Purpose: As children develop skills in writing across academic contexts, clinicians and educators need to have a fundamental understanding of typical writing development as well as valid and reliable assessment methods. The purpose of this study was to examine the progression of linguistic elements in school-age children’s narrative and expository writing development.

Method: Narrative and expository writing samples produced by 89 children in Grades 2 through 4 were analyzed at the microstructure and macrostructure levels. Measures of receptive vocabulary, word-level reading, and reading comprehension were obtained.

Results: Exploratory factor analyses revealed 4 microstructure factors (e.g., productivity, grammatical complexity, grammatical accuracy, and lexical density) and 1 macrostructure factor (e.g., a combination of organization, text structure, and cohesion). Multivariate analyses of covariance with reading comprehension as a covariate showed that productivity and macrostructure were sensitive to grade-level and genre differences and that expository grammatical complexity was sensitive to grade-level differences.

Conclusions: Findings are discussed in light of grade-level standards for narrative and expository writing and current practices in writing assessment. Multiple suggestions are offered for clinical and educational implications, and specific directions are provided for future research.

Speech-language pathologists (SLPs) working with school-age children and adolescents have important roles in prevention, identification, assessment, and intervention for problems involving oral and written language (American Speech-Language-Hearing Association [ASHA], 2010). The school context is heavily influenced by the established curricular standards, and SLPs contribute their knowledge of language development to support students who struggle with the prerequisite language skills to achieve those standards. A majority of states have adopted the Common Core State Standards (Common Core State Standards Initiative, 2010), which include many rigorous and integrated written language competencies across grade levels and content areas. Given their specialized knowledge base in the normal development of writing in the context of the general education curriculum (ASHA, 2002), SLPs can inform the assessment, instruction, and intervention of children’s written discourse skills. There is a need to document the nature of children’s written discourse skills across the elementary grades when children are learning to write. There is also a need for validated assessment methods to elicit and analyze children’s writing products. Therefore, the purpose of this study was to examine dimensions of written language produced by children in Grades 2 through 4 with typical language development within two common academic discourse genres (i.e., narrative, expository) using multiple levels of analysis (i.e., microstructure and macrostructure).

Academic Discourse Genres

Written discourse genres represent different forms and styles of writing and reflect a range of purposes and contexts for writing (e.g., Graham & Harris, 2003; Graham & Perin, 2007). In the school environment, narrative and expository genres are the most commonly encountered discourse genres in elementary grades (Donovan & Smolkin, 2006). A frequently employed assessment method to elicit narrative and expository writing from students in the elementary grades is with predetermined topic prompts given by the teacher or examiner. The majority of state writing assessment programs
utilize this approach to eliciting student writing. In the basic procedure, a set of writing prompts is provided to the student. Students may have a choice of prompts for narrative, expository, and persuasive discourse, depending on their grade level. In some state writing assessment programs, only one set of prompts is provided (i.e., no student choice). The students have an established amount of time to write and most often use paper and pencil to complete the task (with the exception of allowable accommodations for students with disabilities). The students’ written work is later evaluated based on criteria established by the state education agency, taking into account grade-level curricular standards for writing and multiple features that are thought to reflect writing proficiency. The prompt-based elicitation method also has been infused in writing instruction in the classroom. Therefore, writing prompts provided to elementary school-age students frequently are intended to elicit either narrative or expository texts.

Narrative discourse involves telling a story, often about personal events or other life experiences (e.g., novels, personal letters, and short stories). Expository discourse involves conveying facts or describing procedures, sharing basic information, relating cause–effect relationships, or arguing a point of view (e.g., essays, editorials). The ability to write proficiently in both narrative and expository genres is a language skill directly linked to academic success (e.g., Beers & Nagy, 2011; Nelson, Bahr, & Van Meter, 2004; Singer, 2007). By the fourth grade, expository discourse is the principal genre of instruction (Common Core State Standards Initiative, 2010; Nippold & Sun, 2010).

Knowledge of discourse genres is acquired in a developmental progression and is related to reading comprehension and writing achievement (e.g., Olinghouse & Graham, 2009; Shanahan, 2006). Awareness and use of narrative discourse in written language typically develops first, often through storytelling experiences (e.g., Nelson et al., 2004). Compared with narrative discourse, expository discourse structure typically is mastered later in the school years and, as a consequence, is more difficult to produce and comprehend for many students (Berman & Verhoeven, 2002). Much of the recent research regarding discourse genres in written language has centered on text comprehension; in contrast, fewer studies have focused on text production (i.e., writing). Furthermore, when researchers have examined linguistic features at the discourse level in written language, their investigations often have been restricted to narrative discourse. There is a need, then, to examine students’ writing skills across additional discourse genres, such as expository, especially considering that 60% of writing assignments are expository in nature by the fourth grade (Graham & Perin, 2007). In this study, we investigated students’ developing written language skills in both narrative and expository genres.

When examining students’ written language skills, researchers have analyzed their writing samples for microstructure (i.e., the text base where the writer’s conveyance of meaning is structured at the word, sentence, and discourse levels) and macrostructure (i.e., the general idea of the writer’s meaning at the discourse level reflected through cohesion, organization, and genre structure). Both microstructure and macrostructure are important characteristics to inspect because they reflect different levels of linguistic prowess in the writing task. There are important implications from previous investigations. Herein, we present a summary of the findings related to the development of writing skills in school-age children.

**Microstructure**

Limited research on the development of narrative microstructure has revealed developmental changes in children’s and adolescents’ use of literate words and complex syntax in narrative discourse across age groups (Sun & Nippold, 2012). Investigations of the development of expository microstructure have shown that commonly employed measures such as mean length of T-unit (MLTU), type–token ration (TTR), and Syntactic Density Score (T-unit and clause length, number of subordinate clauses, embeddings, and verb expansions) were sensitive to differences in grammatical complexity and lexical diversity between successive age levels, beginning with children ages 9 years and older (Morris & Crump, 1982). Furthermore, developmental changes in writing productivity (e.g., total words, total T-units) and syntax (e.g., MLTU) have been detected between Grades 5 and 8; students in Grade 8 were more likely to produce a greater number of subordinate clauses and to embed subordinate clauses within other subordinate clauses in their sentences (Nippold & Sun, 2010).

In a comprehensive analysis of expository microstructure, Puranik, Lombardino, and Altmann (2008) examined the development of expository writing of 120 children in the third through sixth grade, targeting 13 variables of microstructure at the word, T-unit, sentence, and discourse levels. Measures of productivity and grammatical complexity increased with age. Significant differences were evident between the third and fourth grade groups for total words, total ideas, number of T-units, number of clauses, number of sentences, sentence complexity, and number of different words (NDW). In addition, factor analysis confirmed that the 13 microstructure variables examined clustered into four dimensions of written language microstructure: productivity, complexity, accuracy, and mechanics.

