2012

Development of Writing: Key Components of Written Language

Patricia Thatcher Kantor

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

DEVELOPMENT OF WRITING: KEY COMPONENTS OF WRITTEN LANGUAGE

By

PATRICIA THATCHER KANTOR

A Dissertation submitted to the
Department of Psychology
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2012

Patricia Thatcher Kantor defended this dissertation on October 29, 2012.
The members of the supervisory committee were:

Richard K. Wagner
Professor Directing Dissertation

Barbara Foorman
University Representative

Christopher Schatschneider
Committee Member

Jon Maner
Committee Member

Young-Suk Kim
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

I dedicate this work to my husband for his unwavering encouragement throughout my graduate school experience, to my children whose love knows no bounds, and to my parents for always encouraging me to reach my full potential.

ACKNOWLEDGEMENTS

I would like to thank Rick Wagner for his guidance, mentorship, and patience throughout my graduate career. I could not have asked for a better advisor. I also thank the members of my doctoral committee for their valuable input and support throughout this project.

I thank the members of Team Wagner for their role in data collection, entry, and coding. I particularly thank Danielle Lopez for her organization and management of the project. Additionally, I thank Yusra Ahmed for her assistance in submitting writing samples to Coh-Metrix.

Support for this research was provided by grant P50 HD052120 from the National Institute of Child Health and Human Development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study utilized confirmatory factor analyses and latent change score analyses to model individual and developmental differences in a longitudinal study of children's writing. Participants were 158 children who completed a writing sample each year from 1st through 4th grade. At all four time points, a four-factor model of writing provided the best fit to the data. The factors were macro-organization (presence of topic sentence, number of key elements, and order of ideas), productivity (number of words and number of unique words), complexity (average number or words per sentence and number of connectives), and vocabulary (average number of syllables and average number of characters per word, and percentage of multisyllabic words). The latent change score analyses demonstrated significant relations among the intercepts of macro-organization, productivity, and complexity factors, indicating that children with higher initial levels of one skill were also likely to have higher initial levels of the other. Productivity was also identified as a leading indicator of complexity, such that higher levels of productivity predicted subsequent increases in complexity over time.

INTRODUCTION

Writing is an important skill for all learners. It allows for communication with others across time and space (Graham & Perin, 2007a). Writing can be used to explain, persuade, or convey experience, or as a form of self-expression. In an academic setting, writing is utilized as a record of what one knows or thinks about a particular topic. Learning to write is a remarkable achievement as writing utilizes multiple cognitive resources as the writer generates ideas, organizes them, executes the physical acts of writing, and makes revisions (Kulikowich, Mason, & Brown, 2008).

Although the reading aspect of literacy has received much of the attention in scholarship and policy (e.g., National Institute of Child Health and Human Development, 2000; the Reading First component of No Child Left Behind, 2001), there has recently been an increased focus on writing research based on calls for states to teach writing based on best practices (National Commission on Writing in America's Schools and Colleges, 2003; Riley, 1996; Strickland et al., 2001). The Common Core State Standards Initiative, which has been adopted by nearly all states, includes a framework of writing knowledge and skills that students are expected to learn. Utilizing the standards, it is expected that each year students will increase their writing skills, including syntax, vocabulary, organization of ideas, planning, revising, and editing, as well as use of content and sources (National Governors Association for Best Practices, Council for of Chief State School Officers, 2010). The Writing Next report (Graham & Perin, 2007b) recommended a number of strategies for improving writing instruction in adolescents based on a large-scale review of the research on writing instruction, such as collaborative writing, prewriting, and study of models of good writing. Additionally, the *Writing to Read* report concluded that writing techniques and practice can be utilized to improve reading ability (Graham & Hebert, 2010).

The most recent results of the National Assessment of Educational Progress (NAEP) lend further support for an increase in focus on writing. According to the NAEP results, the majority of students do not write well enough to meet the expectations of higher academics and the workforce. More specifically, in grades 4 (Persky, Daane, & Jin, 2003), 8, and 12 (2011), less than one-third of students were at or above the “proficient” level, which indicates “solid academic performance.” Furthermore, there continues to be significant achievement gaps based on race, gender, and socioeconomic status.

Struggling writers are at a particular disadvantage in academic settings as this is one of the primary means of assessment of content knowledge (Graham & Perin, 2007a). Furthermore, struggling writers are likely to face difficulties in college and in the workforce where writing skills play an important role in hiring and promotion decisions (National Commission on Writing for America's Families, Schools, and Colleges, 2004, 2005). Identification of struggling writers is thus an important and necessary endeavor.

The majority of tests currently in use, including statewide tests, utilize holistic scoring that provides a single score based on a rubric (Huot & Neel, 2006). Although this method identifies general proficiency of writing, it does not capture the complex nature of writing and is insufficient for targeting individualized instruction and intervention needs (Nelson & Van Meter, 2007). A more useful writing assessment tool is one that identifies specific developmental differences in core components of writing and distinguishes typical versus atypical writing performance. A first step in developing such a tool is to determine the key components in written composition and development relations among them.

Components of Writing

In previous research on the factors of writing, a variety of variables has been used to measure quality of writing, including measures of writing productivity (e.g. number of words or sentences), spelling and grammar, organization, vocabulary, ideas and content, and overall quality.

A1961 study was one of the first to attempt to classify the important features of writing. In this study, 11,000 reader comments on 3,557 essays were analyzed and five main factors were identified that explained 43% of the variance in scores (Diederich, 1974). The first and largest factor was ideas expressed, with comments focusing on clarity, development, and relevance to the topic and purpose. The next factor identified was mechanics, which consisted of errors in usage, sentence structure, punctuation, and spelling. The third factor was organization and analysis. Wording and phrasing made up the fourth factor and included the choice and arrangement of words. Diederich noted that this may be thought of a vocabulary factor. The final factor was termed "flavor" or style and consisted of the personal qualities revealed in the writing, such as individuality, originality, and interest.

In their study of reading and writing relations, Mehta, Foorman, Branum-Martin, and Taylor (2005) developed a writing ability estimate. The estimate was calculated using a two-parameter graded response model that was based on quality ratings (0 = poor to 4 = excellent) of eight aspects of writing, including: addressing the prompt, unity and logical organization, vocabulary usage, sentence completion, grammar usage, use of capitalization, use of punctuation marks, and spelling conventions.

The NAEP includes writing assessments that are given to students in grades 4, 8, and 12. Scoring is based on the NAEP writing framework which includes objectives for student writing. The NAEP uses a focused holistic approach to scoring. A scoring rubric is established for each of the communicative purposes that are assessed by NAEP: explanation, persuasion, and conveying experience. The main domains that are considered in the NAEP writing scores are development of ideas (effective depth and complexity of ideas, specific use of details and examples), organization of ideas (effective text structure, coherence, focus), and language facility and use of conventions (sentence structure, word choice, voice and tone, grammar, usage, and mechanics) (ACT, Inc., 2007). Beginning with the 2011 assessment, NAEP writing has utilized computerized assessment for grades 8 and 12.

A number of researchers have measured writing using a levels of language approach that considers writing at three levels: discourse, sentence, and word (e.g., Abbott, Berninger, & Fayol 2010; Nelson & Van Meter, 2007; Puranik et al., 2008, Sanders & Schilperoord, 2006; Whitaker, Berninger, Johnston, & Swanson, 1994). Discourse level features are often scored using holistic scoring methods that emphasize the organization, ideas and content, sentence fluency, and conventions of the writing sample. Sentence level measures include sentence length, counts of connectives, number of grammatical errors, clause density, and number of T-units. Word level features include vocabulary diversity, measured as number of different words, as well as letter formation elements and spelling (Nelson & Van Meter, 2007).

The popular 6+1 Trait Writing system includes seven domains of writing: ideas, organization, voice, word choice, sentence fluency, conventions, with the addition of presentation as the seventh trait. This system was developed by Education Northwest more than 20 years ago, is used in nearly every state in the U.S., as well as numerous other countries, and is the source model used to score papers in numerous state and district assessments across the U.S. (Education Northwest, 2012). Proponents of the 6+1 Trait model report that use of the system increases teacher and student understanding of the components of good writing, creates a shared vocabulary for discussion of writing, and aids in accurate assessment (Culham, 2003).

The use of more sophisticated analysis techniques, such as factor analysis and structural equation modeling, has also been included in several studies of writing. These techniques are useful for testing alternatives models of individual and developmental differences and have recently been included in studies of writing.

Exploratory factor analyses were used to explore the internal structure of written composition across grade levels in a study that utilized samples of writing from students in grades, 4, 6, 8, 10, and 12 (Attali & Powers, 2008). Compositions were scored using the e-rater V.2, which scores essays on seven features: essay length, style, grammar, usage, mechanics, vocabulary, and word length. However, the initial EFA results were not supportive of a single scale as indicated significant differences in the communalities at the feature level. Consequently, two- and three- factor solutions were explored. Based on these results, a three-factor solution consisting of fluency (essay length and style), sentence-level conventions (grammar, usage, and mechanics), and word choice (vocabulary and word length) seems most appropriate for higher grade levels (8, 10, and 12). In lower grade levels (4 and 6), a two-factor solution with combined fluency and conventions seemed most appropriate. When multiple-group confirmatory factor analyses were conducted, both the two- and three-factor solutions showed reasonable fit. However, the expected cross-validation index (ECVI) was greater for the three-factor solution and thus it was determined to be the better solution.

Also using exploratory factor analysis, Puranik, Lombardino, and Altmann (2008) determined that a framework of three factors, productivity, complexity, and accuracy best represented the writing of a study of 120 children in grades 3 through 6. The productivity factor consisted of total number of words, total number of ideas, number of t-units, and number of clauses. Complexity was made up of mean length of t-unit and clause density. Accuracy included percentage of spelling errors, writing conventions (use of periods and initial capital letters), and percentage of grammatical t-units.

In an expansion of Puranik et al.'s (2008) findings, Wagner, Puranik, Foorman, Foster, Wilson, Tschinkel, and Kantor (2011) utilized confirmatory factor analyses and explored the addition of a macro-organization factor. Alternative models of written composition and handwriting fluency were compared utilizing writing samples from 98 first- and 88 fourth-grade students. They found that for both groups of children a four-factor of model of written composition plus a handwriting fluency factor best fit the data. Written composition consisted of macro-organization, complexity, productivity, and spelling and punctuation. Macro-organization included whether a topic sentence was present, rated logical ordering of ideas, and number of key elements present (i.e., main idea, body, and conclusion). Complexity was represented by mean length of T-unit and clause density. The productivity factor consisted of total number of words and number of different words. Spelling and punctuation included number of spelling errors, number of capitalization errors, and number of punctuation errors involving correct placement of a period. Wagner et al. further concluded that the largest developmental differences between the two grades as measured by effect sizes were found for productivity and handwriting fluency, followed by complexity and macro-organization. Only minimal differences were found for spelling and punctuation. The same four factor plus handwriting fluency model was also found in a sample of Chinese children (Guan, Ye, Wagner, & Meng, 2012).

Goals of the Present Study

The current study extends the four factor model of writing found by Wagner et al. (2011) in two ways: expanding the model through the addition of another writing factor and examining the developmental nature of the writing constructs through the use of a longitudinal sample.

To expand the model, the potential addition of a vocabulary/word choice factor, was explored. Given the emphasis on vocabulary in Common Core, 6+1 Trait Writing, NAEP, and elsewhere, this appeared to be an important aspect of writing that was not featured in the previous study. In their study, Olinghouse and Leaird (2009) found that three measures of vocabulary (vocabulary diversity, less frequent vocabulary, and mean syllable length) were significantly correlated with writing quality in second grade and that the same three measures, plus number of polysyllabic words were significantly correlated in fourth grade for the picture prompt writing. Additionally, all four measures of vocabulary were significantly correlated with standard scores on the TOWL-3 in second grade and vocabulary diversity and less frequent vocabulary were significantly correlated in fourth grade. It was further found that vocabulary measures, especially vocabulary diversity, were significant predictors of writing quality beyond compositional length and spelling. Additionally, a variety of measures of vocabulary have been used in other studies of writing (e.g. Mehta et al., 2005; Nelson and Van Meter 2007) though none included them as a separate latent variable. However, another scenario might be that children who are more prolific writers also have a larger bank of words to choose from and so vocabulary and productivity may best be described as a single factor (e.g. Wagner et al., 2011).

Thus, confirmatory factor analyses were used to examine whether vocabulary is best modeled as a separate factor of writing or if the measures should be included with productivity. The confirmatory factor analyses also served the purpose of replicating selected factors found in the Wagner et al. (2011) study, with the exception of the conventions factor.[1]

The current study also examined the development of writing skills over time through the use of dynamic modeling approaches. As Abbott, Berninger, & Fayol (2010) note, to date, few longitudinal studies of writing exist and when longitudinal data has been collected, the results have been analyzed cross-sectionally (e.g. Berninger et al, 2010; Juel, 1988; Juel et al., 1986). Thus, the current study represents an important contribution to the field as it looks at writing longitudinally through the use of modern statistical techniques that capture the dynamic interplay in mean changes among multiple constructs (Grimm, An, McArdle, Zonderman, & Resnick, 2012). More specifically, latent change score (LCS) models were utilized as they allow the modeling of within-person changes (growth) in the individual variables and their interrelationships over time. LCS modeling allowed for the assessment of growth in each of the writing skills, as well as determining the effects of improving one skill on the outcome of the other. LCS models combine features of latent growth curve and cross-lagged models. One advantage that LCS models have is the ability to test hypotheses about lead-lag associations (Ferrer & McArdle, 2010). A leading indicator represents within-person changes in that variable that occur prior to changes in the other (lagging) variable. Based on previous research on written language development, a developmental pattern that may be expected is earlier writing showing shorter texts and less sophisticated vocabulary and organization, with later grades showing extended writing and more complex structure and organization (e.g., Berninger, Abbott, Whitaker, Sylvester, & Nolen, 1995; Nelson & Van Meter, 2007; Wagner et al, 2011). Determining the leading indicators of writing development and including assessments of these skills may be useful in identifying children at risk for writing difficulties, as children whose skills are lower in a leading indicator skill may be at a greater risk of writing problems than children whose skills are lower in the lagging indicators. This information is also useful in assessment development as it will provide information on which skills should be developing earlier.

[1] The conventions factor was not included in the current study due to data scoring complications and an effort to complete the dissertation in a timely manner. It is recommended that future analyses contain this factor and its related indices.

METHODS

Participants

Participants were 158 children (77 male, 81 female), randomly chosen from a larger sample of 316.[2] The children attended six elementary schools in a moderate-sized city in north Florida. On average, the schools served a 40% economically disadvantaged and 1.5% English language learner population. Permission forms were sent home with all students in the first grade classrooms of the participating schools. All children whose parents granted permission were included in the study.

At Time 1, all children were in first grade and had a mean age of 7.03 years (SD = .45). The majority of the children were White (61.4%), with 23.4% Black, 3.8% Hispanic, 4.4% Asian, 5.7% mixed race, and 1.2% not identified. Age and ethnicity information were provided by the local school district.

At Time 2, approximately one year later, 139 children (74 male, 65 female) remained in the study. Five of the children were retained in first grade. The mean age was 8.01 years (SD = .41). White children remained the majority (61.9%), with 21.6% Black, 4.3% Hispanic, 5.0% Asian, 5.8% mixed race, and 1.4% not identified.

At Time 3, approximately two years after the study began, 135 children (72 male, 63 female) remained in the study. Six children were retained in second grade. The mean age was 9.02 years (SD = .46). White children remained the majority (61.5%), with 23.7% Black, 3.7% Hispanic, 5.2% Asian, 5.2% mixed race, and 0.7% not identified.

At Time 4, approximately three years after the study began, 113 children (57 male, 56 female) remained in the study. Seven children were retained in third grade. The mean age was 9.82 years (SD = .41). White children remained the majority (69.9%), with 17.7% Black, 4.4% Hispanic, 5.3% Asian, 1.8% mixed race, and 0.9% not identified.

Measures

Each year of the study, a compositional writing sample was collected as part of a larger study on reading and writing development with a different writing prompt each year.

---

[2] Due to time constraints in scoring the samples, only half of the sample was scored and analyzed for this dissertation. Though it is not expected that the current sample results differ from what would be obtained using the full sample, it is recommended that future analyses utilize the full available sample.

To obtain the writing sample, the task was introduced by saying:

I am going to ask you to do some writing. First I need you to write your name on your paper.

The child was given a sheet of blank, lined paper and a pencil without an eraser. The child was shown the place on the upper, right-hand corner of the page to write their name. After the child wrote their name, the child (at Time 1) was instructed:

You will write about choosing a pet for your classroom. If you could have any pet in the world for a classroom pet, what would choose? When you are writing, I want you to stay focused and keep writing the whole time. Don't stop until I tell you to. If you think of a word that you don't know how to spell, sound it out and do your best. I can't help you with spelling today. If you make a mistake, cross it out and keep writing. Don't erase your mistake because it will take too long. Do you understand?

Any questions were answered and the child was further instructed:

Remember that you will write about choosing a pet for your classroom. Think about why you would like to have a pet in your classroom and write to explain why this animal should be your classroom pet. Ready, begin.

The child was given 10 minutes to write. If the child stopped writing before the time was up he/she was prompted:

What more could you write about choosing a pet?

At Time 2, the prompt asked the child to:

You will write about a favorite subject to learn about. We all have a favorite subject to learn about. Think about one subject that is your favorite to learn about in school. Write to explain why that subject is your favorite.

If the child stopped before 10 minutes, he/she was prompted:

What more could you write about that subject in school?

At Time 3, the prompt asked the child to:

You will write a story about a time you had a day off from school.

If the child stopped before 10 minutes, he/she was prompted:

What more could you write about a time you had a day off from school?

At Time 4, the prompt asked the child to:

You will write about a time you went on a field trip with your class.

