Matthew Effects in Writing Productivity during Second Grade

Carla Wood

Florida State University

Christopher Schatschneider

Florida State University

Jeanne Wanzek

Vanderbilt University

Published 2020

Wood, C., Schatschneider, C., & Wanzek, J. (2020). Matthew effects in writing productivity during second grade. *Reading and Writing, 33*, 1377-1398.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170203 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ABSTRACT

The primary aims of this study were two-fold: a) to describe average change in the written narrative performance of second grade students from the fall and spring of the school year and b) examine patterns of individual growth to test for Matthew effects. Participants included 299 children in second grade. Microstructural measures were derived from students' written narratives including: number of different words (NDW), total number of words (TNW), and accuracy of spelling and grammar. Significant increases in NDW, TNW, and spelling accuracy were evidenced from fall to spring. Students averaged 55 total words in the fall and averaged 69 words in the spring, with a statistically significant increase of 14 words $t(299)=8.4$, $p<.0001$). The variance in TNW from fall to spring increased from Var=791 to Var=1005, which was a significant increase and the correlation of initial Fall TNW and growth in TNW was also significant ($r = 0.39$). Additionally, results from a two-level hierarchical linear model with students nested within teachers indicated that initial level of TNW predicted the change in TNW from fall to spring, with higher levels of initial TNW being related to larger gains in TNW. Significant predictors of Matthew effects included teacher or classroom and free/reduced lunch eligibility. Written personal narrative measures are sensitive to developmental change across a school year. Evidence of Matthew effects in lexical productivity suggests additional support may be warranted to ameliorate gaps in writing achievement.

Key Words: accuracy, achievement, assessment, grammar, language development, Matthew effects, second grade, spelling, vocabulary, writing

Matthew Effects in Writing Productivity during Second Grade

Written language is well recognized as an essential skill for academic success and performance on high stakes tests (Jenkins, Johnson & Hileman, 2004). Notably, students begin on the journey of learning to write in preschool and early elementary grades, with rapid growth expected in the early grades as students are expected to convey ideas clearly through writing in a relatively few number of years of formal instruction in writing. The development of students' writing skills has gained increased attention globally in the last decade. Research on writing has gained momentum across the globe as researchers study ways to cultivate writing skills (Camache & Alves, 2017) and attempt to model relationships between writing and other cognitive-linguistic skills for speakers from a multitude of language backgrounds (e.g., Harrison, Goegan, Jalbert, McManus, Sinclair & Sparling, 2016; Kim & Park, 2019; Yeung, Suk-han, Wai-ock, & Kein-hoa, 2013).

Among the many motivations to study writing development, are observed gaps and underachievement in students' writing achievement during the elementary school years. As one example, in the United States there has been increased awareness of the frequent failure of students to reach proficiency by fourth grade (National Center for Education Statistics, NCES, 2012). According to the Nation's Report Card (2012), only one-quarter of students perform at the proficient level in writing in fourth grade making it increasingly important to monitor writing development earlier and often. Additionally, in the United States, national statistics show the risk of underachievement in writing is even greater for specific demographic groups (e.g., students eligible for free and reduced lunch, students of from linguistically diverse backgrounds, and students living in rural areas) (NCES, 2012) adding to the need to monitor writing skill development early and frequently for students from disadvantaged backgrounds.

The risk for failure to attain writing proficiency appears to be disproportionally greater for students who enter school with low general language skills (NCES, 2012). This phenomenon has been examined in other domains and described as the *Matthew effect*, in which the rich get richer while the poor get poorer (Cook & Campbell, 1979; Merton, 1995; Rigney 2010). As Walberg and Tsai explained, "those who score higher than others on pretests…at the beginning of an experiment gain absolutely and relatively more than others from the same experience" (1983) (p. 360). The term captures the notion that initial advantage tends to foster further advantage. Further, for students with weaker language skills in the early grades, the disadvantage is associated with increasing disadvantage with widening achievement gaps over time.

Matthew effects have been previously studied in relation to students' reading achievement with inconsistent findings (e.g., Pfost, Hattie, Dorfler & Artelt, 2014). Pfost and colleagues (2014) provided a summary of empirical results of Matthew effects in reading after reviewing 28 articles including 78 separate results on inter-individual differences in reading. Among inclusion criteria, was the report of a covariance or correlation between baseline level and a growth component. Of the 78 results, 42% demonstrated decreasing gaps in reading achievement; 26% reflected stable gaps and 23% showed increasing achievement gaps. The authors identified challenges to detecting Matthew effects including ceiling effects on standardized measures, lack of precision in measures, and a low number of available studies.

Matthew effects have not been widely studied in relation to writing skills. The notion of Matthew effects can be viewed in relation to Merton's social theory of opportunity structure (Merton, 1995). Merton describes opportunity structures as "the scale and distribution of conditions that provide various probabilities for acting individuals and groups to achieve

specifiable outcomes" (p. 25). In short, initial advantages result in subsequent advantages. Applying this theory, students with better initial language skills would be expected to experience greater benefits from opportunities to attain writing skills, creating a fan spread growth or widening gap over time in which the gap widens between students at the low and high range of performance over time. Conversely, it is possible that students with low initial language skills catch up and close the performance gap during the school year. It is also possible that the slopes of change stay relatively stable and the slopes of change between low and high performers are relatively parallel to each other across the school year.

Theoretical Framework

The lexical quality hypothesis (Perfetti & Hart, 2002) provides additional support to suspect the presence of Matthew effects in lexical productivity. Based on this hypothesis, children vary in the quality of their lexical representations. Children with high quality representations demonstrate more complete semantic information, phonetic information, and fully specified orthographic representations (i.e., spelling). In contrast, children with low quality lexical representations may know a meaning but not readily retrieve the corresponding word or have incomplete orthographic representations of the word. The lexical quality hypothesis is generally applied to reading, in which less skilled readers have fewer high quality representations (Perfetti & Hart, 2002). Applying the hypothesis to writing, children with low quality representations may demonstrate less breadth and depth of vocabulary in their writing and may also get stuck when attempting to spell a word due to weak orthographic representations.

Although the lexical quality hypothesis is generally applied to reading, its relevance to lexical productivity in writing is based on the intertwined skills involved across the modalities of reading, writing, and oral language. It is generally accepted that oral language provides support

for writing (e.g., Kim & Schatschneider, 2017). The connectedness of reading, writing, and oral language is theoretically grounded in theories that emphasize the common underlying constructs or constellations of knowledge shared by oral language, reading, and writing (Fitzgerald & Shanahan, 2010; Shanahan, 2006).