Investigators have examined microstructure performance across more than one discourse genre. These studies have revealed genre effects in favor of the narrative genre across productivity and writing fluency measures when compared with the expository genre (Scott & Windsor, 2000), including a trend for children with a mean age of 11.5 (years;months) to include more clauses per T-unit in narrative products but more words per T-unit in expository products. In a comparison of microstructure features across four genres (narrative, descriptive, compare/contrast, persuasive), Beers and Nagy (2011) noted that children in the third, fifth, and seventh grades produced more subordinate clauses in persuasive products than in descriptive products, and more words per clause in descriptive products than in...
persuasive products. Compare/contrast texts were shorter (in number of words) than persuasive texts at each grade level, and narratives were shorter than persuasive texts at Grades 5 and 7. Beers and Nagy’s (2011) study documented an important interaction between syntactic complexity and genre, indicating that syntactic- and text-level requirements vary by the genre structure.

In a cross-linguistic study of seven languages, including English, comparing four age levels (Grades 4, 7, 11, and adult), two genres (narrative and expository), and two modalities (spoken and written), Berman and Verhoeven (2002) examined multiple aspects of the development of narrative and expository macrostructure. Measures included lexical diversity, productivity, and syntactic complexity. Like Morris and Crump (1982), Berman and Verhoeven found that measures of microstructure were sensitive to developmental change across elementary, middle, and high school age levels. Furthermore, when considering the potential influence of genre in the development of microstructure, Berman and Verhoeven noted little to no effect of genre on fourth grade narrative and expository lexical diversity, a result that was similar to Scott and Windsor’s (2000) findings for NDW.

In summary, analysis of elements of microstructure in a written product can occur at the word, sentence, and discourse levels. Microstructure analysis generally includes measures of productivity (e.g., number of words, T-units, or ideas), grammatical complexity (e.g., MLTU, clausal density [CD]), and lexical diversity (e.g., TTR, NDW; Nelson et al., 2004; Puranik, Lombardino, & Altmann, 2007, 2008). Taken together, the results of previous investigations suggest that measures of productivity, grammatical complexity, and lexical diversity are sensitive to age and grade-level differences. However, less is known specifically about children’s development of certain microstructure elements (e.g., productivity, grammatical complexity, lexical diversity) in early grades (second through fourth), particularly across genres (e.g., narrative, expository).

**Macrostructure**

In contrast to microstructure analysis, macrostructure analysis occurs mainly at the discourse level (Scott, 2009). Macrostructure is the “abstract representation of the global meaning structure” which represents the “gist” of the text (Sanders & Schilperoord, 2006, p. 387). Macrostructure analysis examines a writer’s expression of meaning at the discourse level and may include measures of organization, cohesion, and genre-specific text structure. Elements of macrostructure often are included in qualitative writing analyses, such as in holistic or analytic scoring systems, or can be depicted quantitatively by counting cohesive ties or genre-specific text structure elements present in a written product (e.g., counting story grammar elements in a narrative text, or marking whether an introduction, body, and conclusion are present in an expository text).

Researchers have examined the development of elements of macrostructure in products written in either a narrative or an expository genre. These studies revealed the following:

- Approximately half of third and fourth grade students exhibiting typical development produced complete stories on the basis of story grammar analysis (Laughton & Morris, 1989).
- There were no developmental differences between scoring procedures for narrative macrostructure when comparing a procedure for parsing and categorizing propositions versus a holistic rating of cohesion, organization, and episodic structure of the story on a 5-point Likert scale (Montague, Maddux, & Dereshiwsky, 1990).
- Fourth and fifth grade students were better able to organize compare/contrast expository compositions than explanation products and included a higher total number of ideas in the explanation genre (Englert, Raphael, Anderson, Gregg, & Anthony, 1989).

Due to the emphasis of curricular standards on learning to write in a variety of academic discourse genres, there is value in examining how macrostructure develops in the writing produced by the same children across more than one genre. When investigators have examined the development of macrostructure features across more than one discourse genre, they have found the following:

- Relations among reading performance and cohesion in writing produced by students in the third and fifth grades, grade effects for use of cohesive ties, and genre effects in favor of the narrative genre for developmental changes in cohesive harmony (i.e., reflection of functional relations through noun and verb chains; Cox, Shanahan, & Sulzby, 1990)
- Possible age effects between students in the fifth and eighth grades were detected by using a trait scoring system for ideas, organization, sentence fluency, and conventions for written narrative, imaginative, persuasive, and expository products (Crawford, Helwig, & Tindal, 2004)

Although it is important to consider students’ performance in a variety of genres, looking solely at macrostructure development across genres ignores the role of microstructure features. In many instances, investigators have sought to document developmental trends across both microstructure and macrostructure variables within a single genre (e.g., Gillam & Johnston, 1992; Mackie & Dockrell, 2004; Nelson & Van Meter, 2007; Nodine, Barenbaum, & Newcomer, 1985). The results revealed differences between grade-level groups (e.g., Grades 1 and 2 vs. Grades 4 and 5) and nonadjacent grade-level groups (e.g., Grades 4, 8, and 11) and suggested a relation among features of microstructure (e.g., productivity, grammatical complexity, lexical diversity) and macrostructure (e.g., story grammar, coherence, cohesion). To date, only one investigation has been conducted to examine the development of linguistic features by using a combination...
of microstructure and macrostructure measures across two or more discourse genres. Koutsoftas and Gray (2012) measured microstructural elements (e.g., productivity, lexical diversity, grammatical complexity, spelling accuracy) by using an analytic scale within narrative and expository writing samples of fourth and fifth grade students (30 students exhibiting typical language development, 26 students with language-learning disabilities). The researchers also examined story grammar elements in the narrative samples (a measure of macrostructure). Significant differences were found between the students with typical language versus those with language-learning disabilities on measurement type (e.g., analytic vs. holistic scales) as well as genre (e.g., narrative vs. expository). Students with language-learning disabilities received lower holistic scores in both genres and lower scores on all six analytic measures used for narrative and three of five analytic measures used for expository writing. Results also reflected a significant relationship between analytic scores for productivity, sentence complexity, and lexical diversity with overall holistic scores. It is important to note that the analytic measures for expository writing did not measure expository macrostructure specifically. Therefore, macrostructure differences could not be examined across genres, an important level of analysis.

In our attempts to understand what is known about children’s development of microstructure and macrostructure features in a variety of academic discourse genres, we must consider some important limitations in the current body of literature: examination of a limited range of variables, inclusion of small sample sizes, analysis on only one level of the written product (microstructure vs. macrostructure), focus on one discourse genre, and examination of linguistic features in students in upper grades only (fourth grade and above). Current curricular standards reflect the expectations for students in Grades 2 through 4 to develop a range of foundational skills that will lead to more advanced academic discourse skills in the later grades (Common Core State Standards Initiative, 2010). Given the heightened focus on academic discourse skills in the earlier grades, more information is needed about how children develop various linguistic features through their narrative and expository writing during these years of school. In addition, the literature suggests that an association exists between the development of microstructure and macrostructure. However, few investigators have explored this relation systematically across multiple genres.

Thus, the purpose of this study was to document the progression of linguistic elements of microstructure and macrostructure that students in second, third, and fourth grade use in their written narrative and expository compositions. The following research questions guided the present study:

1. Are there differences among grades and between genres in written linguistic microstructure elements?
2. Are there differences among grades and between genres in written macrostructure elements?
3. To what degree is development of microstructure elements related to development of macrostructure elements?

