Final Report

Dec. 1, 2014

Title: Predicting Writing Success Using Multilevel Modeling

Project personnel: Jennifer Lindstrom, Ph.D., Lindsay Peaster, Ph.D., Monica Semrad, M.S., Betsy Short, Ed.D.

Purpose

The purpose of this study was to systematically examine key predictor variables to determine their validity for predicting initial status and growth on a standardized test of writing among first and second graders. This multi-level study examined the validity of student-level variables (e.g., pre-test performance, classification status) and classroom-level variables (e.g., teacher type, level of training in Thinking Maps® and Write…From the Beginning®) in predicting initial and end-of-year writing performance. To capture students' writing growth we used multilevel modeling procedures that examine not only end-of-year writing performance but also predictors of growth across the school year. The following research questions were investigated:

Research Questions

1. To what extent does students' pre-test writing performance predict post-test writing performance?
2. To what extent does classification status (i.e., Specific Learning Disability, Significant Developmental Delay, Autism, Early Intervention Program, Gifted, English Learner) predict writing performance?

3. To what extent does level of teacher training (general overview; 1-day training; week-long training) in Thinking Maps® and Write…From the Beginning® predict writing performance?
4. To what extent does teacher type (general education, Special Education, Early Intervention Program, Gifted, English Learner) predict writing performance?

Methods

Participants and Setting

The study took place at Morgan County Primary School (MCPS) in the Morgan County Charter School System, located in rural central Georgia. The school houses all of the Pre-K through 2nd grade students in the county (approximately 880 students). The school is eligible to receive Title I and Title II funding. All first (N=239) and second (N=202) graders attending MCPS participated in the study. Of the first grade participants, 136 (56%) had no diagnosis or Speech Impairment (SI) only, 78 (33%) were classified as English Learner (EL), early intervention program (EIP), or EL/EIP, 10 (.04%) were classified as Specific Learning Disability/Speech Impairment (SLD/SI) or Significant Developmental Delay/Speech Impairment (SDD/SI), 9 (.03%) were classified as SLD or SDD, and 6 (.025%) were gifted. Among the second grade participants, 129 (64%) had no diagnosis or SI only, 67 (33%) were classified as EL, EIP, or EL/EIP, none were classified as SLD/SI or SDD/SI, 2 (.01%) were classified as SLD or SDD, and 4 (.02%) were gifted. Information on classification status for the sample is provided in Table 1.

Table 1

Number of Participants by Classification

Classification Number	Classification	Grade 1	Grade 2
1	None; SI	136	129
2	Gifted	6	4
3	SLD; SDD	9	0
4	EL; EIP; EL/EIP	78	67
5	SLD/SI; SDD/SI	10	2
	Total	239	202

In addition, all first and second grade teachers (N=27) participated in the study, including general education (n=21), EIP (n=1), English as a Second Language (ESOL) (n=1), gifted (n=1), and Special Education (SPED) (n=3) teachers. Among the 27 teacher participants, 2 held a bachelor's degree, 12 held a master's degree, 9 had earned a specialist's degree, and 1 had a doctorate degree (the highest degree earned was unknown for 4 teachers). Teaching experience among the participants ranged from four years to 34 years. The type and level of training each of the first and second grade teachers had received in Thinking Maps® and Write…From the Beginning® varied across participants. For instance, all teachers participated in the Thinking Maps®: A Language for Learning workshop (13 hours) in the fall of 2007, and four faculty attended the Thinking Maps®: Train the Trainer workshop in 2008. Similarly, in the fall of 2008,

all first and second grade teachers participated in the Thinking Maps® – Write…From the Beginning® workshop (10 hours), and then in 2012, four faculty participated in the Write…From the Beginning®: Train the Trainer workshop (24 hours). Teachers who joined the faculty after 2008 received an overview of one or both programs by the school review team and have since been given all of the manuals and corresponding materials. For the purposes of the analyses, level of teacher training was ranked 0-5 (0=none; 1=1-10 hours; 2 = 11-15 hours; 3 = 16-20 hours; 4 = 21-25 hours; 5 = 25+ hours). Information on teacher title, highest degree earned, teaching experience, and level of training in Thinking Maps® and Write…From the Beginning® is presented in Table 2.