Applying the theoretical perspective that writing is the encoding of oral language (Berninger & Amtmann, 2003; Kim & Schatschneider, 2016), writing has been viewed as componential skills of transcription, or the written production of letters and words (Singer & Bashir, 2004), and composing or generating ideas (Kim, 2016; Kim & Schatschneider, 2016; Kim, Park, & Park, 2015). From this perspective, writing and oral language skills are integrally intertwined, as writing relies on interconnected language skills including word knowledge. As such, we suspect that differences in quality of lexical representations may be detectable in early writing performance. This notion is supported by the fact that measures of writing have also been studied relative to the overlap in knowledge area with oral language and reading (Fitzgerald & Shanahan, 2000).

Microstructural Measures of Writing

Observable developmental change in writing during the early grades may be influenced by which aspects of writing are considered and how such components are measured. Dimensions of writing vary across studies. In a study of 186 first and fourth grade students by Wagner et al. (2011), results of a confirmatory factor analysis supported four composition factors including: macro-organization, complexity, productivity, and mechanical errors. Other studies have focused on the linear process of planning, writing, and revising (e.g. Goertz, Duffy, & LeFloch, 2001) and included measures of productivity, complexity, accuracy, and mechanics within such phases of the writing process (Koutosoftas & Gray 2013). Authors Koutsoftas and

Gray (2013) examined dimensions of writing within and across phases of writing through structural equation modeling for 267 typically developing students in 6th grade. Results supported a linear writing process of planning, translating ideas, and revising and the conceptualization of multiple factors to translating ideas including productivity, complexity, accuracy and mechanics. Although writing components of interest vary across studies, to investigate Matthew Effects we focus on several microstructural aspects of writing including lexical measures (e.g., diversity and productivity) and accuracy (e.g., spelling and grammar).

Among common indices of writing, lexical diversity and productivity are widely reported in the literature. Lexical productivity, generally measured by total number of words (TNW), has been used as a standard measure of fluency and productivity in curriculum based writing measures for decades (Martson, 1989) with reported correlations with the Test of Oral and Written Language as high as .84 (e.g., Deno, Marston, & Mirkin, 1982). Other studies have found weaker relationships depending on the assessment measure compared (e.g., Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). Similarly, lexical diversity, or number of different words (NDW) is highly correlated to productivity (authors, 2018) , and widely utilized in previous studies (Fey, Catts, Proctor-Williams, Tomblin, Zhang, 2004; Wagner et al., 2011) as a measure sensitive to developmental change. Additionally, previous studies have shown that greater diversity of word use is correlated with language proficiency levels (Grant & Ginther, 2000; Jarvis, 2002; Yu, 2009).

The rationale for attending to the lexical measures in the exploration of Matthew effects is multifaceted. First, lexical count measures offer potential sensitivity to Matthew effects given that TNW is not inhibited by ceiling effects and demonstrates a developmental progression in school age students (Fey et al., 2004; author et al., 2018). Further, the meaningfulness of writing

productivity is supported by significant relationships with standardized measures of vocabulary knowledge (Grant & Ginther, 2000; Miller, Andriacchi, & Nockerts, 2015; author et al., 2018) and relationships to writing quality for children in the elementary grades (e.g., Abbott & Berninger, 1993; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Wagner et al., 2011).

Among other measures considered in the constellation of writing, accuracy is generally included (Goertz et al., 2001; Koutosoftas & Gray 2013; Wagner et al., 2011). The consideration of spelling accuracy, specifically, is supported by the wide recognition of spelling as an essential component of writing (Berninger, Abbott, Nagy, & Carlisle, 2010; Devonshire & Fluck, 2010). Furthermore, spelling skills have been reported to predict text composition in students in first through seventh grade (Abbot, Berninger, & Fayol, 2010), are sensitive to developmental change in writing (Dockrell, Connelly, Walter, & Critten, 2015) and differentiate children with language learning difficulties (Broc, Bernicot, Olive, Favart, Reilly, Quémart, & Uzé, 2013).

In addition to spelling accuracy, grammatical accuracy is often considered in measuring students' writing. A number of previous findings support that measures of correct writing sequences are sensitive to student progress over time (Dockrell et al., 2015; Malecki & Jewell, 2003). Further, grammaticality, or proportion of grammatical errors, has been found to be sensitive to achievement differences and differentiates between children who are typically developing and children with language impairments (Eisenberg, & Guo, 2013; Scott & Windsor, 2000). In a study by Scott and Windsor (2000), the extent of grammatical error was the only measure that distinguished children with language learning disabilities from their peers.

Previous studies have coded accuracy in a number of different ways, including categories of spelling errors (e.g., Bahr, Silliman, Berninger, & Dow, 2012; Quick & Erickson, 2018; Masterson & Apel, 2013) and categories of grammaticality (Eisenberg & Guo, 2013). Other

studies have reported proportion of spelling errors (Dockrell et al., 2015). Similarly, grammaticality measures are often calculated using an error ratio (Eisenberg & Guo, 2013; Scott & Windsor, 2000) in which the number of grammar errors is measured in relation to total number of words.

Detecting Matthew Effects in Writing

Given that detection of Matthew Effects requires the use of measures that are sensitive to change across the school year and the use of measures and have an indefinite range to minimize constraints of potential ceiling effects, we focus in the current study on examination of Matthew effects in lexical productivity and accuracy. Evidence for the developmental sensitivity of lexical productivity measures of writing is provided by findings of previous studies that have examined average change in written language (e.g., authors, 2017). In one such study (Malecki & Jewell, 2003), investigators examined writing production on a three- minute writing task for 946 students in first- through eighth-grade administered at two time points (fall and spring). Students demonstrated significant increases in writing production and improvements on accuracy production indices from fall to spring time points, supporting the utility of such writing measures for being sensitive to change in writing across a school year. Similarly, in a study by Dockrell et al (2015), investigators examined written texts of 192 students who were in 3rd, 4th, and 5th grade. The authors reported significant effects of time across a 5-month period for total words produced in students' expository and narrative writing samples. Finally, similar findings were reported for students in early elementary grades in a study of average one year change in productivity and lexical diversity between subsequent grade levels for 749 children in first through eighth grade (author et al., 2018). Findings indicated that lexical productivity in written narratives was sensitive to one-year developmental change for students in first through third grade. These

previous findings support the expectation of growth in writing productivity and accuracy across the school year (e.g., Dockrell et al., 2015; Malecki & Jewell, 2003; Walter, & Critten, 2015; authors, 2018) but none that have considered Matthew effects to our knowledge.

Influencing Factors

In a study of Matthew effects in reading, by Morgan, Farkas, and Hibel (2008), the authors examined additional child- and family-level factors associated with differing availability of resources known to support literacy development (e.g., access to print at home). The authors considered gender, race/ethnicity, and social class background which have been established as predictors of reading growth rate (McCoach, O'Connell, Reis, & Levitt, 2006) and access to books and literacy related resources at home (Dickinson, McCabe, & Anastasopoulos, 2002). Growth slopes in reading were significantly lower for males and students from minority or low SES backgrounds, suggesting Matthew effects may be influenced by these factors. In contrast, students from high SES backgrounds and majority race/ethnicity backgrounds maintained their relative rankings but did not demonstrate fan spread effects. Based on the overlap in underlying skills discussed previously, it is possible child or classroom-level factors may influence writing growth; however this has not been examined. Additional research is needed to examine typical growth in microstructural measures of writing, explore differences in growth between students, and, if present, investigate potential predictors of fan spread effects in writing productivity.