We hypothesized that there would be grade-level differences for narrative and expository productivity and grammatical complexity. Performance in the narrative genre was anticipated to be superior to expository performance, especially for students in the second and third grades who generally possess less knowledge of and experience with the expository genre (Duke, 2000). A grade effect was anticipated for narrative and expository macrostructure on the basis of variables of organization, coherence, and text structure. Levels of macrostructure were expected to be superior for the fourth grade students. The potential genre effects were more difficult to anticipate given the range of findings in previous investigations. However, we anticipated that performance on macrostructure measures would be similar in both genres for the oldest students (fourth grade), if the assumption held true that experience and knowledge of various text structures and text cohesion increase with age. Few researchers have directly examined the potential relationship between microstructure and macrostructure. As such, we expected that development of these elements would be related in the sense that increased productivity (more words per written product) allows more opportunities for a writer to incorporate the necessary text structure elements and genre-specific organizational structure.

Method

Participants

Participants were recruited from a public elementary school located in a mid-sized city, with a student body that was representative of the state through the school’s use of stratified sampling. Participants were recruited in conjunction with a larger investigation examining an experimental spelling intervention. Approval was obtained from the University Institutional Review Board and the school for the procedures and consent forms for this study. Consent forms were sent home to all second, third, and fourth grade students. Participants had to be monolingual English-speaking, enrolled in general education, with no history of sensory impairments as determined by school records. Consultation between the first author and school administrators confirmed whether participants with parental consent met the inclusionary criteria. School records confirmed participants’ language status on the basis of the state-required Home Language Survey (Florida Department of Education, 1990) and related language proficiency testing to identify students who were bilingual or who were determined to be English language learners.

A total of 93 participants enrolled in general education were recruited, 89 of whom completed the writing samples (e.g., four left due to transferring out of the school). The final sample included 37 boys (41.6%) and 52 girls (58.4%): 28 participants were in the second grade, 28 participants were in the fourth grade, and 33 participants were
in the third grade. Participants ranged in age from 7;0 to 10;11 (M = 8;6, SD = 10;9) and represented a range of ethnic backgrounds, including 55% White, 20.2% African American, 11.2% Hispanic, 3.4% Asian American, 7.9% multiethnic, and 2.2% unreported ethnic backgrounds. The participants had average receptive vocabulary skills (M = 101.10, SD = 14.14), as a proxy for general language skills, as measured by the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4; Dunn & Dunn, 2007).

Measures

Reading Task
The Group Reading Assessment and Diagnostic Evaluation (GRADE; K. T. Williams, 2001) was administered to obtain participants’ reading levels. The GRADE is a norm-referenced assessment that may be administered in groups. Grade-level forms of the Word Reading, Sentence Comprehension, and Passage Comprehension subtests were administered, and standard scores were calculated. According to the test manual (Williams, 2001), the GRADE has an internal consistency of .95 to .99 (coefficient alpha [α]), and test–retest reliability of .89 to .98.

Receptive Vocabulary Task
The PPVT-4 was administered to determine participants’ receptive vocabulary levels and to corroborate teacher report of receptive language skills within typical limits. The test manual indicates that the PPVT-4 has an internal consistency (split-half reliability) of .94 and test–retest reliability of .92 to .96 (M = .93).

Writing Tasks
Each participant produced one narrative and one expository writing sample (15 min each). Writing samples were elicited during group sessions by the first author and by trained research assistants. Prompts were selected in accordance with parallel forms of the state-mandated writing assessment system protocol (Florida Comprehensive Assessment Test [FCAT]; Florida Department of Education, 2010) and the topical interests of the children enrolled in the classrooms. The prompts provided for narrative and expository genres, respectively, were as follows:

- Tell me about a time that someone surprised you and what happened.
- Pretend you are a super hero and you are being interviewed on the news. Tell everyone what special powers you would have. Also, explain what you would do with them to help the world.

Writing sample elicitation procedures mirrored those of the state’s writing assessment program for elementary grades (FCAT-Writing; Florida Department of Education, 2010), which follows a scripted, generated elicitation method. The writing scale designed for this study, consisting of nine items for microstructure and three for macrostructure, had good internal consistency, with a Cronbach’s α coefficient of .80.

Procedure
Reading (GRADE subtests) and writing tasks were completed in two classroom-wide sessions, and receptive vocabulary was measured per participant in one individual session. Assessments were altogether completed in 6 weeks in the fall of the academic year. Individual sessions were completed first in a quiet testing room on the school campus and lasted approximately 20 min. Group sessions were counterbalanced for writing genre (e.g., expository, narrative). Evaluators were graduate students in speech-language pathology who were trained for each task in small groups by the authors.

Coding and Scoring
The first author transcribed the writing samples into a computer database according to Systematic Analysis of Language Transcript (SALT, Version 8; Miller & Chapman, 2005) conventions. The unit of segmentation was the T-unit, as suggested by Nelson et al. (2004) and consistent with previous investigations (e.g., Nelson & Van Meter, 2007; Puranik et al., 2007, 2008; Scott & Windsor, 2000). A T-unit contains one main clause and any subordinate clauses. Examples of coded writing samples for narrative and expository writing in all three grades are provided in the Appendix.

Microstructure
Nine microstructure elements were calculated, representing measures of productivity, grammatical complexity, and lexical diversity. The microstructure productivity measures of number of total words and T-units were calculated automatically in SALT, as was the microstructure grammatical complexity element of MLTU. Another measure of grammatical complexity, the total number of clauses, was calculated by the first author using SALT to compute CD. Both measures have been used in previous examinations of the written product (e.g., Puranik et al., 2007, 2008; Scott & Windsor, 2000). CD was calculated by dividing the total number of clauses (main and subordinate) in the sample by the total number of T-units across the sample. In addition, the number of clauses per sentence (CPS) was measured to capture grammatical complexity at the sentence level. Transcripts within the SALT program were coded for sentence type (complex vs. simple, correct vs. incorrect) and presence of grammatical errors. A simple sentence consisted of one main clause and only one verb, whereas a complex sentence included one main clause plus one or more embedded/subordinate clauses, two main clauses, or one main clause and a verb phrase joined by a coordinating conjunction. Grammatical errors were defined as errors occurring in verb or pronoun tense, agreement or case, omitted or incorrect inflection, omitted or substituted grammatical elements, and violated word order. A sentence without any grammatical errors was considered correct, whereas a sentence with one or more errors was deemed incorrect.
Two final grammatical complexity measures calculated were the number of grammatical errors per T-unit (GET) and percentage of grammatically correct sentences (percent GS; e.g., Mackie & Dockrell, 2004; Nelson & Van Meter, 2007; Puranik et al., 2007, 2008). The training manual for coders included resources for identifying grammatical errors that may have been attributable to nonstandard dialect use so that considerations were made regarding the potential influence of participant dialect on calculation of grammatical errors.