If the child stopped before 10 minutes, he/she was prompted:

What more could you write about when you went on a field trip with your class?

Scoring Variables

Macro-organization. Five variables were coded to represent the higher level organization of the writing sample, with scoring being completed using the proper version to avoid bias due to spelling and grammatical errors (e.g., Graham, 1999).

1. Statement of topic. A score of 1 was given when a topic sentence was present and a score of 0 when it was not. A topic sentence was defined as that related to or restated the prompt and did not have to be the first sentence.

2. Logical ordering of ideas. Logical ordering of ideas was rated on a 1- to 4- point rating scale. See Appendix A for rating scale details.

3. Number of key elements. One point was given for the presence of a main idea, body, and conclusion, yielding a maximum possible score of three.

4. Number of supporting arguments. This was a count of the number of details that supported the topic. Repetitive details were only counted one time.

5. Number of digressions. This was a count of the number of details that departed from the topic. Repetitive details were only counted one time.

Complexity. Four variables were included to represent the complexity with which the writing sample conveys information, with scoring being completed on the clean version.

1. Modifiers per noun-phrase. This was the mean number of modifiers per noun phrase and was calculated using the Coh-Metrix text analysis tool.

2. Words before main verb. This was the mean number of words before the main verb of the main clause in sentences and was calculated using the Coh-Metrix text analysis tool.

3. Number of connectives. This was a count of the number of uses of common connective words: *and*, *but*, *so*, *because*, *or*. Connective words were counted when they combined phrases and clauses. Connective words were not counted when they were used as an adverb (e.g., "It was *so* much fun.") or when the connective words were only connecting lists of nouns (e.g., "I want a cat, dog, *and* fish.")

4. Average number of words per sentence. This is ratio of total number of words divided by the total number of sentences. This was calculated using web-based readability software (Scott, 2012).

Productivity. Three variables were considered to represent how much writing was accomplished. Productivity variables were scored using the clean version.

1. Total number of words. Total number of words was the number of words produced in the writing sample by the child. Number of words was calculated using web-based readability software (Scott, 2012).

2. Total number of sentences. The count of the number of sentences included in the writing sample. The sentence structure was used to determine the number of sentences when punctuation and capitalization were not used, which is not uncommon for beginning writers (Kim et al, 2011). This was calculated using web-based readability software (Scott, 2012).

3. Number of unique words. This was a count of the number of unique words used in the writing sample and was calculated using web-based readability software (Scott, 2012).

Vocabulary. Six variables were used to represent the word choice/vocabulary of the writing samples. Vocabulary variables were measured using the clean versions.

1. Rare spelling words. All words in each writing sample were compared to a basic spelling list for elementary students (Graham, Harris, & Loynachan, 1993). Rare spelling words were counted as the number of words not included on the basic spelling list, excluding proper names.

2. Vocabulary diversity. A type-token ratio was calculated using the Coh-Metrix text analysis tool. This was the number of unique words (called types) divided by the number of tokens of these words. Each unique word in a text was considered a word type. Each instance of a particular word was a token.

3. Average characters per word. This is ratio of total number of alphabetic characters in the sample divided by the total number of words. Average characters per word was calculated using web-based readability software (Scott, 2012).

4. Average syllables per word. This is ratio of total number of syllables divided by the total number of words. It was calculated using web-based readability software (Scott, 2012).

5. Percentage of multi-syllabic words. The ratio of multi-syllable words divided by the total number of words. This was calculated using web-based readability software (Scott, 2012).

6. Low frequency words. Word frequency was calculated using the Coh-Metrix text analysis tool. The log frequency of all content words in the text variable was multiplied by negative one, so that higher values would represent lower frequency words.

Transcription, Coding, and Reliability

The samples were transcribed into three formats, as-is, clean, and proper versions by trained research personnel under the direction of the project manager using the following guidelines. For as-is, writing samples were typed exactly as the children wrote them, with no corrections or additions. For the clean version, writing samples were corrected for spelling, capitalization, and end punctuation. For the proper version, writing samples were corrected for spelling, punctuation, grammar, usage, and syntax, while maintaining the fundamental nature of the child's original text (see Appendix B for details). The various transcriptions were used as appropriate for the scoring variables.

The macro-organization variables, number of connectives, rare spelling word counts were hand coded by the author. Prior to coding, directions were reviewed with the project manager and another member of the coding team to ensure high reliability. To calculate reliability, twenty percent of the writing passages were randomly selected for coding by a second coder from the research team. The interrater reliability ranged from .86 to .99 for coded items across transcripts.

The Coh-Metrix variables were calculated using the clean version of the writing passages. Grades 1 through 3 were submitted to Coh-Metrix by a fellow graduate student for a prior research project and Grade 4 was submitted by a member of the research team. The vocabulary and productivity variables that were calculated using web-based readability software (Scott, 2012) were submitted by the author. Because these variables were calculated using automated programs with no human judgment required, reliability is near perfect. Therefore, these variables were only submitted for coding a single time.

Procedure

The writing samples were collected over the course of four school years, with each child being tested on measures once a year, approximately one year apart. Children were tested individually by trained examiners. Testing took place in the children’s elementary schools in a quiet location that was dedicated to testing. Other measures that were included in the original study, but not in the current analysis were DIBELS, Stanford-Binet Vocabulary, WRAT-3 Spelling, Woodcock-Johnson III Test of Achievement Letter Word ID, Word Attack, Passage Comprehension, and Oral Comprehension, Experimental Reading and Listening passages, TOWRE Words and Nonwords, WASI Vocabulary, WIAT-II Written and Spelling, WRMT Passage Comprehension, Memory Span, and TOSRE. Tasks were administered in a fixed order that provided a mixture of short and more cognitively demanding tasks to ease potential fatigue. Examiners were allowed to divide the testing into multiple sessions as needed on a case by case basis.

RESULTS

Data Issues and Attrition

Prior to analysis, data were screened for missing and extreme values. Extreme values were determined using the median +/- two interquartile range criterion. Outliers were brought to the boundary (Tabachnik & Fidell, 2007). Many of the variables had a small number of outliers, mostly extreme values above the boundary. Additionally, the outlier analysis revealed a lack of variability for the number of digressions at all four time points, as the majority of compositions did not include digressions. Therefore, number of digressions was dropped from all further analyses. See Table 1 for specific outlier adjustments. Visual inspection of scatterplots revealed no bivariate outliers. Evaluation of skewness and kurtosis statistics, as well as visual inspection of frequency histograms, showed distributions to be normal.

A common concern in longitudinal studies is the loss of participants over time. As previously noted, 19 children were no longer in the study at Time 2, an additional four were lost at Time 3, and 22 more at Time 4. This yielded 113 participants who participated in all waves of data collection. Children who did and did not complete the study did not differ significantly on demographics or most Time 1 variables. However, there were significant differences in Time 1 measures of average number of characters per word $F [1,156] = 9.206$, $p < .01$, total number of connectives $F[1,156] = 4.784$, $p = .03$, and percentage of multisyllabic words $F[1,156] = 8.142$, $p = .004$, with completers having higher scores. The findings of differential attrition should be considered when interpreting the results and may limit the generalizability of the current findings. As these differences were only a few of the variables, it is believed data is missing at random. Therefore, all participants, including those with missing data, were included in the analyses with missing data handled utilizing maximum likelihood estimation within Mplus (Muthén & Muthén, 1998-2007). This approach was chosen because maximum likelihood estimates of missing data provide the least biased estimates (Little & Rubin, 2002; Peugh & Enders, 2004). Based on previous work (see Kantor & Wagner, 2012), it was not anticipated that this level of missingness would influence the estimations.

Descriptive Statistics and Correlations

Basic statistics such as means, standard deviations, ranges and correlations are reported by Wave in Tables 2 through 6. Means on all measures increased from year to year, with the exception of the vocabulary measures. The means on these measures remained relatively stable over the four waves with slight increases and decreases over time.

Correlations were carefully examined within each proposed dimension. Correlation coefficients were used as guidance for determining the best indicators for latent variables in the confirmatory factor analyses. For example, correlations revealed a weak relationship between number of supporting arguments and the other proposed macro-organization variables. Additionally, rare spelling words was more highly correlated with productivity variables than the other vocabulary variables and type-token ratio was negative related to some of the other vocabulary variables.

Confirmatory Factor Analysis

Confirmatory factor analyses (CFAs) were conducted using Mplus 5.2 (Muthén & Muthén, 1998-2007) to evaluate relations among the writing skills and to explore the addition of vocabulary as a separate indicator of writing.

The first step in conducting the CFA was to evaluate the model fit and loadings of the various indicators of writing onto the proposed latent constructs through the use of a measurement model for each construct of interest. Using the prospective loadings and correlations, each factor was individually determined as follows. Model comparison was conducted by evaluating the Chi-squared (χ^2), Chi-squared to degrees of freedom ratio (χ2/df), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) for continuous and Weighted Root Mean Square Residual (WRMR) for models with categorical and continuous variables. The p-value associated with the χ^2 test gives the probability that the model provides a perfect fit to the population variance/covariance matrix. Thus, a non-significant χ^2 test is desired. Normed chi-square values, which are χ^2 values divided by degrees of freedom (df), indicate a good fitting model if the value is less than three (Kline, 2005) or less than two (Tabachnik & Fidell, 2007). Root Mean Squared Error of Approximation (RMSEA) values equal to or less than .05 indicate a good fit, values between .05 and .08 indicate a reasonable error of approximation, and values great than .10 suggest poor fit (Browne & Cudeck, 1993). However, Steiger (2007) has advocated a more conservative RMSEA of less than .07. Values of .90 or greater on the Tucker-Lewis Index(TLI) and Comparative Fit Index (CFI) indicate acceptable model fit and a value of .95 or greater indicates close fit (Hu & Bentler, 1999). Standardized

Root Mean Square Residual (SRMR) values less than .05 indicate good fit (Byrne, 1998; Diamantopoulos & Siguaw, 2000) and less than .08 indicate adequate fit (Hu & Bentler, 1999). Weighted Root Mean Square Residual (WRMR) values less than .90 indicate good fit for models with categorical and continuous outcomes (Muthén, 2004; Yu, 2002). Multiple fit indexes were evaluated as each has liabilities and assets. Because models were not nested, Akaike information criterion (AIC), and Bayesian information criterion (BIC) fit indexes were used when needed to compare models, with the model with smallest AIC and BIC values having the best fit. Also, because single factor models with only three indices are considered just identified, constraints were imposed to hold the residual variances of two factors equal, which released a degree of freedom so that fit indices could be provided by Mplus. Because a constraint was included in this model, this provided a conservative estimate of model fit. When two indices were considered, the factor was paired with another identified factor and fit index differences were compared. Model fit and comparisons were conducted using data from Time 1.

Given the unexpected negative correlations between number of supporting details with the other macro-organization variables at Times 2 and 3, number of supporting details was excluded from the macro-organization factor. This left the three indicators that were included in the Wagner et al. (2011) study: presence of a topic sentence, order of ideas, and number of key elements. Because presence of topic sentence and number of key elements were derived from binary scores, these variables were treated as categorical and the robust weighted least squares (WLSMV) estimator was utilized in Mplus. The model with these three indicators provided a good fit, χ^2 (df =1) = 1.56, p = .21, CFI = 1.00, TLI = .99, RMSEA = .06, and WRMR = .32.

The productivity single factor model with three indicators (number of words, number of unique words, and number of sentences) provided good fit for the data. Fit indices were χ^2 (df =1) = .140, p = .71, CFI = 1.00, TLI = 1.00, RMSEA = .00, and SRMR = .00. Factor loadings for number of words and number of unique words were nearly double the loading estimate for number of sentences.

The complexity single factor model with four indicators (average number of words per sentence, number of connectives, modifiers per noun phrase, and words before main verb) fit the model well. Fit indices were χ^2 (df =2) = 2.497, p = .29, χ^2/df = 1.25, CFI = 1.00, TLI = .99, RMSEA = .04, and SRMR = .03. While the model met criteria for good fit, the modifiers per noun phrase and words before main verb variables did not have significant factor loadings.

Given the non-significant and unexpected negative correlations between the type-token ratios with the other vocabulary variables, this variable was excluded from the vocabulary factor. The remaining five variables were tested in a series of CFAs, first with all five, then eliminating the smallest or non-significant indices one at a time, until a best fitting model was achieved. Table 7 shows the fit indices for the models tested. The model with three indicators, number of characters per word, average number of syllables, and percent of multisyllabic words, for the vocabulary factor was deemed to have the best fit, based on smaller χ^2 and smallest AIC and BIC.

Model comparison. Using the best fitting measurement model for each factor, the structural model of writing was explored. In order to achieve model convergence at each time point, the number of sentences indicator, which had the lowest factor loading, was dropped from the vocabulary factor, leaving the two indices that were used for this factor in the Wagner et al. (2011) study. Additionally, the modifiers per noun phrase and words before verb indices, which both had non-significant factor loadings, were dropped from the complexity factor. Last, the two indices of productivity were adjusted in order to allow the variance to be on a similar scale to the other factor indices, number of words was divided by 100 and number of unique words was divided by 10.

At each time point, two models were tested. The first model posited four separate but correlated factors of writing: macro-organization, productivity, complexity, and vocabulary. In the second model, the vocabulary indices were included on the productivity factor. Potential changes in the structure of writing were examined by comparing the results of the CFA at each of the four time points. χ^2 difference tests were utilized to determine the best fitting of the proposed models at each time point as the three-factor model was nested within the four-factor model. A significant χ^2 difference test would indicate a difference between the two models and that the less restrictive (four-factor) model is best and most parsimonious. A non-significant χ^2 difference test would indicate that the two models did not significantly differ in fit and that the more restrictive (three-factor) model should be accepted as the best and most parsimonious model.

Fit indexes for the models at each time point and χ^2 difference test results are found in Table 8. As shown in the table, the chi-square difference tests were all significant, indicating that the four factor model that included separate variables for vocabulary and productivity was the best fitting model at each time point. The four factor models at each time point are shown in Figures 1 through 4. At Time 1, there were significant positive correlations for macro-organization with complexity (r = .44) and productivity (r = .47), as well as complexity with productivity (r = .48). At Time 2, there were significant positive for correlations for macro-organization with productivity (r = .32) and complexity with productivity (r = .57). At Time 3, there were significant positive correlations for macro-organization with complexity (r = .23) and productivity (r = .48), complexity with productivity (r = .38), and productivity with vocabulary (r = .22). At Time 4, there were significant positive correlations for macro-organization with complexity (r = .76) and productivity (r = .69), and complexity with productivity (r = .53). At all four time points, productivity was positively related to macro-organization and complexity, revealing that children who write longer compositions were also likely to have more complex writing, as well as greater organization. Additionally, at all four time points, vocabulary was found to be unrelated, and in some cases negatively related, to the other factors of writing. The exception was at Time 3, when vocabulary had a small, but significant, relation (r = .22) to productivity.

Longitudinal Measurement Invariance

In addition to attrition, another concern with longitudinal studies is whether the measures are assessing the same attributes over time. This issue is known as measurement invariance. When using latent variables, measurement invariance is defined as when the factor loadings of indicator variables on their respective factors do not differ significantly across time.

I examined the measurement invariance for the writing model between grades. A multi-group CFA model would not be appropriate here because the same subjects were assessed at differing time points. Thus, I employed a CFA with the Time 1 variables loaded on the latent factors corresponding to Time 1, Time 2 variables loaded on the latent factors corresponding to Time 2, Time 3 variables loaded on the latent factors corresponding to Time 3, and Time 4 variables loaded on the latent factors corresponding to Time 4. Given that the same manifest variables were used for each time, residuals of the corresponding variables were first allowed to be correlated and then excluded from the final model when found insignificant. Time 4 average words per sentence had significant correlated residuals with Time 2 (r = .25, p = .04) and Time 3 (r = .36, p < .01) average words per sentence. Time 4 total words had significant correlated residuals with Time 1 (r = .36, p = .02) and Time 3 (r = .44, p < .01) total words.

The model fit of the restrictive model constraining the factor loadings to be the same for the corresponding variables was compared against the unrestrictive model with no such constraints. The baseline model provided adequate fit χ^2 (df = 57) = 84.93, p <. 01, CFI = .90, TLI =.90, RMSEA=.06, and WRMR=.72. The restrictive model with equal loadings had a decrease in fit χ^2 (df = 51) = 85.92, p <. 01, CFI = .86, TLI =.86, RMSEA=.07, and WRMR=.81. The chi-square difference test between the restrictive model with equal factor loadings and the baseline model without indicated that the model without equal factor loadings fit significantly better, χ^2 diff (df = 9) =28.11, p < .01.[3] I found that all loadings were equal except presence of Topic Sentence at Time 4, as the χ^2 diff (df = 9) =15.62, p = .08[2] was non-significant when this constraint was removed.

Turning to measurement invariance of intercepts, the model with equal intercept loadings did not fit well, χ^2 (df = 59) = 315.08, p <. 01, CFI = .00, TLI =.11, RMSEA=.17, and WRMR= 1.47. The chi-square difference test was significant indicating that the model without equal intercepts fit significantly better, χ^2 diff (df = 11) =858.32, p < .01.[2] Additional equal intercept models were attempted relaxing restrictions variable by variable which provided improvement in model fit, but none of them provided a non-significant difference compared to the model without equal intercepts.

Lack of full measurement invariance may indicate that some items may load on different factors across time or fail to achieve discriminant validity (Steenkamp & Baumgartner, 1998). However, partial measurement variance was satisfied based on the factor structure and the loadings, with the exception of T4 Topic Sentence, being equivalent over time (Thomson & Green, 2006). Statistical inferences are valid as at least one indicator, and in most cases more than one, were metrically invariant (Byrne, Shavelson, & Muthén, 1989).