Research Aims

The importance of monitoring language and literacy skill development is undisputed, and the value of writing skills for academic achievement is widely recognized. Despite this recognized importance and the emphasis on writing in the academic standards, there has been

less attention in the research on writing development compared to reading or oral language (Miller & McCardle, 2011). Additional research is warranted to add to our understanding of typical expected growth in written narratives in relation to standardized oral language measures and examine potential Matthew effects. Moreover, additional research is needed measuring change in writing performance to improve our understanding of the relationships between written narrative measures and standardized assessments across the school year. Perhaps most compelling, is the need to examine change in writing skills for fan spread growth and identify which students may be at increased risk for widening achievement gaps over time. In response, the current study was designed to address the research questions:

1) What is the average change in written narrative performance from beginning to the end of the school year for second grade students?
2) Are there differences in patterns of individual growth in written narrative outcomes for students in second grade? Specifically, is there evidence of Matthew effects in individual lexical growth patterns in writing?
3) What are potential predictors of the Matthew effects in students' writing measures (e.g., school, teacher, gender, free/reduced lunch, race/ethnicity, and language of the home)?

METHOD

Data for the current project were collected as part of a package of assessment measures administered in a larger grant funded by the Institute of Education Sciences, U. S. Department of Education. The study procedures were reviewed and approved by two universities' committees on research involving human subjects (HSC#:212777). The current project used extant data from one year of the funded project with the 13 participating elementary schools. The larger

project examined teachers' language but did not explicitly teach narratives or include structured or unstructured writing activities between fall and spring time points as part of the study.

Participants

The sample for this study included children from thirteen schools, in 33 classrooms in urban and rural areas of northern Florida and north central Tennessee. Twelve students were randomly selected for assessment from consented students in each class. The data set for the current paper included 299 participants in second grade from the larger study who had complete data for the written narrative sample at both fall and spring time points. The sample was comprised of 159 girls (52%) and 143 boys (47%) with an average age of 7 years and 6 months old (*SD* = 0.37 years). Of the 299 participating children, 58% were reported as White, 20% African American, 14% Hispanic, 5% Asian, and 2% mixed race. Additionally, a small percentage of the students (16%) were exposed to another language at home by one or more caregivers who spoke Spanish (*n* = 32), Arabic *(n* = 4), Mandarain (*n* =2) Amharic (*n* = 2), Korean (*n* = 1), and Dinka (*n* = 1). The percent of students receiving free or reduced lunch in the sample was 36%, with 108 receiving free or reduced lunch, two with missing data, and 189 students (63%) not eligible.

The investigators administered assessments of global language performance to describe students' general language skills, allow for considerations of generalizeability, and to further verify that students were considered to have typically developing oral language skills. Performance on the *Clinical Evaluation of Language Fundamentals- Fifth Edition* (CELF-5; Wiig, Secord, & Semel, 2013) was used to evaluate if the language skills were within normal range. Scores on language and literacy assessments were not used for inclusionary or exclusionary decisions but for descriptive purposes. Students demonstrated an overall mean

performance on the core language score of the *CELF-5* of 101.35 (*SD* = 15.87) which indicated they were within expected average range. No one was excluded based on his or her performance.

Materials for Assessment of Change across the School Year

Lexical Diversity and Productivity of Written Narratives. In the current study, number of different words (NDW) and total number of words (TNW) were calculated as a standard output measure of vocabulary using Systematic Analysis of Language Transcripts (Miller, 2011). NDW and TNW were considered to be advantageous for measuring Matthew effects because they were not constrained by ceiling effects. Additionally NDW and TNW are among the most frequently used metrics of vocabulary in writing samples (Danzak, 2011; Hall-Mills & Apel, 2013; Koutsoftas & Gray, 2013; Price & Jackson, 2015). Lexical productivity and diversity or the "range of vocabulary in a language sample" (Malvern, Richards, Chipere, & Durán, 2004, p.16) are commonly used as indicators of compositional productivity in writing (e.g., Abbott & Berninger, 1993; Berman & Verhoeven, 2002; Mackie & Dockrell, 2004; Puranik, Lombardino, Altman, 2008; Scott & Windsor, 2000; Wagner et al., 2011).

Accuracy of Written Narratives. Two measures of writing accuracy were included as additional language measures in the current study. The specific types of errors were not the focus of the current study. As such, a broad measure of accuracy based on the proportion of spelling and grammar errors was utilized. Investigators calculated proportion of errors within two broad categories of error types, including errors of spelling and errors of grammar. Calculation of errors as a ratio of errors to total number of words was consistent with measures established in previous studies (Eisenberg & Guo, 2013; Scott & Windsor, 2000).

Standardized Measures of Language

Given that children draw on language abilities when composing written narratives (Kim, 2016; Kim & Schatschneider, 2016; Kim et al., 2015), we have included performance on assessments of general language performance in the fall and spring. Three measures were utilized to describe participants' language skills and reported in Table 1. These include standardized measures of expressive vocabulary, receptive vocabulary knowledge, and formulating oral sentences. Each measure is described below.

Expressive Vocabulary. We assessed children's oral expressive vocabulary using the *Expressive Vocabulary Test*-Second Edition (EVT-2; Williams, 2007) in September using Form A and again in April using Form B. The EVT-2 is an individually administered, norm-referenced test assessing expressive vocabulary and word retrieval. The EVT-2 includes practice/example items and 190 test items arranged in increasing difficulty. For each item, the examiner presents a picture and reads a question intended to elicit a single word response. Correct responses require the child to label (e.g., what shape is this?) or to provide a synonym for a word appropriate to the image. Items include different parts of speech, home and school vocabulary, and different levels of specificity (e.g., tier 2 and tier 3 words). The test is untimed and takes approximately 15 minutes to establish the basal and ceiling (5 consecutive incorrect responses). According to the manual, the test-retest reliabilities yielded correlations between .94 and .97. Internal consistency for each form is reported with high split-half reliability of .94 for Form A and .93 for Form B. The reliability between Form A and Form B is reported to be .83-.91.

Receptive vocabulary. We assessed receptive English vocabulary based on children's recognition of spoken words on the *Peabody Picture Vocabulary Test-IV* (PPVT- IV; Dunn & Dunn, 2007). The test provides an array of four color pictures for each vocabulary item. The examiner asks the child to point to the picture that matches the spoken word from a four-picture

array. The child's response is scored dichotomously, as correct or incorrect. The items are arranged in sets of 10 items that are intended to become increasingly difficult. A basal is established (a set containing one or no errors) and the child continues until the ceiling of eight or more errors in a set is reached. The PPVT- IV is an untimed test normed through a sample of 3,540 participants for use with individuals 2 to 90 years old. Split half reliability by age for Form A and Form B was M = .94 (*SD* = 3.6), and range from .90-.97 for ages 5-11.