Two of the three microstructure elements representing lexical diversity—the NDW in the written text and the TTR (ratio of different word types to overall words)—were automatically calculated by SALT. However, previous researchers have suggested that NDW or TTR are most accurately interpreted when sample size is controlled for (Scott, 2009; Scott & Windsor, 2000). For this reason, an additional related measure of lexical properties was used that was not confounded by writing sample size. Lexical density (LXD) was the proportion of content words (e.g., nouns, verbs, adjectives) to total words (Scott, 2009). By taking a proportion of content words to total words, each sample was then measured for LXD on the same scale regardless of overall sample length, thereby reducing the impact of sample size.

Macrostructure

Each writing sample was reviewed and scored for the dependent measures related to macrostructure variables (organization, genre-specific text structure, cohesion) on the basis of an analytic scoring system (see Hall-Mills, 2010 for operational definitions and protocol). The operational definitions for examining levels of organization, text structure, and cohesion were formed on the basis of key features of informal writing inventories used in previous investigations (Crawford et al., 2004; Moats, Foorman, & Taylor, 2006; Nelson et al., 2004). Organization was examined within the introduction, body, and conclusion of the product. Writing samples also were examined for use of an appropriate text structure (genre-specific), and overall cohesion. Each item (organization, genre-specific text structure, cohesion) received a score ranging from 1 to 4. The individual trait scores were combined for an overall macrostructure composite score.

Reliability

Coding reliability of the microstructure and macrostructure measures was established by using a randomly selected subsample of writing samples equaling 25% of the total number of samples collected, balanced across genre and grade level. Percent agreement and Cohen’s kappa coefficients were calculated for the following variable characteristics that required coding in SALT to produce the scores for each of the dependent variables: T-unit segmentation, clauses per T-unit, CPS, sentence codes to indicate grammatical complexity (simple vs. complex) and accuracy (correct vs. incorrect) of the sentence structure, identification of content words, and identification of grammatical errors. Percent agreement ranged from 83% to 98% for the microstructure variables, and from 84% to 93% for macrostructure variables. Kappa coefficients of >.6 were required to establish adequate reliability. Kappa values may be interpreted as follows: .41 to .60 is fair, .61 to .80 is good, and > .80 is very good reliability among raters (Warner, 2008). Kappas ranged from .80 to .98 for microstructure variables and from .72 to .90 for macrostructure variables, thus indicating suitable reliability for all coded dependent measures.

Results

In our preliminary analyses, we surveyed the data for normality (including a search for outliers) and assumptions of the statistical models we employed to answer the research questions. In particular, we considered conformity with four assumptions for multivariate analysis of covariance (MANCOVA) due to potential effects on Type I errors rate and power (e.g., independent observations, multivariate normal distribution in each group, covariance matrices for all dependent variables, and homogeneity of regression slopes (Field, 2005; Stevens, 1997). Multivariate normality could not be checked in the Predictive Analytic Software (PASW) program, so the assumption of univariate normality was checked for each dependent variable by using the Shapiro–Wilk test (Stevens, 1997). To determine whether the assumption of homogeneity of covariance matrices was met, the univariate tests of equality of variances between groups was checked by using Levene’s test for each of the dependent variables (Field, 2005). Due to unequal group sizes, the homogeneity of the variance–covariance matrices was checked by using Box’s test (Field, 2005).

Two exploratory factor analyses (EFAs) were conducted for the purposes of data reduction for the microstructure variables (nine variables) and macrostructure variables (three variables) and to confirm a priori association of dependent writing variables with factors that are conceptually meaningful to clinicians. The resulting factor scores were used in two separate MANCOVAs to address the research questions. We used a principal component analysis factor extraction method. Based on the scree plots and presence of eigenvalues over 1.0, four factors were rotated by using direct oblimin (δ = 0), an oblique rotation technique, yielding a four-factor solution for microstructure. The results for the microstructure factors are reported in Table 1. For both genres, three variables loaded on the first factor of productivity (total words, total T-units, NDW), three variables loaded onto the second factor of grammatical complexity (CD, CPS, MLTU), two variables loaded onto the third factor of grammatical accuracy (percent grammatical sentences, GET), and one variable loaded onto the fourth factor of lexical diversity (LXD). Two grammatical factors (Factor 2, grammatical complexity; Factor 3, grammatical accuracy) were identified. In comparison to previous research, the second factor in the present analysis resembles the accuracy factor (including variables of syntactic errors per T-unit and % GS identified by the factor analysis reported by Puranik et al. (2008). As such, it was determined.
that the two grammatical variables loading onto the second factor in the present study could actually serve as a fourth factor measured within the microstructure measure. Therefore, this fourth factor level was named grammatical accuracy. The results also indicated that NDW, initially proposed as a measure of lexical diversity, actually loaded onto the first factor, with total words and total T-units as measures of productivity.

The dependent variables for the macrostructure measure (organization, text structure, and cohesion) were selected initially on the basis of review of previous research to examine the macrostructure of written language. An EFA utilizing a principal component analysis factor extraction method resulted in the extraction of only one factor for both genres. Examination of a bivariate correlation matrix confirmed that each of the three macrostructure variables was strongly intercorrelated; however, none were greater than .90 where multicollinearity would be of concern. For both genres, all three macrostructure variables loaded on the identified macrostructure factor (organization, text structure, and cohesion). The factor loadings aligned with predictions that the macrostructure measure was in fact unidimensional and yielded one factor measuring macrostructure with three variables (range of narrative loadings = .88-.92, range of expository loadings = .87-.89).

Table 1. Factor loadings for exploratory factor analysis with oblique rotation.

<table>
<thead>
<tr>
<th>Microstructure variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative, expository</td>
<td>.96, .98</td>
<td>.09, .12</td>
<td>-.01, .01</td>
<td>-.01, .05</td>
</tr>
<tr>
<td>Total words</td>
<td>.97, .97</td>
<td>-.22, -.25</td>
<td>-.01, .04</td>
<td>-.02, -.06</td>
</tr>
<tr>
<td>Total T-units</td>
<td>.02, -.39</td>
<td>.05, .24</td>
<td>-.13, -.03</td>
<td>.96, .84</td>
</tr>
<tr>
<td>Lexical density</td>
<td>.96, .97</td>
<td>.12, .13</td>
<td>.01, -.03</td>
<td>.05, .02</td>
</tr>
<tr>
<td>Number of different words</td>
<td>.02, .01</td>
<td>.92, .90</td>
<td>.02, -.01</td>
<td>.05, -.01</td>
</tr>
<tr>
<td>Clausal density</td>
<td>.02, .01</td>
<td>.72, .70</td>
<td>-.23, .02</td>
<td>-.42, -.34</td>
</tr>
<tr>
<td>Clauses per sentence</td>
<td>-.01, -.02</td>
<td>.86, .87</td>
<td>.21, -.04</td>
<td>.14, .21</td>
</tr>
<tr>
<td>Mean length of T-unit</td>
<td>-.01, -.02</td>
<td>.05, -.05</td>
<td>-.92, .86</td>
<td>.11, .19</td>
</tr>
<tr>
<td>Percent grammatical sentences</td>
<td>-.01, -.02</td>
<td>.13, -.04</td>
<td>.92, -.94</td>
<td>-.01, .14</td>
</tr>
</tbody>
</table>

Note. Each variable (e.g., total words, total T-units, etc.) has two scores: the narrative score and the expository score. Factor loadings > .60 are in boldface. Factor 1 = productivity; Factor 2 = grammatical complexity; Factor 3 = grammatical accuracy; Factor 4 = lexical density.