Latent Change Analyses

The use of latent change score modeling allowed for the determination of causal sequences and determinants of mean changes (Grimm et al., 2012; McArdle, 2009). Four potential hypotheses were compared about the origins of change within pairs of factors: 1) the intraindividual changes in both factors are unrelated to each other; 2) factor X is a leading indicator of changes in factor Y; 3) factor Y is a leading indicator of changes in factor X; and 4)

[3] Chi-square difference tests were conducted using the difftest command procedures outlined by Muthén & Muthén when the WLSMV estimator is utilized. This procedure adjusts the chi-square values and degrees of freedom to obtain a correct p-value.

the relationship is bidirectional such that both X and Y predict each other over time (Jajodia, 2012).

A general principle in building complex models is to begin with simple models, evaluating fit and making modifications as necessary, and sequentially building to the most complex model. The simplest models in the proposed study are the univariate change models. In latent change analyses, the measured value of the continuous variable (X) for each individual (i) at each time (t) of measurement (X[i,t]) is modeled as a continuous latent true score (x[i,t]) plus a residual measurement error term (ex[i,t]). The measurement error term is assumed to have a mean of zero and a normal distribution. The residual errors[4] are assumed to have the same variance across time points to represent the assumption of measurement invariance over time. Change from time (t - 1) to time (t) within an individual is modeled as a latent change variable (Δx[t]) (Grimm et al, 2012; Jajodia, 2012). The latent intercept term (x0i) provides the initial starting values for the latent score and has effects along the single headed arrows from (x[t] to x[t - 1])) such that the intercept mean (μx0) and variance (σ^2x0) are part of the expected value of every time point. The first source of individual variation is the latent slope score (xsi) which represents change over time. The latent changes are a function of an additive component (α[t]), which represents a constant influence (which is usually fixed at 1), though it may be set at a different number when intervals are different from the change represented by the slope. α[t] can also be used as a shaping parameter and estimated by the model to indicate changes in slope over time. Additionally, a dynamic proportional change path (β[t]) can be estimated that represents the effect of change at one point based on the previous point. Furthermore, any change that occurs earlier accumulates and is expressed in the later change variables. When this process is repeated for each time point, a layer of (t -1) latent change variables are added to the model (McArdle, 2009). As there are four waves of data in the proposed study, there were 3 latent change variables (Ferrer & McArdle, 2010; Grimm et al, 2012; McArdle, 2009). A basic univariate latent change model is shown in Figure 5.

For each factor, three basic univariate models were fit, using a single indicator for each time point.[5] The indicator was selected based on having the least variance over time based on the

[4] Significant correlated residuals that were found in the measurement invariance models were allowed to be correlated in the latent change models as well.

[5] The models were originally fit using multiple factor indicators but model fit was decreased (e.g., fit indices for the multiple indicator dual change model of complexity were χ^2 (35) =183. 74, χ^2/df = 5.25, CFI = .24, TLI = .39, and

measurement invariance evaluation, as well as being the indicator with highest face validity and use in previous research.[6] Thus, order of ideas was used for macro-organization, number of words for productivity, average words per sentence was used for complexity, and average number of syllables per word was used for vocabulary. The first model posits no change, the second constant change, and the third allows dual change. The no change model posits that the factor does not change for individual children over time. This model is fit by constraining the mean and variance of the slope, as well as the correlation between slope and intercept to zero. The constant change model assumes a linear slope over time and is achieved by allowing the mean and variance of the slope, as well as the correlation between intercept and slope to be estimated. The dual change model allows for nonlinearity by including the estimation of a proportional change term. Additional models that allow multiple slopes, suggesting variation in the slope shape, are creating by placing a spline at certain waves that allows the estimation of multiple change parameters (α).

The constant change model found for macro-organization was found to be the best fitting, as it showed significant improvement over the no change model, but was not improved by the addition of proportional change paths (see Table 9). Fit indices for the constant change model met criteria for good fit, χ^2 (df = 8) = 9.37, p =. 31, χ^2/df = 1.17, CFI = .95, TLI =.96, RMSEA=.03, and SRMR= .07. The significant positive value for the slope mean (μs = .22) indicated that, on average, each year children improved .22 in order of ideas. The mean intercept ($\mu 0$ = 2.07) indicated that average starting level of order of ideas was 2 which represents an emerging structure in the written composition. The significant variance of the intercept indicated significant variability in children's order of ideas score at Time 1.

Turning to the univariate change model of productivity, the three main models, no change, constant change, and dual change did not provide good fit for the data (see Table 10). Upon examination of a plot of the means for the factor indicators, number of words, at each time point (see Figure 6), it appeared that multiple slopes may have been present, with a steeper slope after Time 2. Thus, a spline model was fit that allowed for the estimation of (α) from Times 2

RMSEA = .16 compared to the single indicator dual change model which were χ^2 (7) =10.75, χ^2/df = 1.54, CFI = .73, TLI = .77, and RMSEA = .06). Furthermore, most studies utilizing latent change analysis have relied on single indicator factors (e.g., Gerstorf, Hoppman, Kadlec, & McArdle, 2009; Kuoros & Cummings, 2010; McArdle & Prindle, 2008)

[6] Mean patterns were similar for all indicators per factor, so it is expected that results would be similar if alternate indicators had been utilized.

through 4. The constant change spline model yielded significant model fit improvement over the models without, though the spline with dual change was no better. The fit indices for the constant change with spline at Time 2 model were $\chi2$ (df = 7) = 24.57, p =<.01, χ^2/df = 3.51, CFI = .85, TLI =.87, RMSEA=.13, and SRMR= .15. While the model fit was not ideal, additional models that included multiple splines did not result in improved fit, thus the constant change spline at Time 2 model was accepted as the best description of the data. The spline at Time 2 model indicated a constant slope until Time 2 (α = 1.00), and then a different constant slope that was three times greater for Times 2 through 4 (α = 3.33). The significant positive value for the slope mean (μs = .11) indicated that children wrote 11 more words each year for the first slope, Times 1 to 2, and 33 words per year from Times 2 through 4.[7] The significant variance in the slope term indicated there was variability in how much the number of words changed over time. The mean intercept ($\mu 0$ = .44), indicated an average of 44 words written at Time 1.[7] The significant variance of the intercept indicated significant variability in the number of words written at Time 1.

For complexity, the three main models were fit, as well as a spline at Time 2 model, as it appeared that there may have been a decline in speed of growth after that point when the plot of means was examined (see Figure 7). However, while the models with spline do show a second slower slope (α = 1 for Times 1-2, α = .32 for Times 2-4 in the constant change model), this model did not provide an improvement in fit over the dual change model when fit indices were compared, larger AIC and χ^2/df and smaller CFI and TLI (see Table 11). Thus the dual change model without spline was accepted as the best fitting model. The fit indices for this model were χ^2 (df = 7) = 10.75, p = .15, χ^2/df = 1.54, CFI = .73, TLI =.77, RMSEA=.06, and SRMR= .12. While the CFI, TLI, and SRMR do not meet criteria of good fit, the non-significant χ^2, χ^2/df less than 2, and RMSEA of .06 do meet criteria for good fitting models and so this model was accepted to represent change in complexity. The dual change model included the proportional change or self-feedback component (β = -.55), which indicated a 55% decrease in the slope of average words over time due to the proportional change component. The mean slope was positive, indicating an increase of 5.85 words per sentence each year. The slope and proportional feedback were coupled together to characterize the change in complexity over time. The mean

[7] Number of words was divided by 100 to handle scale variance for analyses. To return to the original scale, estimates are multiplied by 100.

intercept ($\mu 0 = 8.09$), indicated an average of eight words per sentence at Time 1. The significant variance of the intercept indicated significant variance in the number of words written by children at Time 1.

As can be seen in Figure 8, the means of the vocabulary indicator grow and decline, then growth again. In order to model this pattern, in addition to the three main change models, a latent change model was fit with splines at Time 2 and Time 3.[8] None of the models provided good fit (see Table 12). The negative TLI indicated that the base model, which specifies no relation among the variables, provided a better fit than the specified models. Upon further investigation, two findings were considered as explanations for the lack of fit. First, the magnitude of change was very small between the time points (mean differences ranging from .04 to .13). Second, it was found that there was no correlation for any of the vocabulary indicators across time (see Table 13).[9] Given this finding of no relationship among the vocabulary indicators over time, no further change models were fit for vocabulary.[10]

Next, the best fitting change models for the individual factors of macro-organization, complexity, and productivity[11] (see Table 14) were then coupled to study multivariate change processes and identify the lead-lag relationships. For each model, coupling parameters (γ_{xy} and γ_{yx}) can be fixed or estimated to test the various lead-lag hypotheses. Coupling parameter (γ_{xy}) represents the effect of latent x[*t*] on Δy[*t*]. Coupling parameter (γ_{yx}) represents the effect of latent y[*t*] on Δx[*t*]. For the no-coupling model, changes in variables are unrelated; the coupling parameters were fixed to zero. For testing unidirectional influences, only the coupling parameter that was expected to be the leading indicator was estimated. Lastly, in the bidirectional coupling model, both X and Y affect each other and both coupling parameters were estimated (Ferrer & McArdle, 2010; Grimm et al, 2012; Jajodia, 2012; McArdle, 2009). An example of a bivariate model is found in Figure 9.

[8] The vocabulary latent change analyses were conducted using each of the indicators, which had the same mean pattern. Model fit for average syllables per word provided the best fit and had the least amount of variance over time in the measurement invariance models, so it was deemed best for the latent change analyses.
[9] The pattern of correlation among the three indicators with each other was similar at each separate time point, so this does not change the results of CFA analyses.
[10] The lack of significant relationships over time was also found for the Coh-Metrix low word frequency variable. T1-T2 r = -.10, T1-T3 r = .01, T1-T4 r = -.13, T2-T3 r = .00, T2-T4 r = -.04, T3-T4 r = .17). Additionally, average syllables per word was available for the larger study sample (n = 318) and the lack of correlation over time was also found indicating that this was not specific to the subsample (T1-T2 r = .10, T1-T3 r = .06, T1-T4 r = .06, T2-T3 r = .15*, T2-T4 r = .12, T3-T4 r = .27*). *p<.05.
[11] Bivariate relations with vocabulary were not included as the univariate models did not fit the data.

The results of bivariate models are shown in Tables 15 to 17. For the bivariate models of macro-organization with complexity and macro-organization with productivity, the no-coupling model provided the best fit, indicating that changes in the factors were unrelated. For each of these models there was a significant intercept-intercept relationship r = .89 for macro-organization with complexity and r = .94 for macro-organization with productivity, but no significant slope relations. The significant intercept-intercept relationships reveal that children whose initial scores were high on one variable also had high initial scores on the other variables, and children with lower scores on one had lower scores on the other. Thus children with higher initial levels of macro-organization skills also had higher initial levels of complexity and productivity. Likewise, children with lower macro-organization skills at Time 1 also had lower complexity and productivity scores at Time 1.

For the bivariate model of productivity-complexity, the model with productivity as a leader for complexity provided a significantly better fit χ^2 (df =1) = 5.21, p = .02, indicating that productivity, as measured by number of words, was predicting the changes in complexity, as measured by average number of words per sentence. Additionally, the bidirectional model showed near significant improvement (p = .07) over the no coupling model. The value of the coupling parameter was positive (γ_{xy} = 2.00) suggesting that children who write longer compositions also increase the complexity of their writing over time. The value of β_y, which represents proportional change in complexity, was negative and suggests that complexity levels were not increasing as fast for these children. Thus, productivity affects change in complexity as well as change in itself (self-feedback). The mean slopes of both were positive, indicating that most children increase in productivity (μs = .11) and complexity skills (μs = 7.32) over time. Additionally, there was a significant positive correlation (r = .42) between the slope of productivity and the intercept of complexity, indicating that children who grow faster in productivity also tend to have higher initial levels of complexity and children with lower initial levels of complexity, grow slower in productivity.

DISCUSSION

The current study of writing development in children from first through fourth grade found that writing was best conceptualized by four separate factors: macro-organization, productivity, complexity, and vocabulary. Additionally, the developmental trajectories of the writing factors were explored using latent change analyses.

As expected, the earlier writing of children included shorter texts and less sophisticated organization and complexity. Later grades showed extended writing and more complex structure and organization. However, measures of vocabulary did not show consistent growth and indeed were found to be unrelated to one another.

Structure of Writing

Turning to the confirmatory factor analyses, there was a consistent finding at each time point that the separate abilities, four-factor model of macro-organization, productivity, vocabulary, and complexity fit best, as evidenced by the significant chi-square difference tests at each time. This is consistent with the findings of Wagner et al. (2011) and Guan et al. (2012) that also found writing to best be modeled as separate factors.

Consistent positive correlations were found between productivity with macro-organization ($r = .32$ to $.69$), and productivity with complexity ($r =. 38$ to $.57$). Thus, children with greater levels of productivity also had greater macro-organization and complexity in their compositions. This finding is consistent with expectations that children who write more would have greater opportunity for organization. Furthermore, the relationship of the current indices of complexity, average words per sentence and a count of the number of connectives is expected to be highly related due to an emphasis on word count. The factor correlation patterns somewhat differ from those found by Wagner et al. (2011). Wagner et al. found similarly moderate correlations of productivity with macro-organization ($r = .34$) and productivity with complexity ($r = .31$) in first grade, but did not find the strong relationship that the current study did in fourth grade($r = .18$ and $-.14$, respectively). Guan et al. (2012) did find large correlations between productivity with macro-organization ($r = .56$ and $.71$) in fourth grade Chinese writers, but found mixed results for productivity with complexity ($r = .06$ and $.46$). The lack of consistent findings in the relations among the factors is likely due to differences in the study designs. Wagner et al. and Guan et al. both used cross-sectional samples and both used expository prompts, whereas the current study included a longitudinal sample and used a narrative prompt in Year 4. Additionally, there were differences in indicators for the complexity factor. The current study utilized average number of words and number of connectives as indicators, while Wagner et al. (2011) and Guan et al. (2012) used mean length of T-units and clause density as factor indicators.

The current study found macro-organization and complexity to be significantly related at Times 1 ($r = .47$), 3 ($r = .23$), and 4 ($r = .76$). Because one of the measures of macro-organization, order of ideas, included an emphasis on transition words, this relationship may have been influenced by the number of connectives indicator of complexity. Indeed, the correlations between order of ideas and number of connectives was significant at Time 1 ($r =.29$) and Time 3 ($r = .30$) at Time 3. Similarly to the relations with productivity, Wagner et al. found a similar correlation between macro-organization with complexity in first grade ($r = .38$), but there was almost no relationship in fourth grade writing ($r = .07$). Guan et al. (2012) also found a small relationship between macro-organization with complexity ($r = .13$ and .16) in fourth grade. Again, the differences in factor correlations between the studies are likely explained by study design differences.

The vocabulary factor was not significantly correlated with the other factors of writing. The exception was Time 3, when vocabulary had a small significant correlation ($r = .22$) with productivity. The indicators that were utilized to represent the vocabulary factor all represented the complexity of the word, number of characters, average number of syllables, and percentage of multisyllabic words. Thus, the lack of correlation may be based on the vocabulary indicators focus on word-level features, while the other factors represented sentence- and discourse-level features (Nelson & Van Meter, 2007). Word choice variables were also found to form a distinct factor in a study of over 12,000 students in grades four through twelve (Attali & Powers, 2008). Small correlations revealed the word choice factor (vocabulary and word length) to be distinct from the other factors, $r = .12$ with conventions (grammar, usage, and mechanics) and $r = .07$ with fluency (essay length and style).

Development of Writing

The univariate models of latent change provided information about the initial levels and growth patterns of the children's writing skills in each factor. For macro-organization, a constant change model was best fitting, indicating linear growth in children's order of ideas. The intercept mean (2.07) indicated that children's initial level of order of ideas was a score of 2. Looking at the rating scale guidelines, this indicates that, for most children, composition structure was just beginning to emerge. For most of the writing samples, the compositions were a string of sentences connected loosely by the prompt or the sentences could be moved around at will due to lack of sequencing and transition words. The slope of .22 indicated that, on average, children's order of idea scores increased by .22 each year. This means that on average, by Time 4, children would have only increased their scores by .66. Thus, it appears that over the course of our study, while improvements in ordering of ideas were made by the children, it was not large in terms of mean gains. This may indicate a lack of focus on organization skills in instruction, difficulty in obtaining these skills, or that our measurement scale is not nuanced enough to capture specific developmental changes in this skill during the elementary years.

The productivity univariate latent change model that provided the best fit was one with a constant slope until Time 2 and a second slope for Times 2 through 4. The second slope was three times greater, indicating that children's growth in productivity sped up between Times 2 through 4, as measured by the number of words written. Thus from Time 1 to Time 2, on average, children added 11 words, and from Time 2 through Time 4, children added 33 words per year.

The dual change model provided the best model of complexity. The dual change model is a nonlinear model of growth that included a proportional change component which allowed the slope to change over time. For complexity, the average initial value was eight words per sentence. The slope was 5.85 words per minute. This slope was combined with the proportional change value of $\beta = -.55$, which indicated that the slope slowed by 55 percent over time. Thus, growth in complexity, as measured by average words per sentence, slowed over the course of the study. One possible explanation for this finding is that children were more likely to use longer run-on sentences at Time 1 and, as their writing skills increased, they learned to no longer include these long run-on sentences. Additionally, it may be that elementary children were not yet taught the complex structures that allow for sentences to include more than 10-11words without becoming run-ons.