Formulation of Spoken Sentences. To further describe children's language skills in relation to their growth in written language, we administered the Formulating Sentences subtest of the Clinical Evaluation of Language Fundamentals- Fifth Edition (CELF-5; Wiig, Secord, & Semel, 2013) in the fall and spring of the school year. In the Formulating Sentences task, the examiner presents a word and asks the child to construct a grammatical sentence using the word. The subtest includes words that result in syntactically complex sentences.

Procedures

Standardized Oral Language Measures. The standardized measures of language were individually administered in random order across two testing sessions of 30-45 minutes each. All examiners had completed training on the administration procedure and met a researcher made proficiency criterion to ensure fidelity of implementation using the standardized test administration procedures and protocols. Completed test protocols were double scored to ensure accuracy in calculation of scores following rules for basals and ceilings.

Writing Task. Investigators administered the writing task as a large group (whole classroom) during the first seven weeks of instruction and the last seven weeks of the school year. For the written personal narrative task, the same prompt was administered in both the fall and spring time points. The typed prompt: *One day when I got home from school…*was provided

at the top of a double-sided lined piece of paper, which was similar to written prompts used in other studies (Connelly, Dockrell, Walter, & Critten, 2012; Dockrell, Ricketts, Charman, & Lindsay, 2014; McMaster & Espin, 2007; author et al., 2017). A research assistant read the instructions aloud to the class, informing students that they had 10 minutes to write a response. Specifically, the directions stated,

> *Do your best writing and please write neatly so we can read it later. Now you are going to write a story. I am going to read a sentence to you first, and then I want you to write a story about what happens. You will have 10 minutes to write your story. Do your best work. If you don't know how to spell a word, you should do your best and keep writing. You are going to write a story that begins, "One day when I got home from school…"*

Transcription. Research assistants trained in the university's speech-language pathology program transcribed the written samples following traditional procedures in accordance with conventions established for *Systematic Analysis of Language Transcripts* (*SALT*) (Miller & Iglesias, 2010). Lexical items in the students' writing samples included recognizable real words regardless of spelling errors. In the case that a written word was deemed illegible, it was not included in the analysis for number of different words but instead coded as illegible and not included in the number of different words. NDW and NTW were derived by generating standard measures reports in *SALT*.

Coding. Children's use of capitalization, misspellings, and punctuation was maintained. Research assistants inserted codes proximal to each deviation from Standard English that represented spelling or grammar errors. Examples of commonly occurring grammatical errors include the omission of past tense, omission of conjunctions and possessive markers, lack of verb-tense agreement, lack of singular and plural subject-verb agreement markers. Spelling

errors included any deviation from Standard English spelling rules. Coding rules were established to operationalize the assignment of error codes. For example, if the possessive marker was present but the apostrophe was missing (e.g., principals office), it was considered a punctuation error and not marked as a grammar error. In contrast, if the possessive marker was not present (e.g., principal office) is was counted as a grammatical error. However, one lexical item was allowed to have more than one error attributed. For example, if *cousin* was mis-spelled and lacking a possessive marker (e.g., we played my cuzin game), the spelling error and grammar error codes were both assigned.

Research assistants trained on identifying errors entered the error codes as described above. The SALT software was also utilized to aggregate the occurrence of each type of error code. The first author reviewed one of every ten transcripts to double score the error codes. It was expected that coding errors would occur at a rate of between 3%-6% during sample analysis based on what is commonly reported in the literature (e.g., Fey et al., 2004; Gillam & Johnston, 1992; Windsor, Scott & Street, 2000). Agreement was 97% for spelling and grammar errors, not including formatting errors. Any disagreements in error assignments were discussed to resolve errors as coding continued. Due to the high rate of agreement, further double scoring of spelling and grammar was determined to be unnecessary.

Analyses

For the first research question, descriptive statistics were examined for microstructural measures of writing at fall and spring time points. We used a mixed model to examine for significant effects of time using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). For the second research question we examined individual patterns of growth to test for Matthew effects. Given only two time points we were limited in the types of analyses that could

be used. More sophisticated methods such as quasi-simplex modeling or latent growth curves could not be utilized due to the availability of two time points only. However, with only two-points there are a number of analyses that could be used to examine the potential differences of Matthew effects. Bast and Reitsma (1997) discussed the properties that should be present if a Matthew Effect were operating. First, variance over time should be increasing along with the presence of a strong correlations over time. Secondly, we argue that additional evidence for a Matthew effect would be obtained if an initial score on an assessment was positively related to the change from initial score to final score. That is, we operationally defined a Matthew effect as a significant and positive relationships of initial status with change. With these criterion in mind, we tested for Matthew effects by examining the fall to spring correlations as well as changes in variance in lexical productivity from fall to spring on the writing samples of the students in second grade. Additionally, we fit a model that predicted the change score from initial status while taking into account the nested structure of the data. Finally, for our third research question, we examined potential predictors of Matthew effects (initial writing productivity, school, teacher, gender, free/reduced lunch eligibility and language of home) using a two-level hierarchical linear model with students nested within teachers and schools.

RESULTS

Descriptive Statistics

Our first research question was to examine the average change in written narrative performance from beginning to the end of the school year for second grade students. To address this question, we first report descriptive statistics at the beginning and end of the school year. Means and standard deviations for narrative measures of the full sample are provided in Table 1. For the total number of words, the students averaged 55 total words in the fall and averaged 69

words in the spring, with a statistically significant increase of 14 words t (298.0) = 8.32, p<.0001, d = 0.47. For number of different words, the students averaged 33 unique words in the fall and 41 unique words in the spring. This change was also significant t (298.0) = 9.17, p<.0001, d = 0.53. For proportion of spelling errors, there was a significant decrease from fall to spring time points t (280.6) = 7.38, p < .0001, d = 0.26. The ratio of spelling errors decreased from 15% to 12% on average, relative to the total number of words. There was no significant change in proportion of grammatical errors from fall to spring in the current study t (282.9) = 1.32, p = .190.