Effects of Grade and Genre on Microstructure and Macrostructure

The GRADE Comprehension Composite scores, after being converted to z scores, were used as a covariate for MANCOVA because reading comprehension is strongly associated with writing skills (e.g., Cox et al., 1990; G. J. Williams, Larkin, & Blaggan, 2013). A preliminary analysis to evaluate the homogeneity of slopes assumption indicated that the relationship between the covariate (z score for GRADE Comprehension Composite) and the dependent variables did not differ significantly as a function of the independent variable (grade level) in either the narrative genre, $F(3, 82) = 0.25$ to 1.6, $p = .19$ to .86, $\eta^2 = .01$ to .06, or the expository genre, $F(3, 79) = 0.16$ to 3.03, $p = .03$ to .92, $\eta^2 = .01$ to .10.

Two MANCOVAs were conducted to measure the effects of grade level within each genre on the various microstructure and macrostructure variables, controlling for reading comprehension scores. Table 2 shows the final factors of analysis (i.e., resulting from EFA) with the respective dependent variables. Factor scores of productivity, grammatical complexity, grammatical accuracy, lexical diversity, and macrostructure were examined in each genre for differences between grade levels. Within the factor of productivity, there were three variables (total words, total T-units, NDW); the factor of grammatical complexity included three variables (CD, CPS, and MLTU); grammatical accuracy had two variables (percent grammatical sentences, GET); lexical diversity was represented by one variable (LXD); and macrostructure factor was represented by three variables (organization, text structure, and cohesion).

Tables 3 and 4 contain the grade-level means and standard deviation for the dependent measures for each genre. Two separate, one-way MANCOVAs were conducted to determine the effect of grade in both genres (narrative, expository).
expository) on the identified factor scores for the five factors of productivity, grammatical complexity, grammatical accuracy, lexical diversity, and macrostructure. The Box M test (using $\alpha = .01$ as the criterion for significance) satisfied the assumption of homogeneity of variance/covariance matrices. Intercorrelations between measures ranged from .01 to .52 for the narrative genre and $-\cdot08$ to .40 for the expository genre. None of the correlations among outcome variables was sufficiently large to raise concerns about multicollinearity. Pillai’s trace was selected as the multivariate test statistic due to unequal group sizes (Field, 2005). Pairwise comparisons were conducted to compare performance across grades, utilizing an adjustment for multiple comparisons and corrected for Type I error rate using a Bonferroni correction ($p < .003$).

The MANCOVAs (one for each genre) to examine the effects of grade level revealed a significant multivariate effect of grade for both the narrative genre, Pillai’s trace = .60, $F(10, 154) = 6.623, p < .001$, $\eta_p^2 = .30$, and the expository genre, Pillai’s trace = .55, $F(10, 148) = 5.583, p < .001$, $\eta_p^2 = .27$, with large effect sizes. By analyzing the component measures, a significant main effect was observed for the narrative productivity factor, $F(2, 80) = 30.745, p < .001$, $\eta_p^2 = .44$, with a large effect size. This main effect was indicated by an overall increase in productivity at every grade level, as reflected by total words, total T-units, and NDW. Pairwise comparisons revealed that children in second grade scored significantly lower on productivity ($M = -0.86$) relative to the third ($M = 0.05$) and fourth ($M = 0.77$) grade groups. The third and fourth grade groups differed

### Table 3. Descriptive statistics for dependent measures; narrative genre.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Grade 2</th>
<th></th>
<th>Grade 3</th>
<th></th>
<th>Grade 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total words</td>
<td>24.27</td>
<td>12.22</td>
<td>53.24</td>
<td>25.57</td>
<td>77.59</td>
<td>34.56</td>
</tr>
<tr>
<td>Total T-units</td>
<td>3.46</td>
<td>1.75</td>
<td>6.45</td>
<td>3.28</td>
<td>9.59</td>
<td>4.42</td>
</tr>
<tr>
<td>NDW</td>
<td>18.31</td>
<td>7.05</td>
<td>36.88</td>
<td>14.58</td>
<td>47.33</td>
<td>17.06</td>
</tr>
<tr>
<td>Grammatical complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length T-unit</td>
<td>7.51</td>
<td>2.87</td>
<td>8.49</td>
<td>2.39</td>
<td>7.98</td>
<td>2.02</td>
</tr>
<tr>
<td>Clauses (per sentence)</td>
<td>1.83</td>
<td>0.77</td>
<td>1.96</td>
<td>0.55</td>
<td>1.74</td>
<td>0.50</td>
</tr>
<tr>
<td>Clausal density</td>
<td>1.40</td>
<td>0.40</td>
<td>1.43</td>
<td>0.34</td>
<td>1.46</td>
<td>0.29</td>
</tr>
<tr>
<td>Grammatical accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% grammatical sentences</td>
<td>0.81</td>
<td>0.29</td>
<td>0.84</td>
<td>0.19</td>
<td>0.82</td>
<td>0.19</td>
</tr>
<tr>
<td>Grammar errors per T-unit</td>
<td>0.22</td>
<td>0.29</td>
<td>0.23</td>
<td>0.40</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>Lexical density</td>
<td>0.52</td>
<td>0.08</td>
<td>0.53</td>
<td>0.05</td>
<td>0.55</td>
<td>0.05</td>
</tr>
<tr>
<td>Macrostructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>4.92</td>
<td>1.29</td>
<td>6.79</td>
<td>1.59</td>
<td>6.78</td>
<td>1.39</td>
</tr>
<tr>
<td>Text structure</td>
<td>1.54</td>
<td>0.86</td>
<td>2.76</td>
<td>0.83</td>
<td>2.52</td>
<td>0.85</td>
</tr>
<tr>
<td>Cohesion</td>
<td>1.31</td>
<td>0.47</td>
<td>2.00</td>
<td>0.88</td>
<td>1.93</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Note.** NDW = number of different words.