Table 2:

Teacher Characteristics

Teacher Characteristics	*n*
Title	
General Education	21
EIP	1
ESOL	1
Gifted	1
SPED	3
Highest Degree Earned	
Bachelor's	2
Master's	12
Specialist	9
Doctorate	1
Unknown	4

Teaching Experience	
4-10 years	8
11-20 years	7
20+ years	10
Unknown	2
Level of Training	
None	2
1-10 hours	3
11-15 hours	2
16-20 hours	18
21-25 hours	1
26+ hours	1

Measure

Participants' scores from a standardized mock Writing Assessment, Georgia's performance-based writing assessments, were used for analyses. The Writing Assessment was group-administered to all first and second graders in mid-August (2013) and again in mid-April (2014) during students' Language Arts class. Student writing samples for each administration were sent to the Georgia Center for Assessment (GCA; http://gca.coe.uga.edu/) to be independently scored by trained raters. Student writing was assessed analytically by GCA raters in two domains (*Ideas* and *Organization*) for first grade and four domains (*Ideas, Organization,*

Style, and *Conventions*) for second grade. Students were asked to produce a "narrative" response to the prompt: "Tell about something you did this summer." The scoring categories for each of the domains were "Developing", "In Progress", or "Meets" for first grade and "Does Not Meet" or "Meets" for second grade.

Results

Descriptives: First Grade

The number of students in each category on the pre- and posttest Mock Writing Test (*Ideas)* are presented in Tables 3 (total) and 4 (by classification type).

Table 3:

Number of 1st Graders in Each Category - Ideas

Score	Pretest	Posttest	Difference
Developing	n = 56	n=18	-38
In Progress	n = 123	n=85	+38
Meets	n = 28	n=136	+108

Overall, significant gains were made in first graders' performance in the *Ideas* domain of the Mock Writing Test. The vast majority – 92 percent – of first graders scored in the "In Progress" (35%) or "Meets" (57%) category at posttest compared to only 63 percent at pretest.

Table 4

Number of 1st Graders in Each Category by Classification Status - Ideas

Score	None/SI		Gifted		SLD; SDD		EL; EIP; EL/ EIP		SLD/SI; SDD/ SI	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Developing	13	4	0	0	5	3	30	10	8	1
In Progress	87	42	3	1	3	0	29	36	1	6
Meets	22	90	3	5	1	6	2	32	0	3
% In Progress or Meets	89	97	100	100	44	67	51	87	11	90

When considering changes in students' performance in the *Ideas* domain across disability categories, the groups showing the most significant improvement included students classified as SLD or SDD (23% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to posttest), EL, EIP, or EL/EIP (36% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to posttest) and students classified as SLD/SI or SDD/SI (79% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to post-test).

The number of students in each category on the pre- and posttest Mock Writing Test (*Organization*) are presented in Tables 5 (total) and 6 (by classification type).

Table 5

Number of 1st Graders in Each Category - Organization

Score	Pretest	Posttest	Difference
Developing	n = 136	n=21	-115
In Progress	n = 63	n=124	+61
Meets	n = 8	n=94	+86

Considerable gains were also noted in first graders' performance in the *Organization* domain of the Mock Writing Test. The majority (91%) of first graders scored in the "In Progress" (52%) or "Meets" (30%) category at posttest compared to only 63 percent at pretest.

Table 6

Number of 1st Graders in Each Category by Classification Status - Organization

Score	None/SI		Gifted		SLD; SDD		EL; EIP; EL/ EIP		SLD/SI; SDD/ SI	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Developing	69	7	2	0	5	3	51	10	9	1
In Progress	47	60	4	1	3	3	9	54	0	6
Meets	6	69	0	5	1	3	1	14	0	3
% In Progress or Meets	43	94	67	100	44	67	16	87	0	100

Within the *Organization* domain, significant improvements were observed across all groups. The groups showing the most significant improvement included students with no disability or SI only (51% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to posttest), those classified as EL, EIP, or EL/EIP (71% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to posttest) and students classified as SLD/SI or SDD/SI (100% increase in the number of students who scored in the "In Progress" or "Meets" category from pre- to posttest). Students in the Gifted category improved as well (33% increase in the number of gifted students who scored in the "In Progress" or "Meets" category at posttest).