[insert Table 1]

Our second research question was to examine individual patterns of growth across the school year for students in second grade. To that end, we examined changes in variance in TNW and for Matthew effects. Because NDW and TNW were highly correlated, we did not examine NDW separately. For TNW, the variance from fall to spring increased from Var=791 to Var=1005, which was a significant increase (Grambsch Variance Test, Z= -2.2668, p = 0.0234). Additionally, the correlation of TNW from fall to spring was r=.53. Further, there was a significant positive correlation between initial (Fall) TNW and the change in TNW (r = .39). Lastly we tested for a relationship between pretest scores and the change between pretest to posttest performance. The correlation between Fall TNW and growth (change in TNW from fall to spring) was 0.39. Finally, we fit an HLM model where the TNW change score (posttest minus pretest) was the dependent variable and initial status on TNW was the predictor variable. Initially we also modeled the classroom and school level variance on the change score, but there was no variance at the school level, so that random effect was dropped. The final model demonstrated that there was a significant and positive relationship between initial level of TNW

and change (estimate = 11.8, $t(249.4) = 7.23$, $p < .0001$). The random effect of classroom was also significant (variance = 47.3, $X^2_{(1)} = 5.00$, $p = .025$). The student-within-classroom variance (residual) was estimated to be 679.5, with a classroom ICC of 47.3/679.5=7.0%. This implies that the Matthew effect varied by classroom. All together, these pieces of evidence suggest the existence of a Matthew effect for writing as measured by TNW. The distributions of TNW at fall and spring are displayed in the violin plot in Figure 1. A violin plot closely resembles a box plot, but the sides of the plot represent the density of the distribution.

[insert Figure 1]

Our third research question examined potential predictors of the Matthew Effects. We examined potential predictors of the amount of change in TNW nested within classroom. After taking into account initial TNW, we examined gender, race/ethnicity, eligibility for free or reduced lunch, and language of the home as potential moderators of growth rates. We operationally defined a moderator of a Matthew effect as the presence of a significant interaction of initial status and the potential moderator. As displayed in Table 2, free-reduced lunch eligibility was a marginally significant moderator of Matthew effects. Gender, race/ethnicity, language of the home were not significant predictors above and beyond initial performance level in written lexical productivity (TNW).

DISCUSSION

Key Findings

The primary purpose of this study was two-fold: a) to describe the written narrative performance of second grade students at the beginning and end of the school year and b) to test for Matthew effects in lexical productivity and examine predictors of change. Among key findings, students demonstrated significant increases in lexical diversity, productivity, and

proportion of accurately spelled words across the school year. Students demonstrated an increase of 14 total words on average. Additionally, the variance in lexical productivity (TNW) from fall to spring showed significant increases providing evidence for the existence of a Matthew effect for writing productivity as measured by total number of words. Further, the results suggest that Matthew effects are predicted by teacher or classroom-level factors and socioeconomic status as measured by eligibility for free or reduced lunch.

The finding that microstructural writing measures (lexical diversity, productivity, spelling accuracy) were sensitive to change across the school year is consistent with previous reports in the literature (Dockrell et al., 2015; Fey et al., 2004; Malecki & Jewell, 2003, Wagner et al., 2011) and substantiates the utility of short-duration writing samples for progress monitoring. The sensitivity of TNW in the current study was similar to the significant differences in TNW between fall and spring reported by Malecki and Jewell (2003) and growth across a 5-month period for total words as reported by Dockrell and colleagues (2015) for students in 3rd – 5th grade. Additionally, the significant change in lexical diversity across the school year substantiates expected gains in lexical diversity as reported by Fey and colleagues (2004) from 2nd to 4th grade. Further, the finding that spelling accuracy in writing was sensitive to developmental change in writing is consistent with previous studies (Dockrell et al., 2015).

In contrast, the lack of change in the rate of grammatical errors was surprising, but may have been influenced by the fact that errors in the current study were coded broadly without categorization of error types which may have decreased sensitivity to change in severity of error. Overall, findings substantiate that writing productivity (TNW), lexical diversity (NDW), and proportion of spelling errors are sensitive to developmental change in writing for students in second grade.

The unique contribution of the current results is the evidence for the existence of Matthew effects in writing productivity. In the current findings, students who had better written lexical productivity (TNW) at the beginning of the school year showed greater growth across the school year than students who entered second grade with poor written language productivity. This finding is consistent with the notion of Matthew effects described in other domains in which the rich get richer and the poor get poorer (Cook & Campbell, 1979; Merton, 1995; Rigney 2010). To our knowledge, few if any studies have examined Matthew effects in written language. Writing productivity may be particularly sensitive to such effects given that it is not constrained by ceiling effects and writing productivity measures tend to show the largest magnitude of change during early elementary grades, more so than lexical diversity (Wagner et al., 20011).

The finding that teacher/class and socioeconomic status were significant predictors of Matthew effects, appears to be aligned with results reported for Matthew effects in reading (Morgan et al., 2008) which substantiates that some children are more at-risk for Matthew effects than others. The fact that teachers or classroom-level factors explained 7% of the variance in growth highlights the malleability of early writing and the important influence of the environment on writing.

In light of the significant role of environmental factors in predicting writing growth, the current findings lend support for models that emphasize the interaction of environmental factors with individual characteristics and capacity, such as the Revised Writer(s)-Within-Community (WWC) Model of Writing (Graham, 2018). Although children's initial lexical productivity influenced spring performance, consistent with the Lexical Quality Hypothesis; the current findings suggest that teachers or classroom-level factors and socioeconomic resources in the child's environment also underpin children's writing growth. Such predictors seem well aligned

to the WWC Model of writing which proposes that writing is simultaneously shaped by the home and school communities in which writing takes place in addition to individual cognitive and linguistic capacity.

The presence of Matthew effects in written productivity underscores the importance of expanding writing instructional support for students in elementary school with low initial performance on writing. Further, the finding that teacher/classroom and socioeconomic level predicted Matthew effects suggests that certain groups of children may be more likely to require additional writing supports to prevent achievement gaps from widening. Although it is impossible to identify aspects of classrooms or teachers that influenced writing growth, the current findings affirm that classroom features matter for early writing growth. Previous work in literacy has noted the effects of quality of instruction, classroom resources, peer-to-peer support, and time spent in literacy on student literacy outcomes (Cunningham & Stanovich, 1997; Guthrie, Wigfield, Metsala, & Cox, 1999). By better understanding Matthew effects it is hoped that we can identify effective ways to prevent the effects or neutralize opportunity imbalances such as bolstering writing instruction and support to those students at risk.

Limitations

The length of time studied is notably narrow, in examining growth from fall to spring within the same school year. Given only two time points we were limited in the types of analyses that could be used. As such, preferred methods such as quasi-simplex modeling or latent growth curves could not be utilized. Although our primary research aim was to isolate and describe change within second grade, additional longitudinal studies of successive school years are needed to more fully describe development and add to the knowledge base on written language development across the elementary school years. Similarly, although we focused on component

skills in the current study, in future studies it would be interesting to examine students' average change in measures that assess more depth in vocabulary, rather than just breadth, and measures that reflect more quality ratings or holistic aspects with additional focus on development of ideas, cohesiveness, and organization. Although not available in the current study, it would be interesting to examine macrostructural measures such as the six- trait writing rubric (STWR, Education Northwest, 2006) or similar Likert type scales of quality (Dockrell, Ricketts, Charman & Lindsay, 2014; Koustsoftas, 2016; Wechsler, 2005).