### Table 4. Descriptive statistics for dependent measures; expository genre.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Grade 2</th>
<th></th>
<th>Grade 3</th>
<th></th>
<th>Grade 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total words</td>
<td>27.77</td>
<td>12.00</td>
<td>53.97</td>
<td>22.11</td>
<td>73.46</td>
<td>29.55</td>
</tr>
<tr>
<td>Total T-units</td>
<td>3.73</td>
<td>1.22</td>
<td>6.62</td>
<td>2.96</td>
<td>8.84</td>
<td>4.20</td>
</tr>
<tr>
<td>NDW</td>
<td>20.85</td>
<td>7.59</td>
<td>37.56</td>
<td>12.71</td>
<td>47.85</td>
<td>16.92</td>
</tr>
<tr>
<td>Grammatical complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length T-unit</td>
<td>7.58</td>
<td>3.05</td>
<td>8.58</td>
<td>2.23</td>
<td>8.33</td>
<td>1.87</td>
</tr>
<tr>
<td>Clauses (per sentence)</td>
<td>1.49</td>
<td>0.59</td>
<td>1.83</td>
<td>0.44</td>
<td>2.05</td>
<td>0.58</td>
</tr>
<tr>
<td>Clausal density</td>
<td>1.25</td>
<td>0.32</td>
<td>1.58</td>
<td>0.30</td>
<td>1.61</td>
<td>0.43</td>
</tr>
<tr>
<td>Grammatical accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% grammatical sentences</td>
<td>0.78</td>
<td>0.27</td>
<td>0.85</td>
<td>0.21</td>
<td>0.76</td>
<td>0.25</td>
</tr>
<tr>
<td>Grammar errors per T-unit</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>0.27</td>
<td>0.27</td>
<td>0.25</td>
</tr>
<tr>
<td>Lexical density</td>
<td>0.62</td>
<td>0.09</td>
<td>0.60</td>
<td>0.07</td>
<td>0.61</td>
<td>0.06</td>
</tr>
<tr>
<td>Macrostructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>4.04</td>
<td>1.25</td>
<td>4.66</td>
<td>1.49</td>
<td>4.85</td>
<td>1.40</td>
</tr>
<tr>
<td>Text structure</td>
<td>1.31</td>
<td>0.55</td>
<td>1.81</td>
<td>0.78</td>
<td>1.92</td>
<td>0.79</td>
</tr>
<tr>
<td>Cohesion</td>
<td>1.15</td>
<td>0.46</td>
<td>1.34</td>
<td>0.48</td>
<td>1.46</td>
<td>0.58</td>
</tr>
</tbody>
</table>

**Note.** NDW = number of different words.
significantly from each other in productivity as well, with increased productivity levels in fourth grade. A significant main effect also was observed for narrative macrostructure, $F(2, 80) = 16.358, p < .001, \eta_p^2 = .29$, with a large effect size. Second grade scored significantly lower in narrative macrostructure ($M = −.79$) than both third ($M = 0.39$) and fourth ($M = 0.30$) grades; however, third grade did not differ significantly from fourth. In the narrative genre, there were no significant differences among the three grade-level groups in scores on grammatical complexity, grammatical accuracy, and lexical diversity.

Similar results were obtained in the expository genre, with a significant main effect for productivity, $F(2, 77) = 32.318, p < .001, \eta_p^2 = .46$, with a large effect size. Pairwise comparisons indicated that second grade students scored significantly lower on expository productivity ($M = −0.83$) than did third ($M = 0.08$) and fourth grades ($M = 0.75$). The third and fourth grade groups differed significantly from each other in expository productivity as well, with greater productivity in the fourth grade. In addition, there was a significant main effect for expository macrostructure, $F(2, 77) = 4.961, p < .01, \eta_p^2 = .12$, with a medium effect size. Second grade ($M = −0.43$) scored significantly lower than both third ($M = 0.11$) and fourth grade ($M = 0.30$). Even though there was a trend for fourth grade students to score higher on macrostructure, there was not a significant difference between third and fourth grades. Furthermore, a significant main effect was detected for grammatical complexity, $F(2, 77) = 4.922, p < .01, \eta_p^2 = .13$, with a medium effect size. Second grade ($M = −0.51$) scored lower in expository grammatical complexity than did third ($M = 0.18$) and fourth grade ($M = 0.29$). Third and fourth grades were not significantly different from each other. There were no significant differences among the three grade-level groups in scores on grammatical accuracy and lexical diversity in expository samples.

Discussion

Effect of Grade Level on Microstructure

The first research goal was to determine whether there were differences among grades and between genres in linguistic microstructure elements. As hypothesized, differences were found between grade levels (second and third, second and fourth, and third and fourth) for microstructure productivity in both genres. Participants in each successive grade used more words, produced more T-units, and had greater numbers of different words than did students in the previous grade. These findings are consistent with results from previous investigations showing that measures of productivity are sensitive to changes in grade and age levels in more than one genre (Berman & Verhoeven, 2002; Nelson & Van Meter, 2007; Puranik et al., 2008).

Partially confirming the hypothesis for grammatical complexity, differences were found between two adjacent grade levels (second and third), as well as between second and fourth grades, in the expository genre. However, similar to the findings of Puranik et al. (2008), no significant differences in grammatical complexity were indicated between third and fourth grade levels. Levels of expository grammatical complexity, as measured by MLTU, CPS, and CD, may plateau at third grade, at least on writing assignments similar to those used in this study and in Puranik et al. (2008).

In contrast to expectations, no grade-level differences were found for grammatical complexity in the narrative genre. This result is dissimilar from the results of Nelson and Van Meter (2007), who found a significant difference on MLTU between second and third grades in a narrative genre. Examination of grade-level means in Table 4 for CD and CPS clearly indicates nearly the same scores across the three grade levels. Therefore, it may be that CPS and CD were not sensitive enough to detect differences between subsequent grade levels in grammatical complexity in a narrative genre. These values for CPS and CD may have masked the effects of MLTU when these variables were combined into one factor score that was subsequently used in the MANCOVA to detect grade differences. However, previous research with older students suggests that development of CD in younger students consists of multiple periods of slower, sometimes plateaued growth prior to the eighth grade and may be a useful measure for detecting grade effects in later years (Nippold, Ward-Lonergan, & Fanning, 2005). It might be useful in future studies to compare measures of grammatical complexity to determine their utility for detecting significant differences among elementary grade-level groups.

No grade-level differences were found for either grammatical accuracy or lexical diversity. This finding is similar to Puranik et al. (2008), although they utilized different measures for this dimension of microstructure (percentage of grammatically correct T-units, proportion of spelling errors, conventions) and is in alignment with Nelson and Van Meter’s (2007) finding for grammatical error rates among Grades 1 through 5. In the present investigation, grammatical accuracy was measured by the % GS and the number of GET. Puranik et al. (2008) posited that older children may attempt to produce more complex sentence structures and as a result could generate more errors than expected, making their grammatical accuracy scores more parallel with those of younger participants. We did not find a statistically significant pattern between grade levels in grammatical accuracy. However, the nature of the relationship between grammatical accuracy and grade level warrants further exploration. Therefore, future investigators may want to increase the number of items, or variables, measured for both grammatical complexity and accuracy factors and compare the relations among them under varying contexts and task demands.

Effect of Genre on Microstructure

In contrast to expectations, mean performance on all four microstructure measures was rather stable within grades and across genres. This finding initially appeared
different from previous studies that indicated greater productivity and grammatical complexity in a narrative versus an expository genre among elementary school children (Berman & Verhoeven, 2002; Scott & Windsor, 2000); however, a direct comparison to students in the present study cannot be made to Berman and Verhoeven’s (2002) and Scott and Windsor’s (2000) samples. The finding of no genre effect for either lexical diversity or grammatical accuracy is similar to previous findings for elementary students (Berman & Verhoeven, 2002; Scott & Windsor 2000). The developmental progression of lexical diversity and grammatical accuracy may be slower overall, regardless of genre, and therefore more challenging to detect differences between subsequent elementary grades.