Overall, first graders showed significant improvements in their performance in the *Ideas* and *Organization* domains of the Mock Writing Test, regardless of classification status. However, the most significant improvements by far were observed among students classified as SLD/SI or SDD/SI, with 90 to 100 percent of the students in this group scoring in the "In Progress" or "Meets" category in both domains at posttest compared to <1 percent at pretest.

Descriptives: Second Grade

The number of students in each category on the pre- and posttest Mock Writing Test (*Ideas*) are presented in Tables 7 (total) and 8 (by classification type).

Table 7

Number of 2nd Graders in Each Category - Ideas

Score	Pretest	Posttest	Difference
Does Not Meet	n = 132	n= 91	-41

Meets	n = 44	n= 106	+62

Overall, notable gains were made in second graders' performance in the *Ideas* domain of the Mock Writing Test, though not quite as significant as the gains observed among first graders in this domain. Fifty-four percent scored in the "Meets" category at posttest compared to only 25 percent at pretest.

Table 8

Number of 2nd Graders in Each Category by Classification Status - Ideas

Score	None/SI		Gifted		SLD/SI; SDD/SI		EL; EIP; EL/EIP	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Does Not Meet	70	52	2	0	2	2	58	37
Meets	39	73	2	4	0	0	3	29
% Meets	36	58	50	100	0	0	4	44

Within the *Ideas* domain, moderate improvements were observed across all groups except one. The groups showing the most significant improvement included students classified as Gifted (50% increase in the number of students who scored in the "Meets" category from pre- to posttest) and those classified as EL, EIP, or EL/EIP (40% increase in the number of students who

scored in the “Meets” category from pre- to posttest). As a group, students with no diagnosis or SI only showed some improvements as well (22% increase in the number of gifted students who scored in the “In Progress” or “Meets” category at posttest). However, compared to first graders, the degree of improvement in the *Ideas* domain across all groups was considerably less significant.

The number of students in each category on the pre- and posttest Mock Writing Test (*Organization*) are presented in Tables 9 (total) and 10 (by classification type).

Table 9

Number of 2nd Graders in Each Category - Organization

Score	Pretest	Posttest	Difference
Does Not Meet	n = 144	n= 90	-54
Meets	n = 32	n= 107	+75

Similar gains were made in second graders’ performance in the *Organization* domain of the Mock Writing Test. Fifty-four percent scored in the “Meets” category at posttest compared to only 18 percent at pretest.

Table 10

Number of 2nd Graders in Each Category by Classification Status - Organization

Score	None/SI		Gifted		SLD/SI; SDD/SI		EL; EIP; EL/EIP	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Does Not Meet	82	50	2	0	2	2	58	38
Meets	27	75	2	4	0	0	3	28

% Meets	25	60	50	100	0	0	4	57

Within the *Organization* domain, moderate improvements were again observed across all groups except one. The groups showing the most significant improvement included students classified as Gifted (50% increase in the number of students who scored in the "Meets" category from pre- to posttest) and those classified as EL, EIP, or EL/EIP (53% increase in the number of students who scored in the "Meets" category from pre- to posttest). As a group, students with no diagnosis or SI showed some improvement as well (35% increase in the number of Gifted students who scored in the "In Progress" or "Meets" category at posttest). There were no changes observed among students classified as SLD/SI or SDD/SI from pre- to posttest. And similar to the *Ideas* domain, compared to first graders, the degree of improvement in the *Organization* domain across all groups was considerably less notable.

The number of students in each category on the pre- and posttest Mock Writing Test (*Style*) *a*re presented in Tables 11 (total) and 12 (by classification type).