The use of only one sample at each time point was a noted limitation. Because the prompt was the same at each time point, students may demonstrate change simply because they have thought about the topic previously, and not because their underlying written language skills have improved. Although this is a limitation, it is often common practice to administer the same prompt to keep the measure similar across two points (Abbot et al., 2010; Dockrell et al., 2015; Juel, 1988; authors, 2018). Additionally, given a single writing task at each time point, it cannot be assumed that the sample is representative to their typical work or that similar results would be observed using other types of prompts (e.g., persuasive or explanatory). Other authors (e.g., Olinghouse & Leaird, 2009) have reported vocabulary diversity has been shown to remain stable across two different writing tasks, but other measures from children's writing samples vary across writing tasks. In a future study it would be interesting to examine differences in other types of samples (e.g., persuasive and explanatory) and their use as progress monitoring tools for young school age students.

Another limitation for consideration is the lack of available assessment information on other skills that could be related. For example, in the current study we did not have access to information about the students' verbal reasoning skills and working memory which have been

shown to relate to reading and writing achievement in previous studies (Berninger & Richards, 2010; Prifitera, Weiss, Saklofske, & Rolfhus, 2005). Given that prior evidence supports a relationship between verbal reasoning and reading and writing achievement (e.g., Prifitera et al., 2005) and relationship between verbal working memory and achievement (e.g., Berninger & Richards, 2010), we cannot rule out that differences in average change over the school year are in part related to students' verbal reasoning and verbal working memory. In future studies, it would be interesting to consider other measures of language performance to further explore factors that predict the developmental trajectory of writing skills. Although exploring verbal reasoning and memory skills was not an aim of the current study, it would be beneficial for future studies to include such factors as potential moderators of writing performance and growth.

Implications and Suggestions for Future Studies

Despite limitations of the study, the findings substantiate the sensitivity of written personal narrative measures to developmental changes in children's performance across the school year. The developmental changes support the usefulness of written narratives for progress monitoring the language development of young school age children. The current findings also highlight specific components that appear to be malleable across the school year. The identification of components that are expected to show growth across the school year may be particularly useful for teachers and related personnel in progress monitoring and program planning. Further, the important role of initial vocabulary skills on narrative performance adds to our understanding of children's outcomes and predicted performance across the school year.

The presence of Matthew effects in the writing skills of second grade students spurs intellectual curiosities that warrant further study. Additional studies are needed to understand why and under what circumstances Matthew effects occur and do not occur in writing skills. The

source and nature of inequalities require more scientific inquiry. Additional study of the underlying mechanisms or factors sustaining inequalities is also needed (e.g., motivation, writing experience, access to print) in order to identify ways to prevent gaps in writing achievement. In future studies it would be interesting to explore other potential moderators of the effect, such as quantity and quality of language experiences and exposures, duration of writing instruction, and/or writing instructional strategies. The presence of gaps in writing achievement warrants exploration of innovative ways to minimize achievement gaps and expand writing instructional supports for students with low initial performance in writing.

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of Educational Psychology*, *85*(3), 478-508. http://dx.doi.org/10.1037/0022-0663.85.3.478

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

briefs/FPG_CECER-DLL_AssessingDLLs.pdf

Aram, D., & Levin, I. (2001). Mother–child joint writing in low SES: Sociocultural factors, maternal mediation, and emergent literacy. *Cognitive Development*, *16*(3), 831-852.

Bahr, R. H., E. R. Silliman, V. W. Berninger & M. Dow (2012). Linguistic Pattern Analysis of Misspellings of Typically Developing Writers in Grades 1-9. *Journal of Speech Language and Hearing Research, 55*, 1587-1599.

Bast, J., & Reitsma, P. (1997). Mathew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, *32*(2), 135-167.

Bates, D., Maechler, M., Bolker, M., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Berman, R. A., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy Written*

Language & Literacy Cross-Linguistic Perspectives on the Development of Text-Production Abilities in Speech and Writing. Part 1, 5(1), 1-43. doi:10.1075/wll.5.1.02ber

Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of Psycholinguistic Research*, *39*(2), 141-163.

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345-363). New York, NY, US: Guilford Press.

Berninger, V. & Richards, T. (2010). Inter-relationships among behavioral markers, genes, brain, and treatment in dyslexia and dysgraphia. *Future Neurology, 5,* 597-617.

Broc, L., Bernicot, J., Olive, T., Favart, M., Reilly, J, Quémart, P., & Uzé, J. (2013). Lexical spelling in children and adolescents with specific language impairment: Variations with the writing situation. *Research in Developmental Disabilities, 34*, 3253-3266.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the Quality of Composition and Written Language Bursts From Oral Language, Spelling, and Handwriting Skills in Children With and Without Specific Language Impairment. *Written Communication, 29*(3), 278-302. doi:10.1177/0741088312451109

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago, IL: Rand McNally College.

Cunningham, A. E. & Stonovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934-945.

Danzak, R. L. (2011). The integration of lexical, syntactic, and discourse features in bilingual adolescents' writing: An exploratory approach. *Language, Speech, and Hearing Services in Schools, 42,* 491-505. doi:10.1044/0161-1461(2011/10-0063)

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48,* 368-371.

Devonshire, V., & Fluck, M. (2010). Spelling development: Fine tuning strategy-use and capitalising on the connections between words. *Learning and Instruction, 20,* 361–371. https://doi.org/10.1016/j.learninstruc.2009.02.025

Dickinson, D. K., McGabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95,* 465-481. doi: 10.1037/0022 0663.95.3.465

Dockrell, J. E., Connelly, Walter, Critten (2015). Assessing children's writing productss: the role of curriculum based measures. *British Educational Research Journal, 41*(4), 575-595. doi: 10.1002/berj.3162

Dockrell, J. E., Ricketts, J., Charman, T., & Lindsay, G. (2014). Exploring writing products in students with language impairments and autism spectrum disorders. *Learning and Instruction, 32*, 81-90. doi:10.1016/j.learninstruc.2014.01.008

Dunn, L. M., & Dunn, D. M. (2008). *The peabody picture vocabulary test* (4th ed.). Circle Pines,

MN: American Guidance Service.

Eisenberg, S.L., & Guo, L. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools, 44*, 20-31.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research, 47*(6), 1301-1318.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, *35*, 39–50. doi:10.1207/S15326985EP3501_5

Gansle, K. A., G. H. Noell, A. M. VanDerHeyden, G. M. Naquin & N. J. Slider (2002) Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*, 477-497.

Gillam, R., & Johnston, J. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35,* 1303-1315.

Graham, S. (2018). A Revised Writer(s)-Within-Community Model of Writing, *Educational Psychologist, 53*(4), 258-279, doi: 10.1080/00461520.2018.1481406

Graham, S., Berninger, V.W., Abbott, R.D., Abbott, S.P. & Whitaker, D. (1997). Role of mechanics in composing of elementary school students. A new methodological

approach, *Journal of Educational Psychology, 9*(1), 170-182. doi:10.1037/0022-0663.89.1.170.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, *9*(2), 123-145.