No latent change model provided good fit for vocabulary. Two findings were considered as explanations for the lack of fit. First, the mean changes were very small between the time points. The small mean differences may indicate that children's vocabulary choices in writing, as captured by our variables, do not develop significantly during the early elementary period. An explanation for the overall small mean changes is that children's vocabulary choices may be limited by their spelling abilities. Wagner et al. (2011) did not find a significant decrease in spelling errors between first and fourth grade and attributed this to children's choice of words to write that they had a certain level of confidence in spelling correctly. It may be that children's vocabulary choices were similarly constrained. Future research should examine the relations between the vocabulary variables and spelling errors in the written compositions. Second, it was found that there was no correlation for any of the vocabulary indicators across time, which was indicated by the baseline model providing superior fit to the tested models. This indicated that the current measures of vocabulary were not related from one written composition to another. A possible explanation for this finding is that children's word choices were based on the prompt, and given four varying writing prompts in the current study; there was diversity in the vocabulary chosen. For example, the Time 1 prompt asked children about a pet for their classroom, many of the children chose a "cat" or "dog" and said they were "cute" and "good" which are relatively short words. At Time 2, when the indices had the highest mean, the prompt asked about the child's favorite school subject, and words used were more complex, such as "reading", "science", "favorite" and "recess".

The correlations between the macro-organization, productivity, and complexity factors were also modeled in the bivariate latent change analyses. The macro-organization with productivity and macro-organization with complexity bivariate models were best modeled by the no change model, signifying that changes in these variables were not related to each other. However, significant intercept-intercept correlations were present; suggesting that children with higher initial levels of one skill had higher initial levels of the other and children with lower initial values of one had lower initial values of the other. The best fitting model of productivity with complexity was the latent change model with productivity as the leading indicator. This suggests that children's productivity is a strong predictor of changes in writing complexity. This finding is consistent with studies that have found number of words to be the best predictor of writing quality (Korbin, Deng, & Shaw, 2011; Tindal & Parker, 1991). Productivity is also the most prominent feature of Curriculum Based Measurement of Writing (CBM-W, Coker & Richey, 2010; McMaster & Campbell, 2008; Powell-Smith & Shinn, 2004).

Implications for Education, Limitations, and Future Directions

As mentioned in the introduction, current standards in writing have placed an increased emphasis on children's achievement in all aspects of writing. The Common Core State

Standards Initiative states that each year student's will increase their syntax, vocabulary, organization of ideas, planning, revising, editing, and use of content and source skills (National Governors Association for Best Practices, Council for of Chief State School Officers, 2010). Additionally, a recent practice guide developed for the What Works Clearinghouse entitled Teaching Elementary Students to Be Effective Writers (Graham et al. 2012) states there is strong evidence to support teaching students the writing process, which includes using transition words, writing good topic sentences, evaluating their organization for clear beginning, middle, and end. Additionally, the report states there is a moderate amount of evidence to support the teaching of basic writing skills such as handwriting, spelling, and sentence construction (complexity). The current study revealed that, for students in the early stages of writing development, the writing skills of macro-organization, complexity, productivity, and vocabulary are best conceptualized as separate, but related skills. Additionally, productivity (as measured by the number of words written) appears to be a key indicator. Future research should explore interventions based on the developmental scheme found in the latent change analyses, such as the impact of productivity training on complexity. Additionally, future work should look at the predictive ability of each of the skills to determine whether leading indicators, such as productivity, are more predictive of later writing ability. As teacher's aim to improve children's writing abilities, it is clear that there needs to be an emphasis on development of these basic writing skills.

One limitation of the current study was the lack of developmental relations among the vocabulary variables over time, as evidenced by the poor model fit for the latent change models. Future work should include vocabulary variables beyond word complexity, such as counts of specific types of words (e.g. content words). Additionally, not all of the fit indices for the complexity latent change model reached accepted levels. Furthermore, determination of the factor indicators was somewhat exploratory, as potential indicators were eliminated based on low correlations or non-significant factor loadings. This limits the generalizability of the results. Generalizability is also limited by the adjustments made to the data to bring outliers to the boundary. Another limitation of the current study is attrition. Differences were found in Time 1 scores of average number of characters per word, percentage of multisyllabic words, and total number of connectives between study completers and study non-completers. While this only affected three of the variables and none of those were included in the latent change analyses, differential attrition should still be considered when interpreting the results and may limit the generalizability of the current findings. Additionally, the current study began with the premise that macro-organization, complexity, and productivity had been established as separate factors (see Wagner et al, 2011 and Guan et al., 2012). Given the strong correlations between the macro-organization, complexity, and productivity factors at Time 4, as well as the correlational differences between the current study and the previous two, it is recommended that future studies explore the factor structure further. An additional limitation of the study is the inability to separate differences due to prompt effects from the current findings. Future studies should investigate the equivalence of the prompts and the impact of different prompt types (e.g. narrative, expository, persuasive) on the various writing factors.

In conclusion, the current study found vocabulary as a separate and unrelated factor of writing and also confirmed that the writing factors were best modeled as separate factors in a longitudinal sample of children's writing. This supports the cross-sectional results that were found in Wagner et al. (2011) and Guan et al. (2012). Additionally, the developmental interrelations were explored using latent change analysis, which found productivity to be a leading indicator of complexity, as well as strong relationships between the initial levels of macro-organization, productivity, and complexity.

Table 1.
Outlier Analysis

Variables	Time 1			Time 2			Time 3			Time 4		
	Below	Above	Total	Below	Above	Total	Below	Above	Total	Below	Above	Total
Number of unique words		3	3		3	3		4	4	1		1
Rare spelling words		9	9					1	1			
Type token ratio	1		1				2	1	3		1	1
Characters per word		5	5					2	2	1	2	3
Average syllables per word		5	5	1	1	2					1	1
Percentage of multisyllabic words		6	6		3	3		2	2			
Supporting arguments		3	3		2	2		3	3		1	1
Number of digressions		31	31		26	26		15	15		11	11
Mean number of modifiers per noun-phrase		3	3		4	4	5	1	6		2	2
Mean number of words before the main verb of main clause in sentences		9	9		6	6		11	11		4	4
Number of connectives		5	5					6	6		2	2
Number of words		5	5		7	7		1	1		1	1
Number of sentences		7	7		8	8		2	2		1	1
Avgerage words per sentence		10	10		5	5		6	6		5	5
Low frequency words		2	2	1	4	5						

Table 2.
Means and standard deviations by time.

	N	Mean	Std. Deviation	Minimum	Maximum
Time 1					
T1Topic	158	.66	.48	.00	1.00
T1OrderIdeas	158	2.06	.62	1.00	3.00
T1KeyElements	158	2.08	.60	.00	3.00
T1Supporting	158	3.82	2.22	.00	10.00
T1Modifiers per noun-phrase	158	.49	.23	.00	1.11
T1Mean words before verb	158	1.61	.95	.00	3.77
T1Connectives	158	2.40	2.30	.00	9.00
T1AvgWordsSentence	158	8.06	3.68	1.10	16.90
T1Words	158	43.94	20.81	9.00	98.00
T1Sentences	158	6.04	3.05	1.00	13.00
T1UniqueWords	158	28.14	11.82	6.00	57.00
T1Rare spelling words	158	3.79	2.44	.00	9.00
T1Type-token ratio	158	.79	.12	.47	1.00
T1CharactersperWord	158	3.43	.34	2.65	4.25
T1Avgsyllables	158	1.15	.10	.96	1.40
T1PercMultisyllabic	158	.12	.07	.00	.31
T1Low frequency words	158	-2.39	.29	-3.19	-1.60
Time 2					
T2Topic	138	.93	.26	.00	1.00
T2OrderIdeas	138	2.34	.60	1.00	3.00
T2KeyElements	138	2.33	.51	1.00	3.00
T2Supporting	138	4.35	2.27	.00	10.00
T2Modifiers per noun-phrase	138	.49	.24	.05	1.07
T2Mean words before verb	138	1.89	.93	.00	4.19
T2Connectives	138	3.14	2.22	.00	11.00
T2AvgWordsSentence	138	9.34	3.14	3.93	17.56
T2Words	138	55.15	24.51	16.00	115.00
T2Sentences	138	6.25	2.96	1.00	13.00
T2UniqueWords	138	35.04	13.31	12.00	69.00
T2Rare spelling words	138	6.80	3.17	1.00	15.00
T2Type-token ratio	138	.75	.12	.44	1.00
T2CharactersperWord	138	3.79	.35	2.96	4.65
T2Avgsyllables	138	1.28	.11	.97	1.57
T2PercMultisyllabic	138	.19	.06	.06	.33
T2Low frequency words	138	-2.59	.20	-3.11	-2.08

Table 2 (continued)

	N	Mean	Std. Deviation	Minimum	Maximum
Time 3					
T3Topic	135	.91	.29	.00	1.00
T3OrderIdeas	135	2.42	.57	1.00	4.00
T3KeyElements	135	2.44	.53	1.00	3.00
T3Supporting	135	5.97	3.01	.00	14.00
T3Modifiers per noun-phrase	135	.56	.17	.20	1.07
T3Mean words before verb	135	2.35	1.03	.50	4.62
T3Connectives	135	3.75	2.64	.00	10.00
T3AvgWordsSentence	135	9.90	2.93	4.67	16.84
T3Words	135	90.84	35.64	17.00	181.00
T3Sentences	135	9.58	4.17	1.00	19.00
T3UniqueWords	135	56.17	18.96	14.00	101.00
T3Rare spelling words	135	8.84	4.93	1.00	22.00
T3Type-token ratio	135	.77	.08	.57	.97
T3CharactersperWord	135	3.52	.22	2.99	4.12
T3Avgsyllables	135	1.18	.09	.95	1.41
T3PercMultisyllabic	135	.14	.05	.02	.26
T3Low frequency words	135	-2.60	.21	-3.15	-2.02
Time 4					
T4Topic	112	.97	.16	.00	1.00
T4OrderIdeas	112	2.76	.60	1.00	4.00
T4KeyElements	112	2.46	.54	1.00	3.00
T4Supporting	112	12.71	5.29	.00	27.00
T4Modifiers per noun-phrase	113	.64	.15	.35	1.05
T4Mean words before verb	113	2.46	.93	.88	4.75
T4Connectives	113	4.43	3.18	.00	12.00
T4AvgWordsSentence	113	10.56	2.82	5.73	17.19
T4Words	113	125.87	42.09	12.00	239.00
T4Sentences	113	12.37	4.52	2.00	24.00
T4UniqueWords	113	76.05	21.72	13.00	128.00
T4Rare spelling words	113	14.21	6.64	1.00	31.00
T4Type-token ratio	113	.78	.08	.59	.99
T4CharactersperWord	113	3.63	.20	3.10	4.09
T4Avgsyllables	113	1.22	.07	1.05	1.41
T4PercMultisyllabic	113	.15	.04	.07	.25
T4Low frequency words	113	-2.51	.19	-2.93	-1.98

Note. T1 = Time 1, T2 = Time 2, T3 = Time 3, T4 = Time 4.

Table 3.
Time 1 Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. T1Words	1																
2. T1UniqueWords	.938**	1															
3. T1Sentences	.587**	.559**	1														
4. T1Topic	.287**	.354**	.031	1													
5. T1OrderIdeas	.366**	.417**	-.018	.325**	1												
6. T1KeyElements	.271**	.303**	-.053	.246**	.686**	1											
7. T1Supporting	.471**	.505**	.338**	.244**	.391**	.310**	1										
8. T1AvgWordsSentence	.292**	.246**	-.523**	.211**	.307**	.290**	.019	1									
9. T1Connectives	.531**	.437**	-.079	.143	.292**	.226**	.315**	.623**	1								
10. T1Modifiers per noun-phrase	.007	.016	-.125	.105	-.008	-.133	-.236**	.143	-.014	1							
11. T1Means words before verb	.153	.165*	-.191*	.096	.182*	.262**	-.068	.369**	.082	.131	1						
12. T1CharactersperWord	-.038	.061	-.019	.032	.079	.038	.101	.004	.049	.014	-.077	1					
13. T1Avgsyllables	.081	.144	.069	-.009	-.042	-.027	.008	.012	.058	.207**	-.021	.719**	1				
14. T1Rare spelling words	.586**	.648**	.418**	.070	.269**	.080	.424**	.057	.279**	.114	-.037	.290**	.378**	1			
15. T1Percent multisyllabic	-.072	-.012	-.026	-.029	-.059	-.035	-.046	-.026	.015	.197*	-.022	.696**	.868**	.279**	1		
16. T1Type-token ration	-.372**	-.161*	-.301**	.032	.070	-.074	-.040	-.095	-.181*	.016	-.100	.088	-.036	.032	.083	1	
17. T1LowFRQCLacw	-.167*	-.185*	-.095	-.178*	-.232**	-.279**	-.187*	-.061	-.049	.211**	-.181*	.307**	.265**	.200*	.339**	.115	1

Note. T1 = Time 1.
n = 158
* $p < .05$
** $p < .01$

Table 4.
Time 2 Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. T2Words	1																
2. T2UniqueWords	.930**	1															
3. T2Sentences	.722**	.676**	1														
4. T2Topic	.048	.087	.005	1													
5. T2OrderIdeas	.210*	.325**	.120	.254**	1												
6. T2KeyElements	.170*	.179*	.132	.123	.514**	1											
7. T2Supporting	.481**	.419**	.397**	-.019	.234**	-.017	1										
8. T2AvgWordsSentence	.256**	.243**	-.427**	.072	.067	.002	.083	1									
9. T2Connectives	.634**	.540**	.225**	.094	.061	.022	.345**	.510**	1								
10. T2Modifiers per noun-phrase	-.003	.052	-.109	.090	.031	-.017	-.071	.165	-.104	1							
11. T2Means words before verb	.140	.176*	-.138	.029	.105	.017	-.072	.384**	.159	.276**	1						
12. T2CharactersperWord	-.084	-.007	-.216*	.067	.046	-.014	-.034	.172*	-.027	.401**	.195*	1					
13. T2Avgsyllables	.149	.200*	-.016	.137	.176*	.113	.144	.175*	.073	.263**	.228**	.763**	1				
14. T2Rare spelling words	.639**	.679**	.387**	.169*	.232**	.066	.416**	.278**	.473**	.138	.126	.293**	.416**	1			
15. T2Percent multisyllabic	-.112	-.067	-.180*	.161	.127	.048	-.011	.092	-.044	.359**	.088	.771**	.667**	.265**	1		
16. T2Type-token ration	-.445**	-.227**	-.390**	.033	.068	-.150	-.135	.030	-.271**	.019	-.053	.070	-.119	-.113	.072	1	
17. T2LowFRQCLacw	-.122	-.101	-.135	-.062	-.083	-.203*	.167	.091	-.052	.153	-.015	.341**	.166	.224**	.338**	.345**	1

Note. T2 = Time 2.
n = 138
* p < .05
** p < .01

Table 5.
Time 3 Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. T3Words	1																
2. T3UniqueWords	.958**	1															
3. T3Sentences	.791**	.785**	1														
4. T3Topic	.167	.159	.062	1													
5. T3OrderIdeas	.389**	.423**	.329**	.234**	1												
6. T3KeyElements	.152	.128	.074	.161	.327**	1											
7. T3Supporting	.500**	.477**	.538**	-.073	.388**	.163	1										
8. T3AvgWordsSentence	.123	.087	-.449**	.220*	.019	.049	-.144	1									
9. T3Connectives	.574**	.514**	.206*	.099	.302**	.074	.245**	.425**	1								
10. T3Modifiers per noun-phrase	-.100	-.059	-.159	.063	-.078	-.154	-.047	.166	-.304**	1							
11. T3Means words before verb	.087	.086	-.169*	.176*	.040	.028	.038	.432**	.097	.146	1						
12. T3CharactersperWord	.091	.190*	.130	.019	.015	-.156	.079	-.055	-.189*	.240**	.121	1					
13. T3Avgsyllables	.124	.205*	.169*	-.078	.040	-.145	.091	-.112	-.148	.168	.100	.845**	1				
14. T3Rare spelling words	.690**	.737**	.616**	-.005	.262**	.061	.522**	.005	.280**	.042	.056	.374**	.332**	1			
15. T3Percent multisyllabic	.227**	.289**	.161	.022	.041	-.048	.099	.085	-.055	.206*	.185*	.751**	.747**	.372**	1		
16. T3Type-token ration	-.452**	-.294**	-.387**	-.264**	-.019	-.063	-.128	-.051	-.212*	.112	.033	.015	-.011	-.087	-.076	1	
17. T3LowFRQCLacw	.013	.058	.044	-.228**	.050	-.009	.192*	-.051	-.162	.346**	.083	.306**	.255**	.431**	.234**	.275**	1

Note. T3 = Time 3.
n = 135
* p < .05
** p < .01

Table 6.
Time 4 Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. T4Words	1																
2. T4UniqueWords	.951**	1															
3. T4Sentences	.738**	.723**	1														
4. T4Topic	.254**	.260**	.135	1													
5. T4OrderIdeas	.360**	.459**	.244**	.117	1												
6. T4KeyElements	.084	.061	.022	.248**	.293**	1											
7. T4Supporting	.727**	.692**	.679**	.138	.282**	.046	1										
8. T4AvgWordsSentence	.249**	.213*	-.431**	.229*	.133	.124	-.009	1									
9. T4Connectives	.569**	.493**	.176	.232*	.066	.131	.435**	.524**	1								
10. T4Modifiers per noun-phrase	-.015	.039	-.145	-.060	.023	-.070	-.100	.151	-.171	1							
11. T4Means words before verb	.165	.117	-.194*	.071	.212*	.101	-.005	.524**	.097	.117	1						
12. T4CharactersperWord	.041	.164	.023	-.249**	.287**	-.116	.087	-.015	-.177	.159	-.017	1					
13. T4Avgsyllables	.021	.136	.050	-.160	.187*	-.103	.049	-.067	-.200*	.067	-.021	.861**	1				
14. T4Rare spelling words	.692**	.778**	.520**	.131	.467**	-.004	.600**	.144	.220*	.148	.109	.395**	.366**	1			
15. T4Percent multisyllabic	.040	.154	.022	-.153	.140	-.150	-.002	-.010	-.195*	.151	.010	.708**	.723**	.368**	1		
16. T4Type-token ration	-.363**	-.167	-.303**	-.166	.134	-.149	-.164	-.107	-.230*	.117	-.180	.272**	.257**	.089	.223*	1	
17. T4LowFRQCLacw	-.057	-.006	-.037	-.039	.042	.035	.178	-.060	-.039	.270**	-.085	.338**	.312**	.284**	.257**	.214*	1

Note. T4 = Time 4.
n = 113
* p < .05
** p < .01

Table 7.
Determining indices of vocabulary factor at Time 1

	χ^2	df	*p-value*	χ^2/df	CFI	TLI	RMSEA	SRMR	AIC	BIC	AdjBIC
5 indicators	13.22	5	0.02	2.64	0.98	0.96	0.10	0.03	-105.47	-59.54	-107.02
Char/Word, AvgSyll, RareSpell, PercMultisyll, LowFreq											
4 indicators	5.82	2	0.05	2.91	0.99	0.97	0.11	0.02	-800.32	-765.19	-803.16
Char/Word, AvgSyll, PercMultisyll, LowFreq											
4 indicators	4.20	2	0.12	2.10	0.99	0.98	0.08	0.02	-156.14	-119.39	-157.38
Char/Word, AvgSyll, RareSpell, PercMultisyll											
3 indicators	0.17	1	0.68	0.17	1.00	1.01	0.00	0.00	-871.42	-846.92	-872.24
Char/Word, AvgSyll, PercMultisyll											

Note. Char/Word = number of characters per word, AvgSyll = average number of syllables per word, RareSpell = number of rare spelling words, PercMultsyll = percentage of multisyllabic words, LowFreq = low frequency words, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, AIC = Akaike information criterion, BIC = Bayesian information criterion.