Effect of Grade Level on Macrostructure

As hypothesized, we found grade-level differences in both genres for macrostructure between second and third grades, and between second and fourth grades. Second grade students scored lower on narrative and expository organization, text structure, and cohesion than third and fourth grade students. However, in contrast to expectations, third and fourth grade students did not differ significantly from each other in either genre. In examining the grade-level means in the narrative genre, a trend was noted for the third grade mean to exceed fourth in text structure and cohesion, although these differences were not statistically significant. In the expository genre, grade-level means indicated a slight trend between third and fourth grades on organization, text structure, and cohesion.

Upon further analysis, the grade-level trends in macrostructure observed in the expository genre between second and third, and second and fourth appear to be due to text structure scores more so than organization or cohesion scores. It may be that an expository writing sample could be somewhat organized and cohesive yet not reflect the targeted genre structure. This is understandable, especially for second grade students, whose knowledge of expository text structure is thought to be less established than for older students (e.g., Berman & Verhoeven, 2002; Nelson et al., 2004). A single second grade participant scored at a Level 3 for expository text structure on the macrostructure rubric, whereas the remaining second grade students scored at Level 1 (n = 19) or Level 2 (n = 6). This was in contrast to the number of third grade students (n = 5) and fourth grade students (n = 5) scoring at a Level 3 or higher and also was dissimilar from the number of students in the third and fourth grades scoring at a Level 2 (third grade, n = 15; fourth grade, n = 13). Thus, the grade-level trends appeared to be impacted by floor effects for the second grade students on measures of macrostructure. In the future, researchers may want to develop more sensitive measures to examine young children’s knowledge of genre-specific text structures across grade levels, particularly their knowledge of text structure requirements for the genres most commonly assessed in the classroom, to help further define this developmental progression.

Effect of Genre on Macrostructure

In contrast to expectations, our results did not reveal any cross-genre differences for performance on the macrostructure measure. The mean macrostructure values were very similar across genres for each grade level. The potential genre effects on macrostructure were difficult to anticipate given the range of findings in previous investigations. However, we had anticipated that performance on macrostructure measures would be similar in both genres for the oldest students (fourth grade) on the basis of the assumption that experience and knowledge of various text structures and text cohesion increase with age.

The lack of significant differences in macrostructure scores between the narrative and expository genres may be due to the type of writing prompts utilized to elicit the samples. It is possible that the prompts selected may not have fully represented one genre versus another. With this issue in mind, we conducted a post hoc analysis to determine the degree of match between intended text structure on the basis of the prompts and the text structure produced in the children’s writing samples. We recruited three blind raters (i.e., blind to each other as well as the writing prompts, previous transcription, coding, and scoring of writing samples for this investigation) and asked them to review a randomly selected subset of writing samples (25% of the entire sample; equal proportions by genre and grade) to identify which genre structure the sample most closely resembled, given a set of standard definitions and features for each genre. Raters were instructed to identify a writing sample with the following features as having a narrative text structure (e.g., telling a story, often about personal events or other life experiences; it may be fictional or nonfictional and can include novels, personal stories, and short stories), an expository text structure (e.g., conveying facts, describing procedures, explaining something, sharing basic information, relating cause-effect, compare/contrast, problem-solution structures, and may include reports, term papers, procedural documents, manuals, essays, and editorials), or an ambiguous text structure (e.g., combination of narrative and expository features, or cases where the text structure cannot be readily identified given standard academic definitions of genre). The post hoc analysis with the genre identification task revealed that 72% of the samples were coded reliably, with consensus for genre type by three naive raters (κ = .67). Thus, it appears the majority of writing samples reflected their intended genre structure. When coders did not reach consensus on some of the samples reviewed (18%), they noted instances of “ambiguous structures.” On closer examination, the text structure ambiguity in writing samples was observed equally across narrative and expository topics and thus was not wholly attributable to either the narrative or the expository prompt on the analytic scale. Some children may have produced text in a structure other than the elicited genre, and there are a variety of possible explanations for this occurrence. It may be the case that these children utilized inconsistent or limited genre structures in their writing as they were learning to
write. This issue is addressed further below among important clinical and educational implications. It remains to be determined within a larger sample of school-age students whether writing that contains features of more than one genre reflects maturity versus instability of academic discourse forms. The lack of significant differences in macrostructure also may be due to the use of “hybrid” genres that incorporate a variety of text structure features from one or more genres. In situations such as this, the scoring rubric designed for this investigation would need to be altered to better capture occurrences of “mixed” or hybrid genres. Further exploration is warranted to establish how adequately writing prompts elicit the intended genre.

Limitations and Future Research

One potential limitation to our study was the method used to elicit writing samples. In this investigation, a single elicitation technique was incorporated (i.e., response to writing prompt). Although this method reflects standard educational practice, it is important to consider that grade and genre effects may vary as a result of differences in prompting procedures and targeted genre structure. More work is needed in this area to compare the value of various elicitation techniques to capture the possible relations between elicitation method and writing outcomes. Furthermore, the degree of the relations among elements of microstructure and macrostructure may be shaped by the actual genre structure produced. In our investigation, the post hoc analysis revealed that a small portion of the writing samples could not be characterized specifically as having a narrative or expository text structure due to either mixed text structure elements of both genres or an ambiguous genre structure. Although the occurrence of mixed elements can be expected to some degree in early writing development, it is possible that the expository prompt may have misled some students to compose a fictional narrative instead. The central characteristics of the expository prompt that would cue students to engage in expository discourse were the key words tell and explain. The word pretend (e.g., pretend you are a superhero) may have steered students in a different direction if they did not process the exposition cue words. Given all of the data collectively, including the post hoc analysis, we did not have enough evidence to state that it was the nature of the prompts alone that attributed to the occurrence of mixed genre features or ambiguous structures for a small portion of the sample. Yet, the expository prompt we selected warrants caution for use in future writing activities.

Prompt selection has important implications for research and clinical or instructional practices. Investigators planning future studies may elect to first establish the reliability of selected prompts to elicit the intended genre and plan in advance an analysis to verify the reliability of selected prompts within their sample. Likewise, educators and clinicians need to select prompts for writing assessment carefully. As they plan and conduct writing assessments, they should know the reliability of selected or created prompts to elicit the genre of interest. Otherwise, the likelihood of eliciting the appropriate genre and therefore the opportunity to comprehensively evaluate the child’s writing performance across microstructure and macrostructure features may be compromised.

One controversy that remains in writing assessment is the utility of holistic versus analytic measures of writing performance. The fact that the macrostructure measure was unidimensional (according to the EFA results) would seem to be contradictory to some authors’ recommendations against the use of holistic score ratings of writing performance to inform instruction and monitor growth (Nelson & Van Meter, 2007). However, as noted in the present investigation, a holistic rating scale for macrostructure was a useful method to compare a particular student’s or grade level’s performance in comparison to peers or comparison groups. In contrast, EFA results indicated that the microstructure measure consisted of four distinct factors. As such, microstructure, in contrast to macrostructure, would be best examined with an analytic scoring method, utilizing more than one factor or score. Either way, the purpose for the writing assessment, as well as the reliability of a particular scale to fulfill that purpose (see Koutsoftas & Gray, 2012), should be the focus at the outset. In some states, the statewide progress monitoring measure of writing in the elementary grades is administered multiple times per school year and yields only a holistic score. Four factors often considered in the students’ text are focus, organization, support, and conventions. Educators are first cautioned against using this single score as the sole determinant of a student’s writing proficiency and are encouraged to interpret this score in light of the student’s performance in other writing tasks and contexts.