Table 11

Number of 2nd Graders in Each Category - Style

Score	Pretest	Posttest	Difference
Does Not Meet	n = 156	n= 105	-51
Meets	n = 19	n= 92	+73

Some second graders' performance in the *Style* domain of the Mock Writing Test did improve from pretest to posttest, however, the majority (53%) of second graders still performed in the "Does Not Meet" category at posttest.

Table 12

Number of 2nd Graders in Each Category by Classification Status - Style

Score	None/SI		Gifted		SLD/SI; SDD/SI		EL; EIP; EL/EIP	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Does Not Meet	92	57	3	0	2	2	59	46
Meets	16	68	1	4	0	0	2	20
% Meets	15	54	25	100	0	0	3	30

Moderate to significant improvements were seen within the *Style* domain across all groups except one. The groups showing the most significant improvement included students classified as Gifted (75% increase in the number of students who scored in the "Meets" category from pre- to posttest), those with no diagnosis or SI only (39% increase in the number of students who scored in the "In Progress" or "Meets" category at posttest), and those classified as EL, EIP, or EL/EIP (27% increase in the number of students who scored in the "Meets" category from pre- to posttest). There were no changes observed among students classified as SLD/SI or SDD/SI from pre- to posttest. Compared to the first two domains (*Ideas* and *Organization*), there was much more significant improvements in *Style* among second graders classified as Gifted or with no diagnosis than those with any type of diagnosis.

The number of students in each category on the pre- and posttest Mock Writing Test (*Conventions*) are presented in Tables 13 (total) and 14 (by classification type).

Table 13

Number of 2nd Graders in Each Category - Conventions

Score	Pretest	Posttest	Difference
Does Not Meet	n = 144	n= 120	-24
Meets	n = 25	n= 71	+46

While a small percentage of second graders improved in the *Conventions* domain of the Mock Writing Test, only 37 percent of second graders performed in the "Meets" category at posttest. This is noticeably different than the trends that were observed in the former three domains.

Table 14

Number of 2nd Graders in Each Category by Classification Status - Conventions

Score	None/SI		Gifted		SLD/SI; SDD/SI		EL; EIP; EL/EIP	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Does Not Meet	84	66	2	3	2	2	60	54
Meets	24	59	2	1	0	0	1	12
% Meets	22	47	50	25	0	0	1	22

Within the *Conventions* domain, little to no improvement was observed, which is in stark contrast to the improvements that were observed in the *Ideas* and *Organization* domains in particular, and somewhat in the *Style* domain. Specifically, only two groups showed slight

improvements – students with no diagnosis or SI only (25% increase in the number of students who scored in the "In Progress" or "Meets" category at posttest) and students classified as EL, EIP, or EL/EIP (21% increase in the number of students who scored in the "Meets" category from pre- to posttest). There was also a 25% decrease in the number of Gifted students who scored in the "Meets" category, a trend that was not observed in any of the other three domains. There were no changes observed among students classified as SLD/SI or SDD/SI from pre- to posttest. Noticeably weaker performance across all groups in the *Conventions* domain suggests the focus of instruction was geared more towards the higher level processes involved in writing – ideation, organization, and style – compared to the mechanical elements of writing.

Similar to first graders, overall, second graders showed significant improvements in their performance in the *Ideas* and *Organization* domains of the Mock Writing Test, regardless of classification status. However, the most significant improvements by far were observed among students classified as SLD/SI or SDD/SI, with 90 to 100 percent of the students in this group scoring in the "In Progress" or "Meets" category in both domains at posttest compared to <1 percent at pretest.

Statistical Analyses and Model Descriptions

A repeated measures generalized linear mixed model (GLMM; e.g., see Berridge and Crouchley, 2011) was used to evaluate the research questions. The MIXED procedure in SAS 9.4 was used implement the repeated measures GLMM.