Table 8.
Fit indices and Chi-square difference tests for four- versus three-factors models of writing

	χ^2	df	*p-value*	χ^2/df	CFI	TLI	RMSEA	WRMR	χ^2diff[a]
Time 1									
Four Factor	22.35	10	0.01	2.24	0.95	0.95	0.09	0.64	
Three Factor	100.38	8	< .01	12.55	0.65	0.51	0.27	2.00	92.61**
Time 2									
Four Factor	17.57	9	0.04	1.95	0.94	0.90	0.08	0.68	
Three Factor	71.95	9	< .01	7.99	0.53	0.26	0.23	1.64	60.38**
Time 3									
Four Factor	15.66	10	0.10	1.57	0.95	0.94	0.07	0.62	
Three Factor	81.67	10	< .01	8.17	0.40	0.22	0.23	1.73	68.01**
Time 4									
Four Factor	47.06	9	< .01	5.23	0.68	0.61	0.19	1.14	
Three Factor	97.43	9	< .01	10.83	0.25	0.09	0.30	1.98	69.90**

Note. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, WRMR = Weighted Root Mean Square Residual

[a] Chi-square difference tests were conducted using the difftest command procedures outlined by Muthén & Muthén when the WLSMV estimator is utilized. This procedure adjusts the chi-square values and degrees of freedom to obtain a correct p-value.

* p < .05

** p < .01

Table 9.
Univariate latent change estimates for macro-organization

Fit statistics and Parameter Estimates	No Change Model 1	Constant Change Model 2	Dual Change Model 3
Fit statistics			
χ^2 (df)	105.94** (11)	9.37 (8)	9.21 (7)
χ^2/df	9.63	1.17	1.32
$\Delta\chi^2$ (Δdf)	-	105.57**[a]	.016[a]
CFI	0.00	0.95	0.91
TLI	-1.03	0.96	0.93
RMSEA	0.23	0.03	0.05
SRMR	0.35	0.07	0.08
AIC	1058.11	967.53	969.39
Parameter estimates			
Mean intercept	2.360**	2.068**	2.076**
Variance intercept	.058**	.093*	.091*
Mean slope	=0.0	.217**	0.00
Variance slope	=0.0	.002	.004
Correlation between intercept and slope	=0.0	-.51	-.80
Constant change parameter α[t]	=1.0	=1.0	=1.0
Proportional change path β	=0.0	=0.0	0.096
Residual variance Order of Ideas	0.358**	0.278**	0.278**

Note. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, AIC = Akaike information criterion
[a] Model fit compared to no change model
* $p < .05$
** $p < .01$

Table 10.
Univariate latent change estimates for productivity

Fit statistics and Parameter Estimates	No Change Model 1	Constant Change Model 2	Dual Change Model 3	Spline at Time 2	Dual Change with Spline
Fit statistics					
χ^2 (df)	585.86**(11)	75.75**(8)	42.69**(7)	24.57**(7)	24.46**(6)
χ^2/df	53.26	9.47	6.10	3.51	4.08
$\Delta\chi^2$ (Δdf)	-	510.11**[a]	33.06**[b]	51.18**[b]	0.11[c]
CFI	0.00	0.41	0.69	0.85	0.84
TLI	-1.75	0.55	0.73	0.87	0.84
RMSEA	0.58	0.23	0.18	0.13	0.14
SRMR	1.23	0.20	0.17	0.15	0.15
AIC	653.50	149.38	118.33	100.21	102.10
Parameter estimates					
Mean intercept	.75**	.38**	.42**	.442**	.442**
Variance intercept	0.01	0	0.01	.012*	.011*
Mean slope	=0.0	.27**	-0.02	.107*	.125*
Variance slope	=0.0	.01*	0.00	.001*	0.002
Correlation between intercept and slope	=0.0			.71*	.78
Constant change parameter α[t]	=1.0	=1.0	=1.0	=1.0 for t = 1-2, 3.332 for t = 2-4	=1.0 for t = 1-2, 3.103 for t = 2-4
Proportional change path β	=0.0	=0.0	.48**	=0.0	-.045
Residual variance Order of Ideas	.18**	.06**	.05**	.043**	.044**

Note. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, AIC = Akaike information criterion

[a] Model fit compared to no change model
[b] Model fit compared to constant change model
[c] model fit compared to constant change spline model
* $p < .05$
** $p < .01$

Table 11.
Univariate latent change estimates for complexity

Fit statistics and Parameter Estimates	No Change Model 1	Constant Change Model 2	Dual Change Model 3	Spline at Time 2	Dual Change with Spline
Fit statistics					
χ^2 (df)	68.91**(11)	16.78*(8)	10.75(7)	11.24(7)	9.33(6)
χ^2/df	6.26	2.10	1.54	1.61	1.56
$\Delta\chi^2$ (Δdf)	-	52.13**[a]	6.03*[b]	5.54*[b]	1.91[c]
CFI	0.00	0.36	0.73	0.69	0.76
TLI	-1.29	0.52	0.77	0.74	0.76
RMSEA	0.18	0.08	0.06	0.06	0.06
SRMR	0.38	0.14	0.12	0.12	0.11
AIC	2846.95	2800.82	2796.79	2797.28	2797.37
Parameter estimates					
Mean intercept	9.34**	8.22**	8.009**	8.000**	8.057**
Variance intercept	1.19*	3.40**	5.273**	5.081**	5.921**
Mean slope	=0.0	.82**	5.851**	1.477**	7.338**
Variance slope	=0.0	0.34	0.529	2.026	0.784
Correlation between intercept and slope	=0.0	-.84**	.06	-.86**	.28
Constant change parameter α[t]	=1.0	=1.0	=1.0	=1.0 for t = 1-2, 0.316 for t = 2-4	=1.0 for t = 1-2, 1.074 for t = 2-4
Proportional change path β	=0.0	=0.0	-.554**	=0.0	-.758**
Residual variance Order of Ideas	9.82**	8.19**	7.546**	7.636**	7.393**

Note. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, AIC = Akaike information criterion
[a] Model fit compared to no change model
[b] Model fit compared to constant change model
[c] model fit compared to constant change spline model
* p < .05
** p < .01

Table 12.
Univariate latent change estimates for vocabulary

Fit statistics and Parameter Estimates	No Change Model 1	Constant Change Model 2	Dual Change Model 3	Spline at Times 2 and 3	Dual Change with Spline at Times 2 and 3
Fit statistics					
χ^2 (df)	170.09**(11)	133.85(8)	32.25(7)	19.48(6)	21.14(5)
χ^2/df	15.46	16.73	4.68	3.25	4.23
$\Delta\chi^2$ (Δdf)	-	36.24**[a]	101.60**[b]	114.37**[b]	1.66[c]
CFI	0.00	0.00	0.00	0.00	0.00
TLI	-34.27	-37.37	-7.80	-4.48	-6.84
RMSEA	0.30	0.32	0.15	0.12	0.14
SRMR	1.80	1.27	1.18	1.09	0.99
AIC	-893.90	-924.14	-1023.73	-1034.50	-1030.84
Parameter estimates					
Mean intercept	1.21**	1.19**	1.15**	1.151**	1.152**
Variance intercept	0	0	0	.002*	0.002
Mean slope	=0.0	.01**	2.02**	.129**	.375*
Variance slope	=0.0	0.00	0.00	.006*	-0.001
Correlation between intercept and slope	=0.0		.57**	-.63**	
Constant change parameter α[t]	=1.0	=1.0	=1.0	=1.0 for t = 1-2, .766 for t = 2-3, .291 for t=3-4	=1.0 for t = 1-2, .467 for t = 2-3, .801 for t=3-4
Proportional change path β	=0.0	=0.0	-1.66**	=0.0	-.215
Residual variance Order of Ideas	.01**	.01**	.01**	.007**	.008**

Note. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, AIC = Akaike information criterion
[a] Model fit compared to no change model
[b] Model fit compared to constant change model
[c] model fit compared to constant change spline model
* p < .05
** p < .01

Table 13.
Correlations among vocabulary indices over time

	1	2	3	4
1. T1AvgSyllables	1			
2. T2AvgSyllables	.06	1		
3. T3AvgSyllables	.11	.13	1	
4. T4AvgSyllables	.14	.10	.14	1

	1	2	3	4
1. T1CharactersperWord	1			
2. T2CharactersperWord	-.03	1		
3. T3CharactersperWord	.18*	.14	1	
4. T4CharactersperWord	.31*	-.02	.28*	1

	1	2	3	4
1. T1PercentMultsyllabic	1			
2. T2PercentMultsyllabic	.08	1		
3. T3PercentMultsyllabic	.04	.18*	1	
4. T4PercentMultsyllabic	.08	.09	.17	1

Note. T1 = Time 1, T2 = Time 2, T3 = Time 3, T4 = Time 4.
*$p < .05$

Table 14.
Best fitting univariate models for macro-organization, productivity, and complexity factors

Fit statistics and Parameter Estimates	Macro-Organization Constant Change	Productivity Spline at Time 2	Complexity Dual Change
Fit statistics			
χ^2 (df)	9.37 (8)	24.57** (7)	10.75 (7)
χ^2/(df)	1.17	3.51	1.54
CFI	0.95	0.85	0.73
TLI	0.96	0.87	0.77
RMSEA	0.03	0.13	0.06
SRMR	0.07	0.15	0.12
Parameter estimates			
Mean intercept	2.068**	.442**	8.009**
Variance intercept	.093*	.012*	5.273**
Mean slope	.217**	.107*	5.851**
Variance slope	.002	.001*	0.529
Correlation between intercept and slope	-.51	.71*	.06
Constant change parameter $\alpha[t]$	=1.0	=1.0 for t = 1-2, 3.332 for t = 2-4	=1.0
Proportional change path β	=0.0	=0.0	-.554**
Residual variance Order of Ideas	0.278**	.043**	7.546**

* $p < .05$
** $p < .01$

Table 15.
Bivariate latent change model of macro-organization and complexity

Fit statistics and Parameter Estimates	No Coupling	Macro as Leader	Complexity as Leader	Bidirectional
Fit statistics				
χ^2 (df)	40.52	38.44	40.50	38.44
	(26)	(25)	(25)	(24)
$\Delta\chi^2$ (Δdf)	-	2.08 (1)	.02 (1)	2.08 (2)
CFI	0.78	0.79	0.76	0.78
TLI	0.76	0.77	0.73	0.74
AIC	3752.77	3572.69	3754.75	3754.69
RMSEA	0.06	0.06	0.06	0.06
SRMR	0.10	0.09	0.10	0.09
Parameter estimates for macro-organization variable				
Mean intercept	2.066**	2.066**	2.063**	2.065**
Variance intercept	0.091*	0.096*	0.093*	0.097*
Mean slope	0.217**	0.217**	0.272	0.235
Variance slope	0.002	0.003	0.001	0.003
Correlation between intercept and slope	-.505	-.522	-.479	-.510
Proportional change path β	=0.0	=0.0	=0.0	=0.0
Residual variance	0.279**	0.276**	0.279**	0.276**
Parameter estimates for complexity variable				
Mean intercept	8.012**	8.059**	8.012**	8.059**
Variance intercept	5.267**	6.108**	5.268**	6.110**
Mean slope	5.785**	3.770*	5.799**	3.795
Variance slope	0.542	1.289	0.524	1.285
Correlation between intercept and slope	0.043	-0.04	0.055	-0.034
Proportional change path β	-.547**	-.881**	-.548**	-.884**
Residual variance	7.524**	7.270**	7.544**	7.276**
Cross-variable correlations				
Macro Intercept, Complex Intercept	.891**	.756**	.893**	.759**
Macro Intercept, Complex Slope	-.098	-.29	-.102	-.289
Macro Slope, Complex Intercept	-1.461	-.862	-1.648	-.85
Macro Slope, Complex Slope	.297	-.174	.498	-.158
Correlated residuals	.027	.048	.026	.047
Longitudinal couplings				
Macro[t -1] → ΔComplex[t]	= 0	2.226	= 0	2.227
Complex[t -1] → ΔMacro[t]	= 0	= 0	-0.006	-0.002

Note. Model fit compared to no change model.
* $p < .05$
** $p < .01$

Table 16.
Bivariate latent change model of macro-organization and productivity

Fit statistics and Parameter Estimates	No Coupling	Macro as Leader	Productivity as Leader	Bidirectional
Fit statistics				
χ^2 (df)	52.19	52.02	51.25	51.19
	(27)	(26)	(26)	(25)
$\Delta\chi^2$ (Δdf)	-	0.17 (1)	.94 (1)	1.00 (2)
CFI	.87	.87	.87	.87
TLI	.87	.86	.86	.85
AIC	1020.70	1022.53	1021.76	1023.70
RMSEA	.08	.08	.08	.08
SRMR	.10	.10	.10	.10
Parameter estimates for macro-organization variable				
Mean intercept	2.067**	2.067**	2.083**	2.082**
Variance intercept	0.094*	0.094*	0.094*	0.095*
Mean slope	0.217**	0.218**	0.108	0.113
Variance slope	0.003	0.003	0.004	0.004
Correlation between intercept and slope	-.521	-.526	-.721*	-.715
Proportional change path β	=0.0	=0.0	=0.0	=0.0
Residual variance	0.278**	0.278**	0.278**	0.278**
Parameter estimates for productivity variable				
Mean intercept	0.443**	0.438**	0.442**	0.440**
Variance intercept	0.012**	0.012**	0.012*	0.012*
Mean slope	0.107**	0.100**	0.107**	0.103**
Variance slope	0.001**	0.001**	0.001**	0.001**
Correlation between intercept and slope	.673*	.663*	.691*	.684*
Proportional change path β	=0.0	=0.0	=0.0	=0.0
Residual variance	0.043**	0.043**	0.043**	0.043**
Cross-variable correlations				
Macro Intercept, Productivity Intercept	.939**	.949**	.949**	.956**
Macro Intercept, Productivity Slope	.183	.152	.209	.189
Macro Slope, Productivity Intercept	-.821	-.846	-1.073	-1.078
Macro Slope, Productivity Slope	.855	.884	.318	.364
Correlated residuals	.144*	.142*	.146*	.144*
Longitudinal couplings				
Macro[t -1] → ΔProductivity[t]	= 0	0.008	= 0	0.005
Productivity[t -1] → ΔMacro[t]	= 0	= 0	0.179	0.170

Note. Model fit compared to no change model.
* $p < .05$
** $p < .01$

Table 17.
Bivariate latent change model of productivity and complexity

Fit statistics and Parameter Estimates	No Coupling	Productivity as Leader	Complexity as Leader	Bidirectional
Fit statistics				
χ^2 (df)	59.57	54.36	59.56	54.19
	(26)	(25)	(25)	(24)
$\Delta\chi^2$ (Δdf)	-	5.21* (1)	0.01 (1)	5.38 (1)
CFI	.79	.82	.79	.82
TLI	.78	.80	.76	.78
AIC	2877.95	2874.74	2879.94	2876.57
RMSEA	.09	.09	.09	.09
SRMR	.12	.11	.12	.11
Parameter estimates for Productivity variable				
Mean intercept	0.444**	0.443**	0.445**	0.447**
Variance intercept	0.011*	0.011*	0.011*	0.011*
Mean slope	0.106**	0.106**	0.107**	0.112**
Variance slope	0.001**	0.001**	0.001**	0.001**
Correlation between intercept and slope	.901*	.848*	.902*	.852*
Proportional change path β	=0.0	=0.0	=0.0	=0.0
Residual variance	0.044**	0.044**	0.044**	0.044**
Parameter estimates for Complexity variable				
Mean intercept	8.103**	8.060**	8.015**	8.067**
Variance intercept	5.226**	5.836**	5.239**	5.918**
Mean slope	6.201**	7.323**	6.193**	7.295**
Variance slope	0.619	1.065	0.617	1.066
Correlation between intercept and slope	.101	.258	.100	.255
Proportional change path β	-0.593**	-.852**	-.593	-.852**
Residual variance	7.508**	7.277**	7.506**	7.264**
Cross-variable correlations				
Productivity Intercept, Complexity Intercept	.276	.192	.280	.207
Productivity Intercept, Complexity Slope	.288	.093	.288	.090
Productivity Slope, Complexity Intercept	.351	.415*	.356	.433*
Productivity Slope, Complexity Slope	.054	-.187	.058	-.173
Correlated residuals	.214**	.220**	.214**	.219**
Longitudinal couplings				
Productivity[t -1] → ΔComplexity[t]	= 0	2.001*	= 0	2.049*
Complexity[t -1] → ΔProductivity[t]	= 0	= 0	-0.001	-0.002