Gutierrez-Clellen, V.F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13,* 175-197.

Hall-Mills, S., & Apel, K. (2013). Narrative and expository writing of adolescents with language learning disabilities: A pilot study. *Communication Disorders Quarterly, 34,* 135-143. doi:10.1177/1525740112465001.

Harrison, G., Goegan, L., Jalbert, R., McManus, K., Sinclair, K., Spurling, J. Predictors of spelling and writing skills in first- and second-language learners. *Reading and Writing: An Interdisciplinary Journal, 29*, 69-89. doi: 10.1007/s11145-015-9580-1

Jenkins, J., Johnson, E., & Hileman, J. (2004). When is reading also writing: Sources of individual differences on the new reading performance assessments. *Scientific Studies of Reading, 8*, 125-151.

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. doi:10.1037/0022-0663.80.4.437

Justice, L.M., Bowles, R.P., Kaderavek, J.N., Ukrainetz, T.A., Eisenberg, S.L., & Gillam, R.B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age

children's narrative performances. *American Journal of Speech-Language Pathology, 15,* 177-191.

Kim, Y.S.G., & Park, S.H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing: An Interdisciplinary Journal, 32*, 1319-1341. https://doi.org/10.1007/s11145-018-9913-y

Kim, Y. S. G., Park, C., & Park, Y. (2015). Dimensions of discourse level oral language skills and their relation to reading comprehension and written composition: An exploratory study. *Reading and Writing: An Interdisciplinary Journal*, *28*(5), 633-654.

Kim, Y., & Schatschneider, C. (2016). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology, 109*(1) 35-50. doi.org/10.1037/edu0000129.

Koutsoftas, A. D., & Gray, S. (2013). A structural equation model of the writing process in typically-developing sixth grade children. *Reading and Writing: An Interdisciplinary Journal, 26,* 941-966. doi:10.1007/s11145-012-9399-y.

Lerkkanen, M., Rasku-Puttonen, H., Aumola, K., & Nurmi, J. (2004). The developmental dynamics of literacy skills during the first grade. *Educational Psychology, 24,* 793-810.

Lonigan, C. J., & Shanahan, T. (2008). Executive Summary Developing Early Literacy: Report of the National Early Literacy Panel A Scientific Synthesis of Early Literacy Development and Implications for Intervention. Executive Summary of the report of the national early literacy panel. Retrieved from: https://files.ericed.gov/fulltext/ED508381.pdf

Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research*, *47*, 1469-1483.

Malecki, C. K. & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40*(4). 379-390.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Hapshire, NY: Palgrave Macmillon.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98,* 14-28.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education, 41*(2), 68-84. doi:10.1177/00224669070410020301

Merton, R. (1995). The Thomas theorem and the Matthew effect, *Social Forces, 74*(2), 379-424.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2015). *Assessing language production using SALT software: A clinician's guide to language sample analysis.* Middleton, WI: SALT Software LLC.

Miller, J. F., & Iglesias, A. (2010). Systematic analysis of language transcripts (SALT), English & Spanish (Version XX) [Computer software]. Madison: University of Wisconsin – Madison, Waisman Center, Language Analysis Laboratory.

Miller, B., & McCardle, P. (2011). Reflections on the need for continued research on

writing. *Reading and Writing: An Interdisciplinary Journal*, *24*, 121–132. doi:10.1007/s11145-010-9267-6

Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly, 31,* 187-198.

National Center for Education Statistics. (2012). Institute of Education Sciences, U.S. Department of Education. *The nation's report card: Writing 2011* (NCES 2012-470). Retrieved from: https://www.nationsreportcard.gov/writing_2011/g12_national.aspx

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading & Writing: An Interdisciplinary Journal, 22*(5), 545-565. doi:10.1007/s11145-008-9124-z

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67-86.

Pfost, M., Hattie, J. Dorfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203-244.

Price, J. R. & Jackson, S. C. (2015). Procedures for obtaining and analyzing writing samples of school-age children and adolescents. *Language, Speech, and Hearing Services in Schools, 46,* 277-293.

Prifitera, A., Weiss, L., Saklofske, D., & Rolfhus, E. (2005). The WISC-IV in clinical assessment context. In A. Prifiteria, D. Saklofske, & L. Wiss (Eds.), *WISC IV clinical use*

and interpretation. Scientist-practitioner perspectives (pp.3-32). New York, NY: Elsevier.

Puranik, C. S., Lombardino, L. J., Altman, L. J. P. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*, *17*, 107-120.

Quick, N. & Erickson, K. (2018). A multilingisic approach to evaluating student spelling in writing samples. *Language, Speech, and Hearing Services in Schools, 49*(3), 509-523. doi.org/10.1044/2018_LSHSS-17_0095.

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage.* New York, NY: Columbia University Press

Shanahan, T. (2006). Relations among oral language, reading, and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 83–95). New York, NY: Guilford Press.

Scott, C.M. & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities, *Journal of Speech Language and Hearing Research, 43*(2), 324-449.

Singer, B. D., & Bashir, A. S. (2004). Developmental variations in writing composition skills. In C. Addison Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders*, (pp. 559-582). New York, NY: Guildford 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: An Interdisciplinary Journal, 24*, 203-220. doi: 0.1007/s11145-010-9266-7

Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American Educational Research Journal, 20*, 359-373. doi: 10.3102/00028120200003359.

Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Bloomington, MN: Pearson.

Williams, K. T. (2007). *EVT-2: Expressive vocabulary test, Second Edition*. Mineapolis, MN: Pearson Assessments.

Windsor, J., Scott, C. M., & Street, C. K. (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech Language and Hearing Research, 43*(6), 1322-1336. doi:10.1044/jslhr.4306.1322

Wood, C. L., Bustamante, K. N., Schatschneider, C., & Hart, S. (2018). Relationship between children's lexical diversity in written narratives and performance on a standardized reading vocabulary measure. *Assessment for Effective Intervention*. *44*(3), 173-183. doi:10.1177/1534508417749872

Wood, C. Schatschneider, C. & Hart, S. (2017). Grade level expectations in lexical measures and accuracy of written narrative samples. *Journal of Child Language Acquisition and Development, 5*(2), 127-144. doi: 2148-1997

Yeung, P.S., Ho, C., Chan, D., Chung, K. (2013). Modeling the relationships between cognitive linguistic skills and writing in Chinese among elementary grade students. *Reading and Writing: An Interdisciplinary Journal, 26*(7), 1195-1221. doi: 10.1007/s11145-012-94116

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*, 236-259. doi: 10.1093/applin/amp02

Table 1

Fall and Spring Performance on Writing Measures and Descriptive Standardized Assessments of Language