Grade 1 GLMM. The first GLMM used to quantify the impact of training on the change in first graders' writing scores, when controlling for teacher type, student classification, and maximum training level was specified as:

(1)

Where is the writing test score on testing occasion for participant *p* where for developing, for in-progress, and for meets; is a dummy-coded predictor where equals 1 for the post-test occasion and equals 0 for the pre-test occasion; is a dummy-coded predictor where equals 0 for general teachers and 1 for other teachers; the maximum level of training a teacher received (ranging from 0-5); and is a dummy coded predictor where =1 when student classification = c. There are five types of student classifications (see Table 1). is the intercept of the model for response (i.e., the log-odds of scoring over 5 for the baseline testing [for pre-test with general teacher and 0 training and student in last classification] occasion and reference group); is the regression coefficient for testing occasion effect on the writing score; is the regression coefficient for the teacher type effect on the writing score; is the regression coefficient for training effect on the writing score; and is the regression coefficient for student classification effect on the writing score; is the two-way interaction term for testing occasion and teacher type; is the two-way interaction term for testing occasion and maximum training level; is the two-way interaction term for testing occasion and student classification ; and is the random intercept for a participant such that .

Results: Ideas domain. Results showed that pre-test performance is a significant predictor of post-test performance ($F(1,198) = 18.30$, $p < .0001$). Post-test scores also differed by

student classification, when controlling for teacher type and maximum training level ($F(4,198) = 8.79$, $p < .0001$), although student classification did not significantly impact the degree of change in writing scores from pre- to post-test, controlling for student classification and teacher type ($F(4,198) = 1.25$, $p = .2898$). Similarly, post-test scores did differ by teacher type, when controlling for student classification and maximum teacher training level ($F(1,198) = 3.93$, $p < .05$), although teacher type did not significantly impact writing score gains from pre- to post-test, controlling for student classification and maximum training level ($F(4,198) = .41$, $p = .5228$). Finally, within the *Ideas* domain, the level of teacher training did not significantly impact post-test performance ($F(1,198) = .24$, $p = .6222$) or pre- to post-test writing gains ($F(1,198) = .19$, $p = .6655$) , when controlling for student classification and teacher type. Because significant post-test differences were found across the five classification categories, pairwise comparisons by classification status from pre- to post-test were conducted. However, results indicated that no significant differences in gain scores were observed between any of the five classification groups. The only groups approaching significant differences in their pre- to post-test gain scores were group 1 (no diagnosis or SI only) and group 5 (students classified as SLD/SI; SDD/SI) ($t(198)=1.86$, $p = .0645$).

Results: Organization domain. Results showed that classification status, when controlling for teacher type and maximum training level, is a significant predictor of post-test performance ($F(4,198) = 3.29$, $p < .05$). No other predictor variables (e.g., pre-test performance, teacher type, maximum teacher training) significantly impacted post-test performance. With regard to gain scores, teacher type significantly impacted change in writing scores, when controlling for student classification and maximum teacher training level ($F(1,198) = 5.89$, $p < .$

05). Because significant post-test differences within the *Organization* domain were found across the five classification categories, pairwise comparisons by classification status from pre- to post-test were conducted. Results indicated that of the 10 comparisons that were made, significant differences in gain scores were observed between the classification groups: Group 1 (no diagnosis or SI only) and group 3 (SLD; SDD) ($t(198)= -2.19$, $p < .05$); Group 2 (Gifted) and Group 3 ($t(198)= -2.05$, $p < .05$); and Group 3 and Group 4 (EL; EIP; EL/EIP) ($t(198)= 2.07$, $p < .05$).

Grade 2 GLMM. The second GLMM used to quantify the impact of training on the change in second graders' writing scores, when controlling for teacher type, student classification, and maximum training was specified as:

(2)

Where is the writing test score on testing occasion for participant *p* where for does not meet and for meets; is a dummy-coded predictor where equals 1 for the post-test occasion and equals 0 for the pre-test occasion; is a dummy-coded predictor where equals 0 for general teachers and 1 for other teachers; the maximum level of training a teacher received (ranging from 0-5); and is a dummy coded predictor where =1 when student classification = c. There are five types of student classifications (see Table 1). is the intercept of the model (i.e., the log-odds of scoring 0 over 1 for the baseline testing [for pre-test with general teacher and 0 training and student in last classification] occasion and reference group); is the regression coefficient for testing occasion effect on the writing score; is the regression coefficient for the teacher type effect on the writing score; is the regression coefficient for training effect on the writing score;

and is the regression coefficient for student classification effect on the writing score; is the two-way interaction term for testing occasion and teacher type; is the two-way interaction term for testing occasion and maximum training level; is the two-way interaction term for testing occasion and student classification ; and is the random intercept for a participant such that .