Note. Model fit compared to no change model.
* $p < .05$
** $p < .01$

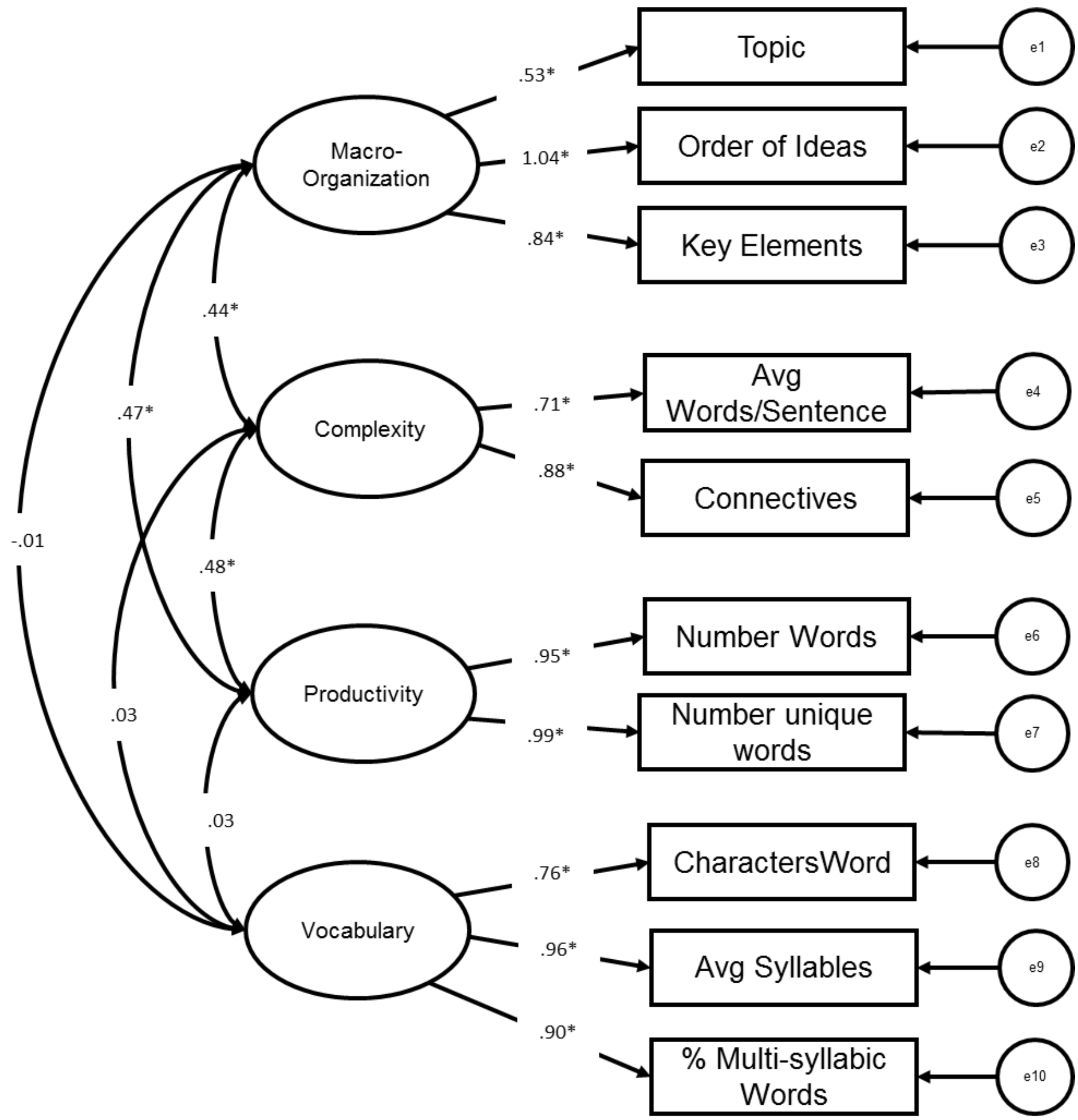

Figure 1. Four factor model of writing at Time 1.
* $p < .05$
** $p < .01$

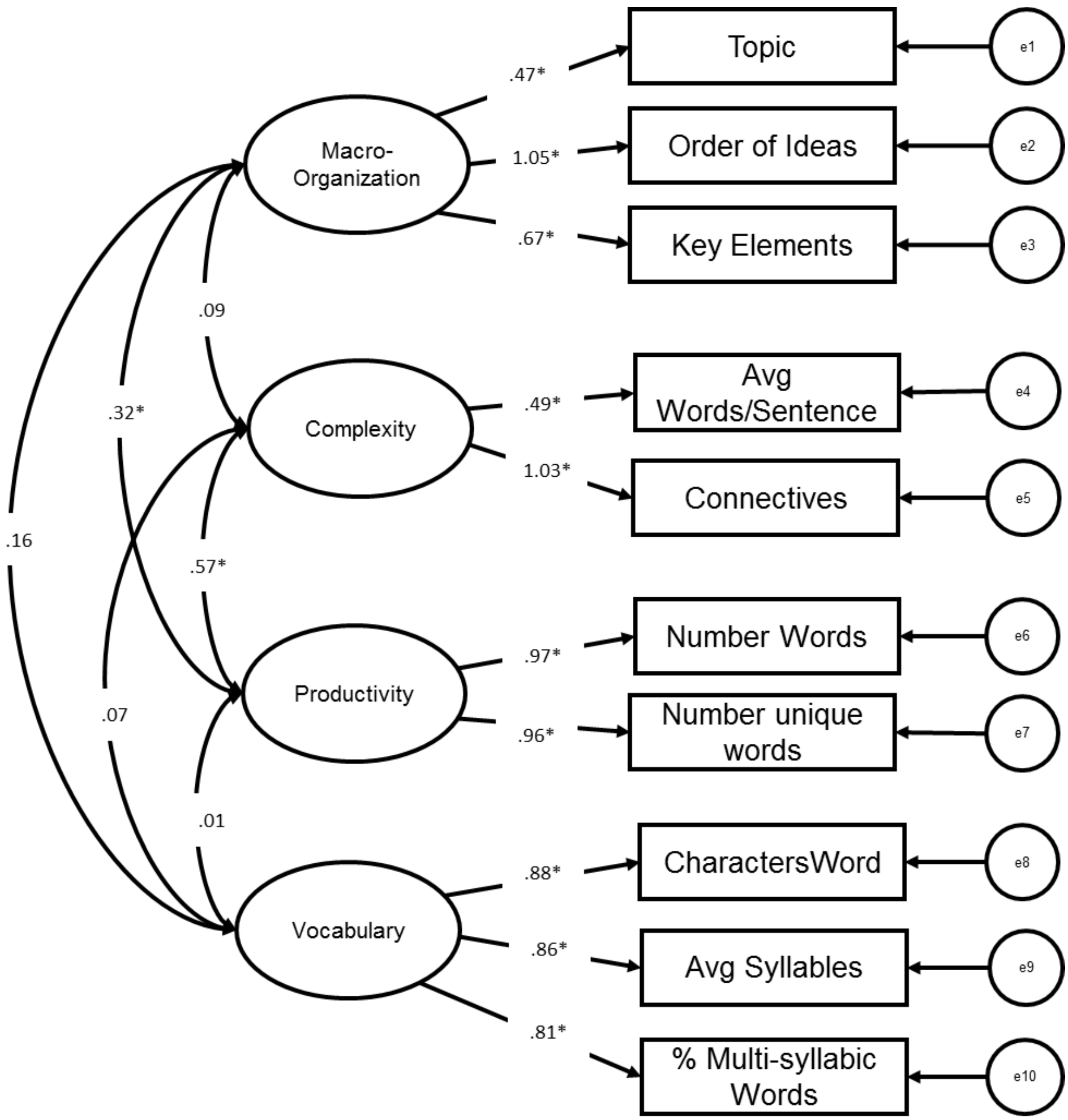

Figure 2. Four factor model of writing at Time 2.
* $p < .05$
** $p < .01$

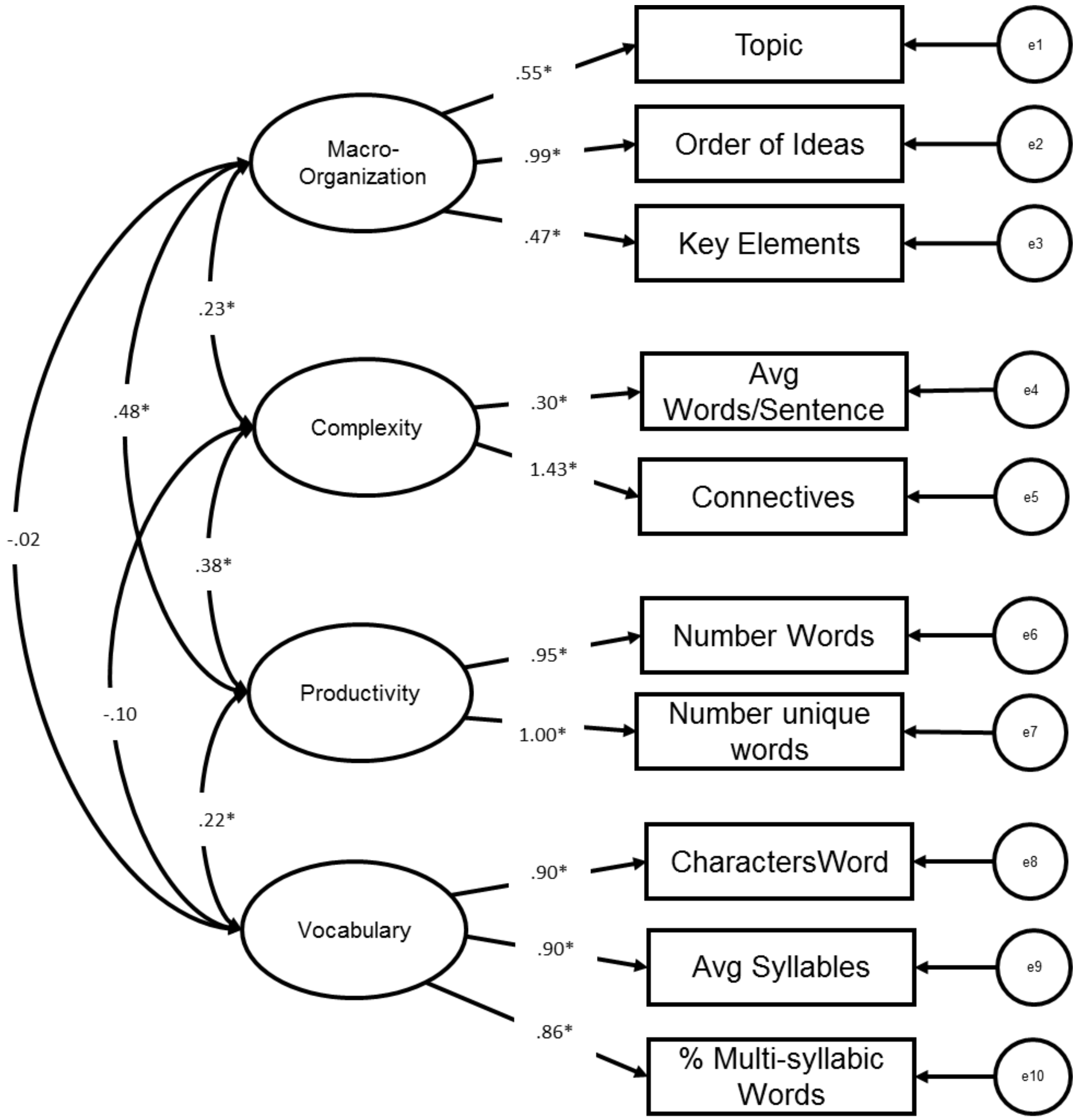

Figure 3. Four factor model of writing at Time 3.
* $p < .05$
** $p < .01$

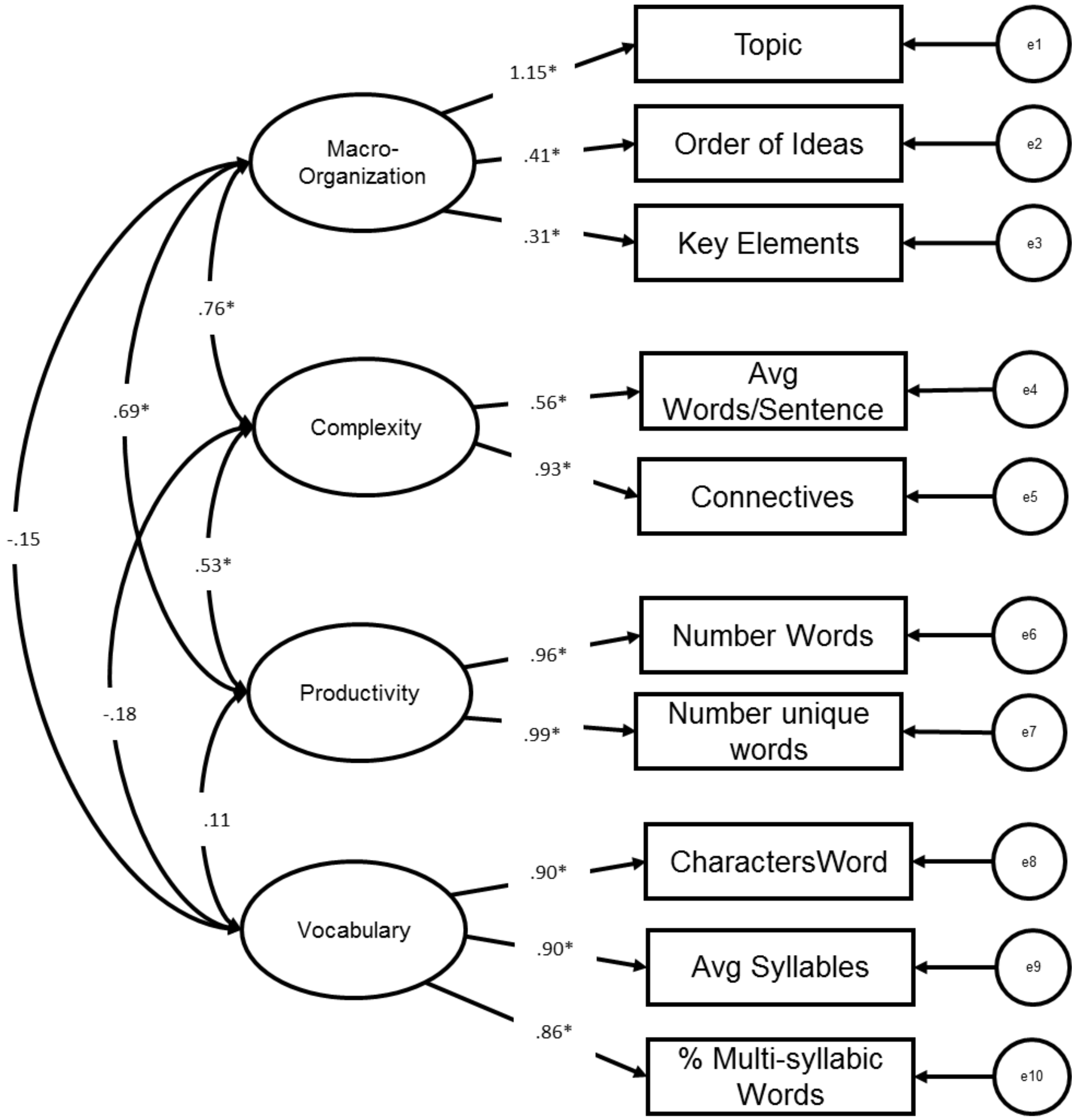

Figure 4. Four factor model of writing at Time 4.
* $p < .05$
** $p < .01$

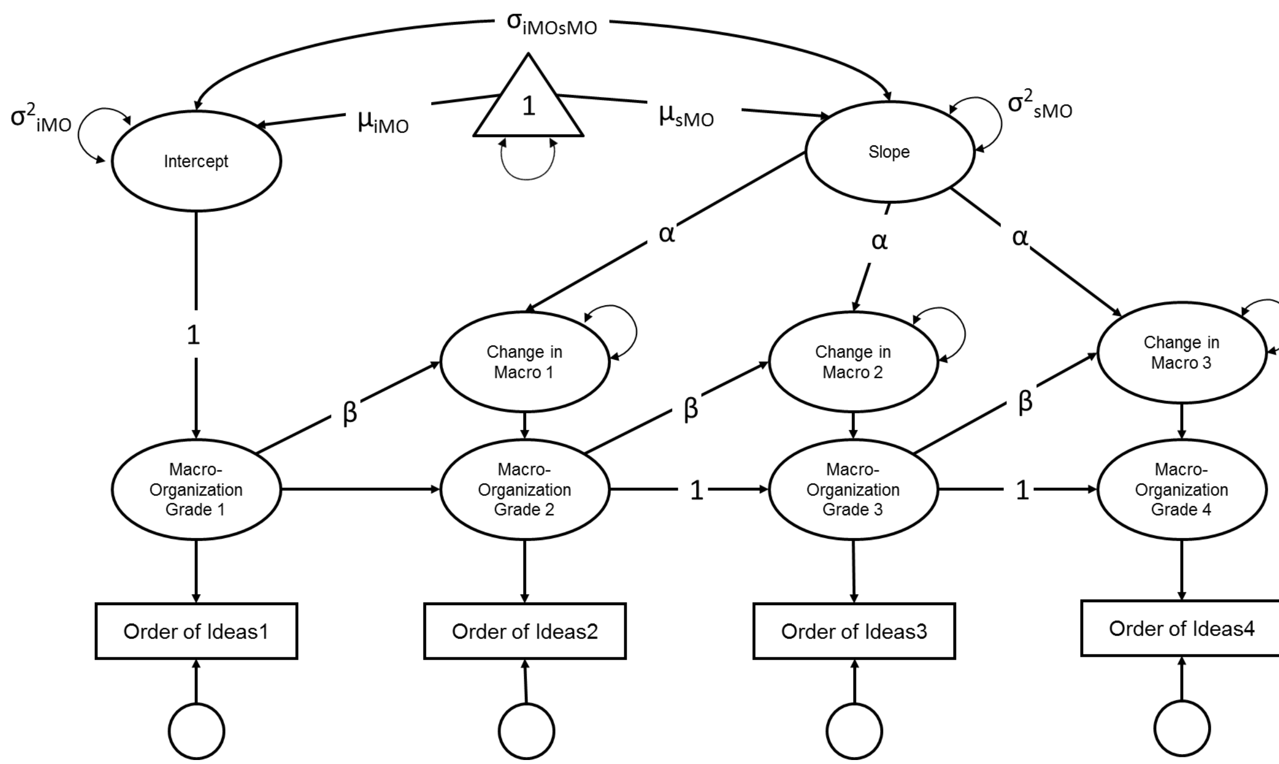

Figure 5. Univariate latent change score model of macro-organization.