	Fall			Spring		
	N	*M*	*SD*	*N*	*M*	*SD*
Writing Measures						
Number of Different Words	299	33.35	14.65	299	41.37	15.56
Number of Total Words	299	55.75	27.95	299	69.70	31.73
Errors of Spelling	299	0.15	0.12	299	0.12	0.11
Errors of Grammar	298	0.03	0.04	298	0.03	0.04
Standardized Assessments						
PPVT Raw Score	299	127.60	18.57	299	138.41	17.94
PPVT Standard Score	299	104.39	13.98	299	105.19	13.59
EVT Raw Score	299	94.34	16.10	299	102.22	14.98
EVT Standard Score	299	100.71	13.69	299	101.55	12.93
CELF-FS Raw Score	299	29.28	8.36	299	31.16	8.20

Note. FRL refers to eligibility for free or reduced lunch. TNW refers to total number words. *Errors of Spelling* refers to the proportion of spelling errors to total words. *Errors of Grammar* refers to the proportion of grammatical errors to total words. *PPVT* refers to the *Peabody Picture Vocabulary Test-IV* (Dunn & Dunn, 2007). *EVT* refers to the *Expressive Vocabulary Test-2* (Williams, 2007). *CELF FS* refers to the formulated sentences subtest of the *Clinical Evaluation of Language Fundamentals-5th Edition* (Wiig, Secord & Semel, 2013).

Table 2

Parameter Estimates Examining Moderators of Matthew Effects

Model	Effect	Estimate	t-value/Chi Squared	df	p-value
Gender	Initial (Fall TNW)	8.75	3.42	274.4	0.001
	Gender	-14.50	-2.08	294.3	0.038
	Initial*Gender	0.194	1.71	293.1	0.089
	Classroom Variance	42.70	4.35	1	0.036
	Residual	676.10			
Race	Initial (Fall TNW)	9.23	1.79	266.9	0.075
	Race		-0.31	284.9	0.757
	Initial*Race		0.29	275.5	0.772
	Teacher Random Effect	44.33	4.06	1	0.044
	Residual	684.33			
Free/Reduced Lunch	Initial (Fall TNW)	9.58	4.94	237.97	<.0001
	FRL	0.85	0.24	171.06	0.812
	Initial*FRL	7.03	1.92	292.98	0.056
	Teacher Random Effect	36.55	2.99	1	0.084
	Residual	679.75			
Language of Home	Initial (Fall TNW)	13.17	7.79	238.81	<.0001
	Language of Home	-2.10	-0.41	202.82	0.679
	Initial*Lang. of Home	-9.87	-1.82	280.53	0.070
	Teacher Random Effect	41.45	4.04	1	0.044
	Residual	651.68			

Note. FRL refers to eligibility for free or reduced lunch. TNW refers to total number words. There is no estimate for race or the interaction of race and initial status because it is a multiparameter test.

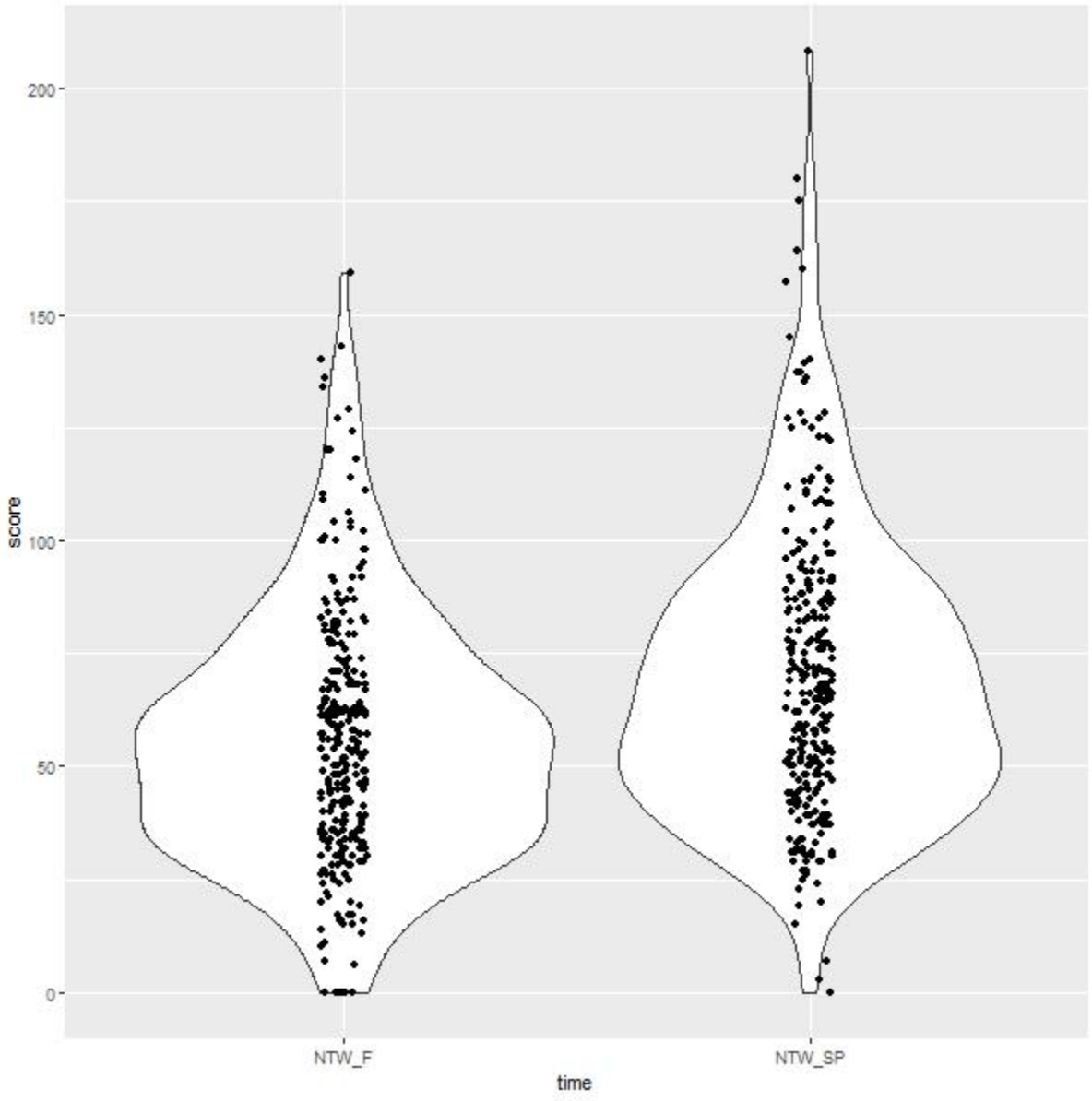

Figure 1.

Matthew Effects in Students' Written Productivity as Measured Total Number of Words.

Note. The figure on the left shows fall lexical productivity as compared to spring on the right.