconditional Logarithmic Growth Model Predicting MI Write Overall Score for Spring Essay. Range of Y-axis range is restricted. Full range of the MI Write Overall Score is 6-30.

RQ 3 & 4 - Transfer Effects & AWE Usage Predicting Gains

Two-Level Model Predicting Gain in Average Initial Draft Scores

Fixed Effects	Model 1 – Unconditional Linear Model: *Coefficient* (S.E.)	*t*	Model 2 – Student Level Predictors: *Coefficient* (S.E.)	*t*	Model 3 – Student Level and Class-Level Predictors: *Coefficient* (S.E.)	*t*	*Effect Size*
Intercept	1.24 (0.92)	1.35	1.22 (0.88)	1.39	1.21 (0.82)	1.48	
	df = 43		df = 43		df = 41		
Level-1 Predictors							
Student Total Number of Essays			0.01 (0.02)	0.23	0.01 (0.02)	0.35	0.01
Student Average Drafts/Essay			0.05* (0.03)	1.86	0.05* (0.03)	1.88	0.03
Total Lesson Minutes (≥ 13 min)			-0.27* (0.12)	-2.29	-0.26* (0.12)	-2.28	-0.02*
Essays * Average Drafts/Essay			-0.004 (0.00)	-1.65	-0.004 (0.00)	-1.66	0.00
			df = 381		df = 381		
Level-2 Predictors							
Class Number of Assignments					-0.55** (0.19)	-2.94	-0.44**
Class Average Drafts/Essay					-0.64* (0.27)	-2.41	-0.32*
					df = 41		
Variance Components							
r Level-1 (students)	0.47		0.46		0.46		
u_0 Level-2 (teachers)	37.61***		37.37		32.42		
Goodness of Fit Statistics							
Deviance	1169.73	df = 2	1164.57	df = 2	1178.48	df = 2	
AIC	1173.73		1188.57		1182.48		
BIC	1181.85		1196.69		1190.60		
SBIC	1175.51		1190.35		1184.26		

Note. $^{\cdot}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Two-Level Model Predicting Gain in Average Growth Rate

Fixed Effects	Model 1 – Unconditional Linear Model: *Coefficient* (S.E.)	*t*	Model 2 – Student Level Predictors: *Coefficient* (S.E.)	*t*	Model 3 – Student Level and Class-Level Predictors: *Coefficient* (S.E.)	*t*	*Effect Size*
Intercept	0.41* (0.14)	2.93	0.44* (0.15)	3.02	0.44* (0.14)	3.02	
	df = 43		df = 43		df = 41		
Level-1 Predictors							
Student Total Number of Essays			-0.01* (0.00)	-2.07	-0.01* (0.00)	-1.99	-0.04*
Student Average Drafts/Essay			0.00 (0.01)	0.58	0.00 (0.01)	0.59	0.00
Total Lesson Minutes (≥ 13 min)			0.00 (0.03)	-0.06	0.00 (0.03)	-0.05	0.00
Essays * Average Drafts/Essay			0.00 (0.00)	0.04	0.00 (0.00)	0.03	0.00
			df = 381		df = 381		
Level-2 Predictors							
Class Number of Assignments					-0.03 (0.03)	-0.84	-0.15
Class Average Drafts/Essay					-0.01 (0.05)	-0.33	-0.03
					df = 41		
Variance Components							
r Level-1 (students)	0.04		0.04		0.04		
u_0 Level-2 (teachers)	0.87***		0.86***		0.89***		
Goodness of Fit Statistics							
Deviance	18.45	df = 2	49.96	df = 2	58.55	df = 2	
AIC	22.45		53.96		62.55		
BIC	30.57		62.08		70.67		
SBIC	24.23		55.74		64.33		

Note. $^{\cdot}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Post-Hoc Analysis

Demographic Comparisons Between the "Few Essay Group" and "Many Essay Group"

Variable	Few Essay Group (N = 73)		Many Essay Group (N = 115)		Chi Square test
	Number	Percentage	Number	Percentage	
Grade					15.84***
4	40	54.8	94	81.7	
5	33	45.2	21	18.3	
Gender					0.64
Male	38	52.1	53	46.1	
Female	35	47.9	62	53.9	
Race					5.37
African American	8	11.0	8	7.0	
Asian	7	9.6	15	13.0	
Hispanic/Latino	22	30.1	17	14.8	
White	58	79.5	92	80.0	
SPED	10	13.7	4	3.5	6.77**
ELL	11	15.1	13	11.3	0.57

Note. $^{**}p < 0.01$; $^{***}p < 0.001$.

Mean Comparisons Between the "Few Essay Group" and the "Many Essay Group" (T-Tests)

	Few Essay Group (N = 73) *M* (*SD*)	Many Essay Group (N = 115) *M* (*SD*)	Mean Difference	*t*	Standard Error of the Difference
Student Average Drafts/Essay	9.25 (5.37)	4.97 (2.71)	-4.27	-6.31***	0.68
Student Lesson Minutes (≥ 13mins)	0.18 (0.39)	0.54 (0.50)	0.36	5.56***	0.06
HMH Fall 2019 Score	754.89 (240.92)	806.55 (239.26)	51.66	1.44	35.90
SBELA 2018 Writing Claim Score	1.12 (0.58)	1.36 (0.58)	0.23	2.70**	0.09
Class Number of Assignments	6.18 (3.16)	13.37 (4.39)	7.20	13.04***	0.55
Class Average Drafts/Essay	7.25 (2.63)	4.54 (1.82)	-2.70	-8.35***	0.32
Gain in Initial Draft Scores between Fall 2018 and Spring 2019	1.14 (5.94)	1.04 (10.08)	-0.09	-0.08	1.17
Gain in Growth Rates between Fall 2018 and Spring 2019	1.04 (1.06)	0.01 (0.94)	-1.04	-7.01***	0.15

Note. $^{\cdot}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. SBELA 2018 = Smarter Balanced ELA test administered in Spring 2018. The reference group for the mean difference comparisons was the "Many Essay Group".

Conclusions

1. Three-level hierarchical growth curve modeling revealed a non-linear, decelerating shape of growth in writing performance across successive revisions to the Fall and Spring essays.
2. Two-level hierarchical gain score modeling indicated no unaided transfer effect for average initial draft scores, but an aided transfer effect for students' gain in growth rates. The post-hoc analysis showed students who completed many essays, but few drafts made lesser gains.
3. Teachers mainly used the process approach to writing instruction, but there was variability among the sample. It is plausible that differences in teachers' instructional approach, such as their adherence to the process-writing approach, led to different ways of implementing MI Write that, in turn, were associated with effects on students' gains in writing proficiency.

Acknowledgements The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H170046 to the University of Delaware. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.