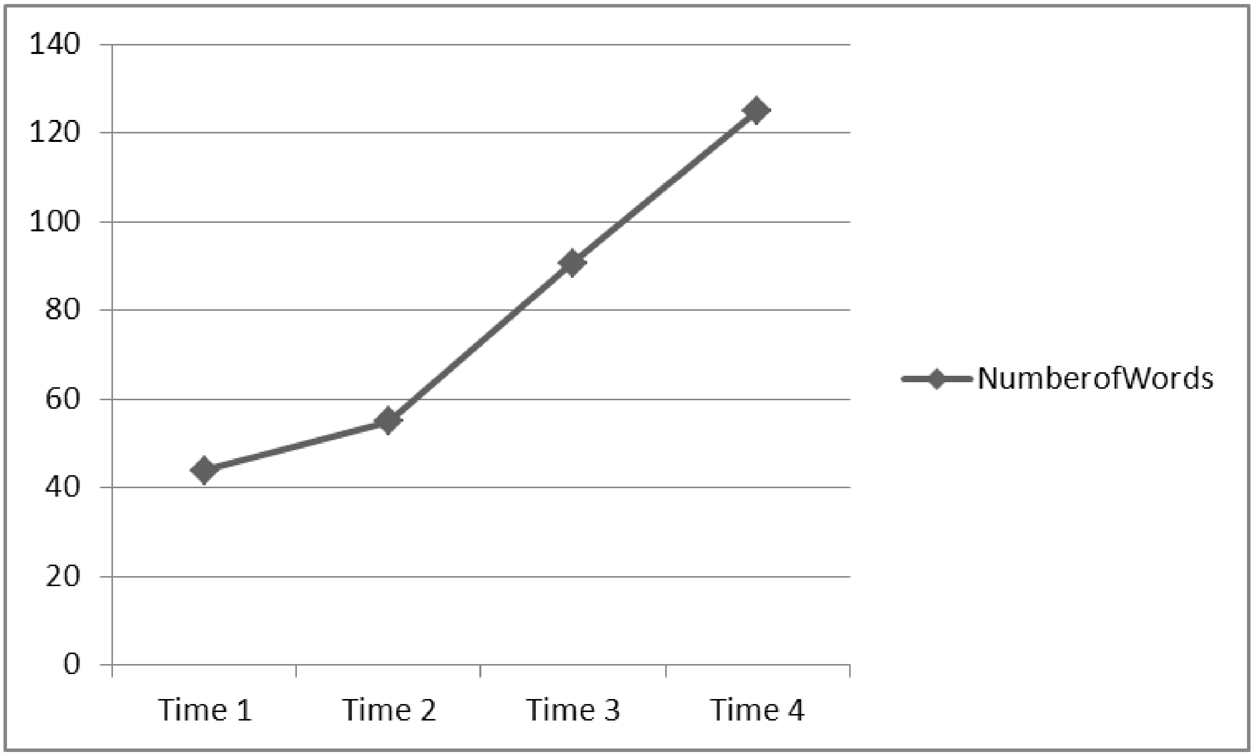

Figure 6. Plot of productivity indicator means over time.

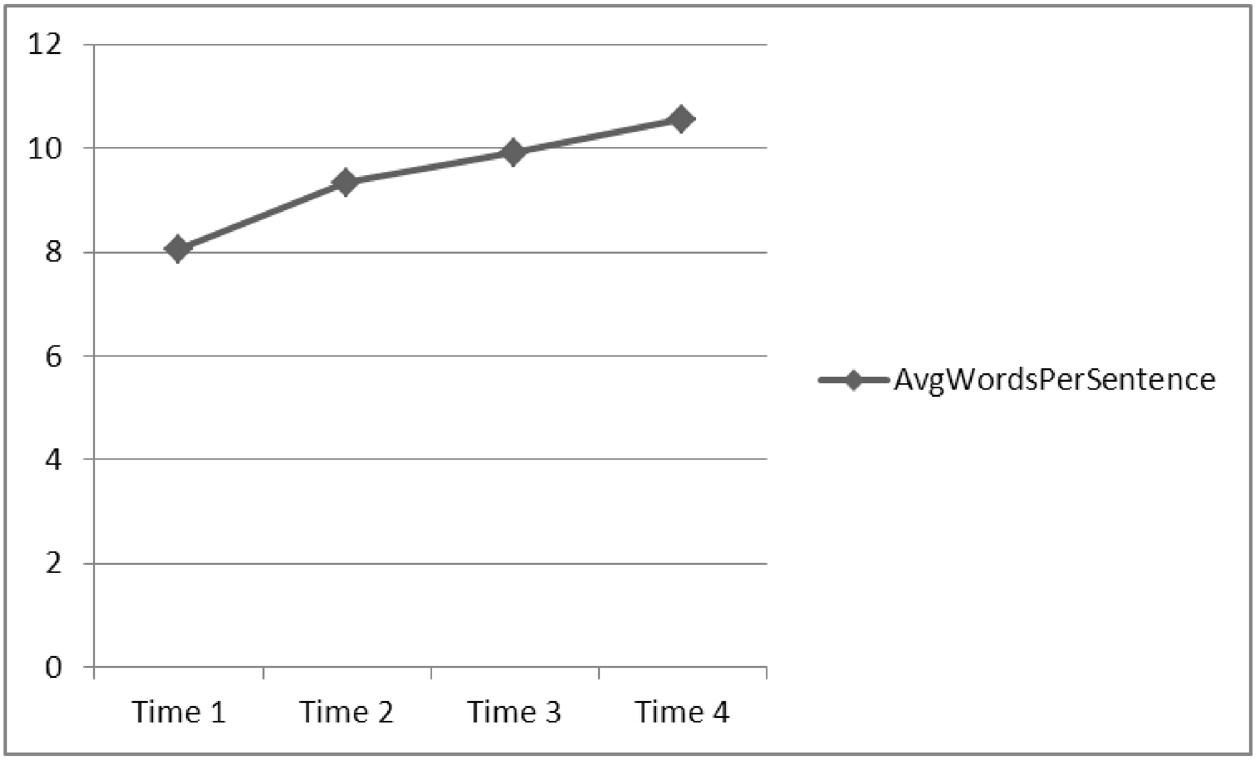

*Figure 7***.** Plot of complexity indicator means over time.

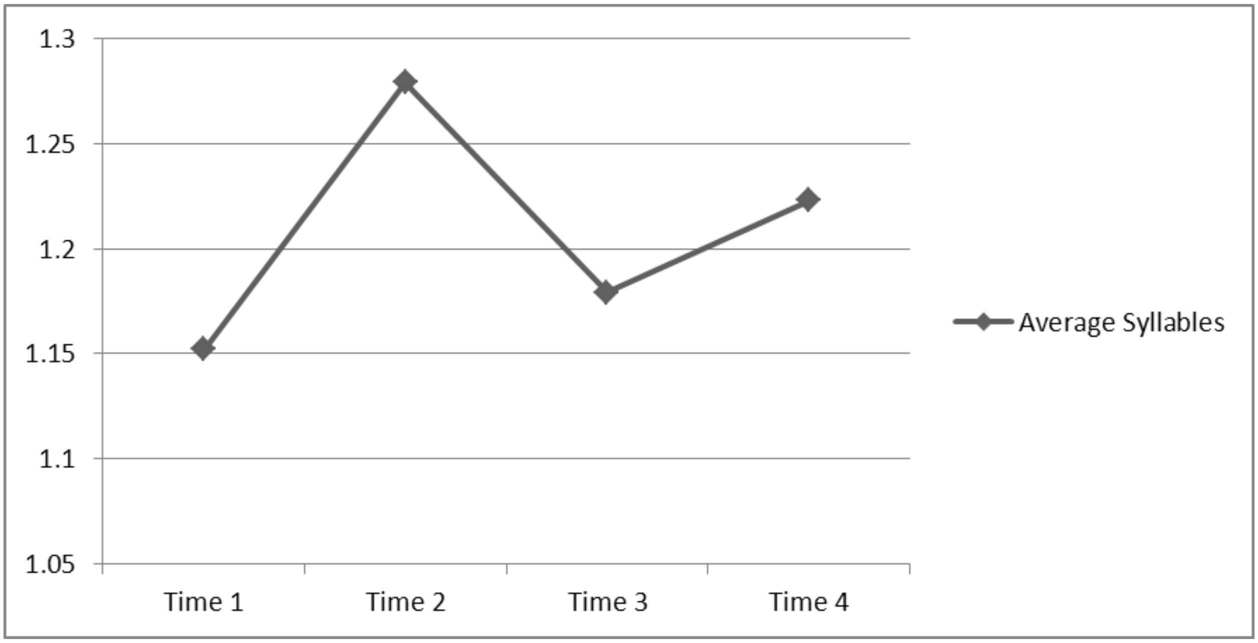

Figure 8. Plot of vocabulary indicator means over time.

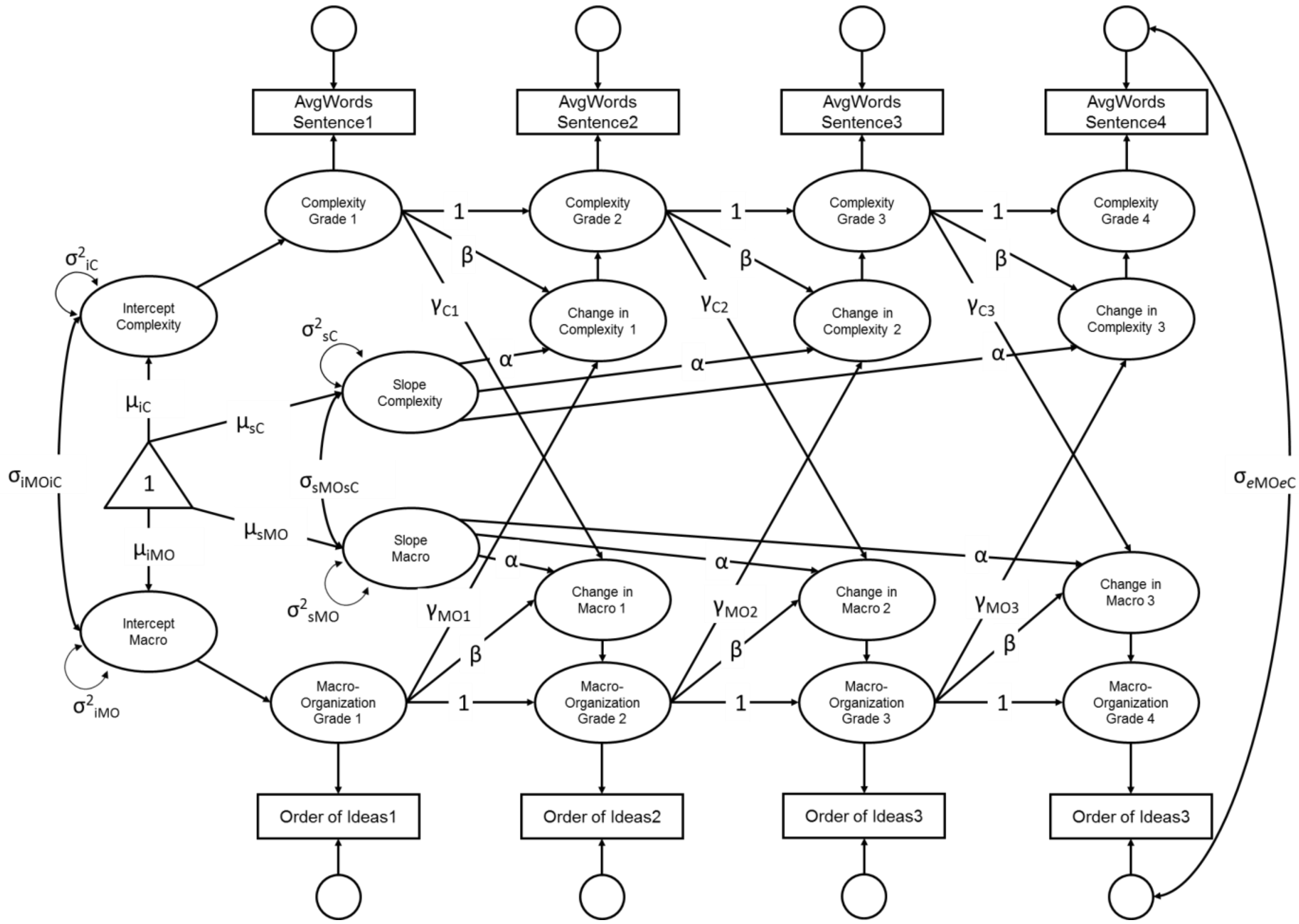

Figure 9. Bivariate latent change score model of macro-organization and complexity.

APPENDIX A

Rubric for logical order of ideas

Expository (Times 1 and 2)

1 = Structure. No sense of order. Transitions/sequencing is not present.
2 =Structure is unclear or only starting to emerge. No use of sequencing words when necessary.
3 = There is basic order with a few missteps. There is a clear use of two of the story structure: beginning, and middle, OR middle and end. Sentence parts may be linked with conjunction or connecting words (but, and, or, so).
4 = Ideas follow a logical sequence. There is a clear beginning, middle, and end. Everything fits together nicely.

Narratives (Times 3 and 4)

1 = Structure is absent. No sense of order. Transitions/sequencing is not present.
2 = Structure is unclear or only starting to emerge. Sequencing is confusing: The piece is little more than a list of sentences (2 or more) connected by theme (i.e., could move sentences around and not change the meaning). No use of sequencing words when necessary.
3 = There is basic order with a few missteps. There is a clear use of two of the story structure: beginning, and middle, OR middle and end. Sentence parts may be linked with conjunction or connecting words (but, and, or, so). Simple sequencing words may be used (e.g., first, next, then) when necessary or deemed appropriate.
4 = Ideas follow a logical sequence. There is a clear beginning, middle, and end. Everything fits together nicely. Clear transitions ((e.g., first, next, then, later, finally, etc.), including uncommon transition words (e.g., recently) connect one sentence to the next.

APPENDIX B

Transcription guidelines

As-is

Writing samples were typed exactly as the children wrote them, with no corrections or additions.

When transcribers had a different interpretation of a letter, the child was given the benefit of the doubt. For example, the letters 'a' and 'u' can be difficult to distinguish due to a child's handwriting style. If the interpretations are 'about' vs. 'ubout', child is given the benefit of the doubt and it was transcribed as 'about'. Another example was the capitalization of letters. Certain letters such as s, c, and p are difficult to interpret as lowercase or uppercase in some cases. The child was given the benefit of the doubt here as well.

Additionally, if the child drew a picture as part of the text, then it was transcribed as such, "I [drew a heart] pizza." Symbols such as "&" and "+" are typed as such.

If a child did not finish his or her sentence after time was called or did not write a complete word, it was left in the as-is version and indicated at the end of the transcript as 'abandoned/end of transcript.' For example, [I wa...] abandoned/end of transcript.

Example 1:

> Books I Like Books Becuse You can read and I Like to read Sometimes it is esy sometimes it is hard. they can Be aBute Animals and PePole and ABwtenathure and space and nuburs and ABC and Frute and vesdibuls. they can Be aBute School and Daycare and cars and the earth and fish. Sharks and math and
> [and...end of transcript]

Example 2:

> I K lrniguboteinsexs it is cool Because thermite be suptheg that you mit not now and sum tims you mitetuch it and it is rile cool to and you mite see its cave and it is rile cool when you git too looc at haw big it is and ispeshle when you go in the eave and see oll the cool thigs in ther and it is rile cool.

Clean

For the clean version, writing samples were corrected for spelling, capitalization, and end punctuation.

If a word was spelled incorrectly in the context of the sentence, but was spelled correctly as an actual word, then it was left as is. For example, "The hole class went." Since "hole' is spelled correctly, regardless of context, it was left as the child spelled it.

Capitalizations of words were corrected. Proper nouns and words that started a sentence were capitalized. Unnecessary capitalizations were also corrected, such as "doG" corrected to dog."

End punctuation marks were corrected and/or added when necessary. If a period was placed incorrectly in a sentence, it was taken out and corrected. Run-on sentences were not corrected as long as they had a conjunction between the clauses. Misuses of commas, quotations, etc., were not corrected in this version, only end punctuation.

If a numeral or digit was written in the sample, it was left as a digit. If it was written as a word, it was kept as a word. For example,"'4" was left as a digit and not corrected to "four."

If a symbol or drawing was used in place a word within the writing sample, it was changed to the word with the most common meaning. For example, "I [drew a heart] pizza," would be changed to "I love pizza." Additionally, symbols such as "&" or "+" were changed in the clean versions to the word "and." Drawings that are meant for decoration or have no word equivalent, such as a smiley face, were removed from the transcription.