Results: Ideas domain. Results showed that classification status, when controlling for teacher type and maximum training level, is a significant predictor of post-test performance ($F(2,165) = 6.23$, $p < .01$). Maximum teacher training also significantly impacted post-test performance, when controlling for student classification and teacher type ($F(1,165) = 5.84$, $p < .05$). Pre-test performance did not significantly impact post-test performance in the *Ideas* domain. With regard to gain scores, classification status significantly impacted change in writing scores, when controlling for teacher type and maximum teacher training level ($F(2,165) = 3.16$, $p < .05$). Because significant post-test differences within the *Ideas* domain were found across the four (no second graders were in Group 3) classification categories, pairwise comparisons by classification status from pre- to post-test were conducted. Results indicated that significant differences in gain scores were observed between the following classification groups: Group 1 (no diagnosis or SI only) and group 5 (SLD/SI; SDD/SI) ($t(165)= -12.46$, $p < .0001$); Group 1 and Group 4 (EL; EIP; EL/EIP) ($t(165)= 2.51$, $p < .05$); and Group 4 and Group 5 ($t(165)=$ Infinity, $p < .0001$).

Results: Organization domain. Similar to the *Ideas* domain, results of the *Organization* domain showed that classification status, when controlling for teacher type and maximum training level, is a significant predictor of second graders' post-test performance ($F(2,165) = 3.74$, $p < .05$). Maximum teacher training also significantly impacted post-test performance,

when controlling for student classification and teacher type ($F(1,165) = 8.56$, $p < .05$). Pre-test performance did not significantly impact post-test performance in the *Organization* domain. With regard to gain scores, maximum level of teacher training significantly impacted change in writing scores, when controlling for classification status and teacher type ($F(1,165) = 5.25$, $p < .05$). Because significant post-test differences within the *Organization* domain were found across the classification categories, pairwise comparisons by classification status from pre- to post-test were conducted. Results indicated that significant differences in gain scores were observed between the following classification groups: Group 1 (no diagnosis or SI only) and group 5 (SLD/SI; SDD/SI) ($t(165)= -13.46$, $p < .0001$); and Group 4 (EL; EIP; EL/EIP) and Group 5 ($t(165)=$ Infinity, $p < .0001$).

Results: Style domain. Results of the *Style* domain showed that maximum level of teacher training, when controlling for classification status and teacher type, is a significant predictor of second graders' post-test performance ($F(1,164) = 6.50$, $p < .05$). None of the other predictor variables (pre-test performance, classification status, teacher type) significantly impacted post-test performance in the *Style* domain. With regard to pre- to post-test gain scores, none of the predictor variables significantly impacted change in writing scores. Pairwise comparisons were conducted by classification status from pre- to post-test, and showed that significant differences in gain scores were observed between the following classification groups: Group 1 (no diagnosis or SI only) and group 5 (SLD/SI; SDD/SI) ($t(164)= -8.08$, $p < .0001$); and Group 4 (EL; EIP; EL/EIP) and Group 5 ($t(164)=$ Infinity, $p < .0001$).

Results: Conventions domain. Results of the *Style* domain showed that pre-test performance is a significant predictor of post-test performance ($F(1,160) = 6.08$, $p < .05$). Likewise, student classification status, when controlling for teacher type and maximum level of training, is a significant predictor of second graders' post-test performance ($F(1,160) = 7.35$, $p < .05$). As a predictor variable, teacher type had minimal impact on post-test performance in the *Style* domain. With regard to pre- to post-test gain scores, only one predictor variable, teacher type, impacted the degree of gain students made from pre- to post-test in the *Conventions* domain. Pairwise comparisons by classification type could not be conducted in this domain.