If the writing sample contained an "abandoned/end of transcript" notation, a period was added to the end of the abandoned sentence. If there was an incomplete or unfinished word, it was corrected or deleted if it could not be interpreted.

Example 1:

> Books. I like books because you can read and I like to read. Sometimes it is easy. Sometimes it is hard. They can be about animals and people and about nature and space and numbers and ABC and fruit and vegetables. They can be about school and daycare and cars and the earth and fish. Sharks and math and
> [and...end of transcript]

Example 2:

> I like learning about insects. It is cool because there mite be something that you might not now and sum times you mite touch it and it is rile cool to and you mite see its cave and it is rile cool when you get too look at haw big it is and especially when you go in the eave and see all the cool things in there and it is rile cool.

Proper

For the proper version, writing samples were corrected for spelling, punctuation, grammar, usage, and syntax, while maintaining the fundamental nature of the child's original text. Commas, quotations, and words were changed, added, and/or deleted and when necessary for syntax. Additionally, the order of sentences could be switched For example, “Me and my friends…” would be changed to “My friends and I…”

If the writing sample contained an "abandoned/end of transcript" notation that could not be corrected into a complete sentence, then it was removed from the transcript.

Example 1:

> Books. I like books because you can read and I like to read. Sometimes it is easy. Sometimes it is hard. They can be about animals and people and about nature and space and numbers and ABC and fruit and vegetables. They can be about school and daycare and cars and the earth and fish. Sharks and math.

Example 2:

> I like learning about insects. It is cool because there might be something that you might not know and sometimes you might touch it and it is really cool too and you might see its cave and it is really cool when you get to look at how big it is and especially when you go in the cave and see all the cool things in there and it is really cool.

APPENDIX C

IRB Approval

Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 5/11/2012

To: Patricia Kantor

Address: Department of Psychology, 1107 W. Call Street, Tallahassee, FL 32306-4301
Dept.: PSYCHOLOGY DEPARTMENT

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Development of Written Language

The application that you submitted to this office in regard to the use of human subjects in the research proposal referenced above has been reviewed by the Human Subjects Committee at its meeting on 05/09/2012. Your project was approved by the Committee.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 5/8/2013 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is FWA00000168/IRB number IRB00000446.

Cc: Richard Wagner, Advisor
HSC No. 2012.8289

APPENDIX D

Sample Parent Consent Letter

The Florida State University
Department of Psychology
1107 W Call St.
Tallahassee, Florida 32306-4301
Voice 850/644-2040 Fax 850/644-7739

PLEASE SIGN AND RETURN TO TEACHER

Dear Parent,

I am a professor in the Department of Psychology at Florida State University. I am conducting a research study to help us better understand the kinds of skills that promote reading and writing development. For the past 2 years, you kindly gave permission for your child to be part of a large group of students who were given a set of reading and writing tasks. Because we would like to check the progress in reading and writing development that each child has made over the past 2 years, we are asking parents of all children who participated before to consent to have their child participate once again.

Your child's participation will involve reading silently, reading out loud, writing, listening to reading passages, and tasks that most children find enjoyable. One of the reading out loud tasks will be recorded. The tasks will be given individually in an open area at your child's school to children whose parents or guardians have given consent. The total time involved is about 90 minutes. We will also look at the results of the standardized reading comprehension test, SAT-10, and the "Writes Upon Request" assignment from previous years that they were administered. These reading and writing assessments are part of your child's regular elementary school work.

Your child's participation in this study is voluntary. Your consent may be withdrawn at any time. There will be no penalty and it will not affect your child's grade. Your child can also choose not to participate without penalty.

The results of the study may be published, but your child's name will not be used. Confidentiality of records will be maintained to the full extend allowed by Florida law. Possible benefits of your child's participation include increased understanding of the kinds of skills that are needed for successful reading and writing.

If you have any questions concerning this study or your participation, please contact me at rkwagner@psy.fsu.edu or at 850-644-1033.

Sincerely,

Richard K. Wagner
Alfred Binet Professor of Psychology

I GIVE permission for my child, ____________________ to participate in the above study and for my child's SAT-10 scores and Writes Upon Request assignments to be provided to the researchers of this study.

Parent's Name: Signature Date: ________

School Your Child is Attending in 2009- 2010: ____________________

Your Mailing Address: ____________________

____

If you have any questions about your child's rights as a participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subject's Committee, Institutional Review Board, through the Vice President for the Office of Research at (850) 644-8633.

FSU Human Subjects Committee approved on 9/10/09 VOID after 9/8/2010 HSC# 2009.3152

REFERENCES

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*, 281-298. Doi:10.1037/a0019318

ACT, Inc. (2007). *Writing specifications for the 2011 National Assessment of Educational Progress*. Iowa City, IA: Author.

Attali, Y., & Powers, D. (2008). *A developmental writing scale (ETS Research Report No. RR-08-19)*. Retrieved from Education Testing Service website: http://www.ets.org/Media/Research/pdf/RR-08-19.pdf

Berninger, V., Abbott, R., Swanson, H. L., Lovitt, D., Trivedi, P., Lin, S., … & Amtmann, D. (2010). Relationship of word- and sentence-level working memory to reading and writing in second, fourth, and sixth grade. *Language, Speech, and Hearing Services in Schools, 41*, 179–193.

Berninger, V. W., Abbott, R. D., Whitaker, D., Sylvester, L., & Nolen, S. B. (1995). Integrating low- and high-level skills in instructional protocols for writing disabilities. *Learning Disability Quarterly*, 18, 293-309.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equations models* (pp. 99-146). Newbury Park, CA: Sage.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466. Doi:10.1037/0033-2909.105.3.456

Culham, R. (2003). *6+1 traits of writing: The complete guides, grades 3 and up*. New York: Scholastic Professional Books.

Diamantopoulos, A. and Siguaw, J.A. (2000). *Introducing LISREL*. London: Sage Publications.

Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Education Northwest. (2012*). About 6+1 Trait Writing*. Portland, OR: Author. Retrieved from Education Northwest website: http://educationnorthwest.org/resource/949

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, 19, 149-154. Doi:10.1177/0963721410370300

Gerstorf, D., Hoppman, C. A., Kadlec, K. M., & McArdle, J. J. (2009). Memory and depressive symptoms are dynamically linked among married couples: Longitudinal evidence from the AHEAD study. *Developmental Psychology, 45*, 1595-1610. Doi:10.1037/a0016346

Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: A review. Learning Disability Quarterly, 22, 78–98.

Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide* (NCEE 2012-4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch

Graham, S., Harris, K. R., & Loynachan, C. (1993). The basic spelling vocabulary list. *Journal of Educational Research*, 86, 363-368.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading. A Carnegie Corporation Time to Act Report*. Washington, DC: Alliance for Excellent Education.

Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476. Doi:10.1037/0022-0663.99.3.445

Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling, 19*, 268-292. Doi:10.1080/10705511.2012.659627

Guan, Q., Ye, F., Wagner, R. K., & Meng, W. (2012). Developmental and individual differences in Chinese writing. *Reading and Writing*, online, 1-26. Doi:10.1007/s11145-012-9405-4

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6* (1), 1-55.

Huot, B., & Neel, M. (2006). Writing assessment: A techno-history. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research (*pp. 417-432). NY: The Guilford Press.

Jajodia, A. (2012). Dynamic structural equation models of change. In J. T. Newsom, R. N. Jones, & S. M. Hoker (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 291-328). New York: Routledge.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Juel, C., Griffith, P., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*, 243–255.

Kantor, P. T., & Wagner, R. K. (2012). *Modeling the development of prereaders' phonological processing skills: A latent variable longitudinal study*. Manuscript submitted for publication.

Kim, Y-S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, 21, 517-525. Doi:10.1016/j.lindif.2011.06.004

Kobrin, J. K., Deng, H., & Shaw, E. J. (2011).The association between SAT prompt characteristics, response features, and essay scores. Assessing Writing, 16, 154-169. Doi:10.1016/j.asw.2011.01.001

Kulikowich, J. M., Mason, L. H., & Brown, S. W. (2008). Evaluating fifth- and sixth-grade students' expository writing: Task development, scoring, and psychometric issues. *Reading and Writing, 21*, 153-175. Doi:10.1007/s11145-007-9068-8

Kouros, C. D., & Cummings, E. M. (2010). Longitudinal associations between husbands' and wives' depressive symptoms. *Journal of Marriage and Family, 72*, 135-147. Doi:10.1111/j.1741-3737.2009.00688.x

Little, R. J. A., & Rubin D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structual Equation Modeling: Present and Future* (pp. 342-380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577-605. Doi:10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Prindle, J. J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychology and Aging, 23*, 702-719. Doi:10.1037/a0014349

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading, 9*, 85-116. Doi:10.1207/s1532799xssr0902_1

Miller, J., & Chapman, R. (2001). *Systematic analysis of language transcripts (Version 7.0) [computer software]*. Madison, WI: Waisman Center, University of Wisconsin-Madison.

Muthén, B.O. (2004). Mplus Technical Appendices: Version 3. Los Angeles, CA: Muthén & Muthén

Muthén, L., & Muthén, B. (1998-2007). Mplus (Version 5.2) [Computer software]. Los Angeles: Muthén & Muthén.

Muthén, L. K., & Muthén, B. (2009). Mplus short courses, topic 1: Exploratory factor analysis, confirmatory factor analysis, and structural equation modeling for continuous outcomes (Slide 168). Retrieved from Mplus website: http://www.statmodel.com/download/Topic%201-v11.pdf

The National Commission on Writing for America's Families, Schools, and Colleges. (2004). *Writing: A ticket to work…or a ticket out, a survey of business leaders.* College Entrance Examination Board. Retrieved from College Board website: http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf

The National Commission on Writing For America's Families, Schools, and Colleges. (2005). *Writing: A powerful message from state government.* College Board. Retrieved from College Board website: http://www.collegeboard.com/prod_downloads/writingcom/powerful-message-from-state.pdf

The National Commission on Writing in America's Schools and Colleges. (2003). *The neglected R: The need for a writing revolution.* College Entrance Examination Board. Retrieved from College Board website: http://www.collegeboard.com/prod_downloads/writingcom/neglectedr.pdf

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from Common Core State Standards Initiative website: http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the* *subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading and Writing Quarterly, 23*, 287-309. Doi:10.1080/10573560701277807

No Child Left Behind Act of 2001, PL 107-110, 20 U.S.C. § 6301*et seq*.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545-565. Doi:10.1007/s11145-008-9124-z

Persky, H. R., Daane, M. C., & Jin, Y. (2003). The nation's report card: Writing 2002 (NCES 2003-529). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from National Center for Education Statistics website: http://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*, 525-556.

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*, 17, 102-120.

Scott, B. (2012). *Readabilityformulas.com: Spache readability calculator[web-based software]*. Retrieved from http://www.readabilityformulas.com/free-spache-readability-test.php.

Riley. (1996). Improving the reading and writing skills of America's students. *Learning Disability Quarterly, 19*, 67-69.

Salahu-Din, D., Persky, H., & Miller, J. (2008). The nation's report card: Writing 2007 (NCES 2008-486). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from National Center for Education Statistics website: http://nces.ed.gov/nationsreportcard/pdf/main2007/2008468.pdf

Sanders, T. J. M., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing: How text analysis provides insights in writing products and writing processes. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 386–402). New York: Guilford.

Steenkamp, J-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research, 25, 78-90.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42* (5), 893-98.

Strickland, D. S., Bodino, A., Buchan, K., Jones, K. M., Nelson, A., & Rosen, M. (2001). Teaching writing in a time of reform. *The Elementary School Journal, 101*, 385-397.

Tabachnick, B.G. and Fidell, L.S. (2007). Using multivariate statistics (5th ed.). New York: Allyn and Bacon.

Tindal & Parker (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research & Practice, 6*, 211-218.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (eds.), *Structural Equation Modeling: A Second Course* (pp. 119-169). Greenwich, CT: Information Age.

Wagner, R. K., Puranik, C. S., Foorman, B. R., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing, 24*, 203-220. Doi:10.1007/s11145-010-9266-7

Whitaker, D., Berninger, V. W., Johnston, J., & Swanson, L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences, 6*, 107-130.

Yu, C.Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation, University of California, Los Angeles.

BIOGRAPHICAL SKETCH

CURRICULUM VITAE
Patricia Thatcher Kantor

EDUCATION
Ph.D., Psychology, Developmental — December 2012
Florida State University — Tallahassee, FL
Dissertation Title: *Development of Writing: Key Components of Written Language*

M.S., Psychology, Developmental — May 2010
Florida State University — Tallahassee, FL
Thesis Title: *Modeling the Development of Prereaders' Phonological Processing Skills: A Latent Variable Longitudinal Study*

M.A., Early Childhood Education — May 2005
University of Central Florida — Orlando, FL

B.S., Psychology — May 2003
Florida State University — Tallahassee, FL
Minor: Women's Studies
Cum Laude, with Honors

AWARDS
Institute of Education Sciences Predoctoral Interdisciplinary Research Training Fellowship (2006-2011), Florida State University
W. Russell and Eugenia Morcom Scholarship (August 2011), Florida State University
W. Russell and Eugenia Morcom Scholarship (January 2012), Florida State University
W. Russell and Eugenia Morcom Scholarship (May 2012), Florida State University

PROFESSIONAL EXPERIENCE
Predoctoral Interdisciplinary Research Training Fellow — 2006-2011
Florida State University, Florida Center for Reading Research — Tallahassee, FL
Duties: Conducted high quality evaluation research; Trained in the use of RCT, strong quasi-experimental and longitudinal designs; Completed interdisciplinary courses in research methodology, measurement, quantitative analyses, core knowledge in reading, and practical issues in educational research; Submitted at least one publication-quality empirical manuscript each year.

RESEARCH
<u>Research Interests</u>
Assessment and development of emergent and evolving literacy skills
Identification and intervention of learning disabilities
Assessment and development of children’s writing skills
Educational applications of structural equation modeling, hierarchical linear modeling, item response theory, and longitudinal data analysis

Publications

Kantor, P. T., & Wagner, R. K. (2012). *Modeling the development of prereaders' phonological processing skills: A latent variable longitudinal study*. Manuscript submitted for publication.

Kantor, P. T., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2011). Comparing two forms of dynamic assessment and traditional assessment of preschool phonological awareness. *Journal of Learning Disabilities, 44*, 313-321. doi:10.1177/0022219411407861.

Ahmed, Y., Wagner, R. K., & **Kantor**, P. T. (in press). How visual word recognition is affected by developmental dyslexia. In J. S. Adelman (ed.) *Visual Word Recognition Vol. 2: Meaning and Context, Individuals and Development.* Hove, UK: Psychology Press.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & **Kantor**, P. T. (2011). Modeling the development of written language. *Reading and Writing, 24*, 203-220. doi: 10.1007/s11145-010-9266-7.

Kantor, P. T. & Kershaw, S. (2010). Parametric Statistics. In N. J. Salkind (Ed.), *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage Publications.

Wagner, R. K., Piasta, S. B., & **Kantor**, P. T. (2010). Latent variables. In N. J. Salkind (Ed.), *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage Publications.

Wagner, R. K., & **Kantor**, P. T. (2010). Dyslexia deciphered. In D. Priess & R. Sternberg (Eds.), *Innovations in Educational Psychology: Perspectives on Learning, Teaching, and Human Development* (pp 25-47). New York: Springer Publishing Company.

Presentations

Kantor, P. T., Wagner, R. K, & Ahmed, Y. (2011, July). *Development of Written Language: A Coh-Metrix Analysis of Writing Samples*. Poster presented at the annual meeting of the Society for the Scientific Study of Reading, St. Pete Beach, FL.

Kantor, P.T. (2010, April). *Dynamic assessment of early phonological* awareness. Talk presented at annual Department of Psychology Graduate Research Day, Florida State University.

Kantor, P. T., & Wagner, R. K. (2009, June). *Development of prereaders' phonological processing skills: A latent variable longitudinal study*. Poster presented at the annual meeting of the Society for the Scientific Study of Reading, Boston, MA.

Kantor, P. T., & Wagner, R. K. (2009, June). *Development of prereaders' phonological processing skills: A latent growth curve model*. Poster presented at Institute of Education Sciences 2009 Research Conference, Washington, D.C.

Kantor, P. T., & Wagner, R. K. (2008, July). *Improving prereaders' phonological awareness measures*. Poster session presented at the annual meeting of the Society for the Scientific Study of Reading, Asheville, NC.

Kantor, P. T., & Wagner, R. K. (2007, June 8). *Comparing assessment strategies for prereaders' phonological sensitivity measures*. Poster presented at Institute of Education Sciences 2007 Research Conference, Washington, D.C.

Quantitative Skills
Proficient in a range of analytic approaches including; Regression, Analysis of Variance, Structural Equation Modeling, Latent Growth Analysis, Latent Change Score Modeling, Hierarchical Linear Modeling, and Item Response Theory. I am skilled with the following statistical packages: *SPSS, Amos, Mplus, NORM, BILOG-MG, HLM*.

Ad hoc journal reviews
Scientific Studies of Reading
Journal of Learning Disabilities

Organizational Memberships
Society for the Scientific Studies of Reading, 2008-present

INSTRUCTION

Adjunct Instructor — July 2011 to May 2012
Tallahassee Community College — Tallahassee, FL
Courses Taught:

CLP1001	Psychology of Personal and Social Adjustment
PSY2012	General Psychology

Curriculum Coordinator and Lead Teacher, VPK — 2004-2006
Winter Park Presbyterian Preschool — Winter Park, FL

Lead Teacher — 2003-2004
Kids R Kids — Orlando, FL

CERTIFICATIONS

Certified Pre-K to Primary Teacher, Florida Department of Education, May 2005