A second limitation of our investigation is the lack of an independent measure of participants’ dialect of speech. The influence of cultural–linguistic factors on writing performance (e.g., ethnicity and dialect) is important to consider. We did not assess dialect directly in the present study. We know from the participant demographics that there were similar numbers of children across ethnic categories within each grade in our sample. However, one must acknowledge that ethnicity and dialect are not equivalent, and each factor warrants separate consideration. Regardless of reported ethnicity of participants, it is possible that dialectal influences may have existed for individual children and could have affected the outcome measures of grammatical accuracy. The dialect-shifting–reading achievement hypothesis suggests that students who successfully shift from nonmainstream to standard dialectal forms in different literacy tasks (including writing) demonstrate better reading outcomes than students who do not make the shift as adequately (Craig, Zhang, Hensel, & Quinn, 2009). Investigations employing larger samples of participants with ethnically diverse backgrounds, and incorporating distinct a priori measures of dialect (e.g., dialect density measures), may have better chances of detecting possible differences. If differences are indeed detected in this manner, investigators can recode the SALT files to capture features of a specific dialect that has been observed in the sample. It would be
worthwhile to compare results for written grammatical and lexical microstructure variables, as well as text structure influences, of dialectal speakers to capture the weight of influence that dialectal differences may exert on dependent writing measures for both microstructure and macrostructure (Terry, 2006; Thompson, Craig, & Washington, 2004). Clinicians should utilize dialect coding procedures as well.

**Educational and Clinical Implications**

Results of the present investigation extend findings from previous studies and add to the existing literature regarding development of and relations among written microstructure and macrostructure features within and across grade levels and genre types. Educators and clinicians are encouraged to consider the lack of differences between grade levels for some of the dependent measures in light of established grade-level expectations that are reflected in state standards for writing. For example, the Common Core Standards for writing (Common Core State Standards Initiative, 2010) require second grade students to write in a variety of genres, including expository and narrative forms. Third and fourth grade students are expected to write routinely in multiple genres with increasing detail, genre-specific organization, and cohesion. In reviewing the present data, it is clear that not all of the writing samples that we collected reflected mastery of the previous grade level’s standards for writing. This raises the question: If the established grade-level expectations are considered reasonable and the Common Core State Standards are evidence based, then how well are current assessment and instructional practices designed to support student achievement of these standards?

The nature of the writing standards is a substantial factor in how writing assignments and assessments are structured within the school context. In particular, the selection of writing prompts and the method by which the examiner chooses to evaluate the written product are influenced by curricular standards for writing across grade levels. Prompts to elicit writing from students often are crafted on the basis of specific curricular standards for the grade level or grade-level group of interest. For example, students in Grade 2 are expected to “Write narratives in which they recount a well-organized sequence of events; details about characters, actions, feelings, in temporal words to signal event order, and provide a sense of closure” (Common Core State Standards Initiative, 2010, CCSS.ELA-Literacy.W.2.3). Therefore, a narrative writing prompt for a second grade writing assessment would need to be designed or selected to ensure that it would cue the students to write a basic story. General assessment prompts provide the topic and the basic purpose of the writing activity but rarely include any additional instructions nor details or lists of components to be included. For instructional or intervention purposes, however, these additional details can be included to provide a greater level of support to students learning to write. For example, clinicians and educators have found it useful to provide graphic organizers, story grammar or exposition checklists, and related resources to young writers to further support their efforts in incorporating a variety of narrative and expository elements in their work.

The method by which the examiner evaluates the written product is also influenced by curricular standards for writing. In the example given above from a second grade writing standard, there would need to be analysis of microstructure features (e.g., lexical properties such as words to signal temporal relations among events) as well as macrostructure features (e.g., story grammar elements: a plot occurring in a sequence; details about characters’ actions, thoughts, and feelings; and a resolution to provide a sense of closure) to comprehensively measure a student’s mastery of the standard. Not all writing assessment methods are designed to measure both levels of children’s written work. The risk is that there can be misalignment of standards, assessment techniques, and instructional or intervention approaches implemented with the school-age population. Clinicians and educators should be encouraged to make these important alignment considerations in their practice of supporting the development of writing proficiency among their school-age students. Moreover, future research needs to determine the extent to which writing instruction, assessment, and progress monitoring adhere to grade-level standards for writing performance. In the meantime, state writing standards that are being developed or revised need to be research based, and educators, clinicians, researchers, and policymakers need to work collaboratively to design assessment and instruction that are reflective of research-based standards for writing.

**Acknowledgments**

This research was partially funded by a grant from the U.S. Department of Education, Office of Special Education and Rehabilitative Services (H8435D030046), awarded to the first author, and a Florida State University Dissertation Research Grant, also awarded to the first author.

**References**


Appendix

Narrative and Expository Writing Samples

**Second Grade Narrative**
Once my sister Lily scared me by hiding on the walls when I was in the restroom [3CL][CC]. When I came out she went "ROAR"! [2CL][CC]. It was scary.
and I screamed [2CL][CC]. Then I started to laugh so hard I fell on the ground [2CL][CC].

**Third Grade Narrative**
One day a man went to McDonald’s to get something to eat through the drive through [3CL][CC]. When he got his lunch the bag started to shake [2CL][CC]. The hamburger popped out, and the french fries popped out.
and they grow [GE] and grow [GE] until they became humongous [4CL][CI]. They took over the world [1CL][SC]. But in a matter of time they shrunk.
and the man had his lunch [2CL][CC].

**Fourth Grade Narrative**
When I got surprised it was when my cousin told me she broke her wrist [4CL][CC]. it really happened yesterday [1CL][SC]. she was playing soccer and running backwards and tripped on grass.
and her hand caught her and broke [5CL][CC]. she is [GE] getting a light green cast [1CL][SI]. Judith wished she broke her other hand.
so she didn’t have to write [3CL][CC]. But she broke her left hand [1CL][SC].

**Second Grade Expository**
I am a superhero.
and I have super powers [2CL][CC]. My power is to turn into lava [1CL][SC].
I can also turn invisible [1CL][SC].
I can use my powers to help save the world [2CL][CC]. When I see a villain, I turn invisible.
and then I burn them up like they are on a grill [4CL][CC]. That is how I save the world [1CL][SC].

**Third Grade Expository**
Hi my name is Gardenia [1CL][SC]. I have the powers to make things grow right after I plant the seeds [3CL][CC]! I would like to help the Earth by turning landfills into beautiful gardens [2CL][CC]. That’s all I have to say for today, bye [2CL][CC].

**Fourth Grade Expository**
I am a superhero named Superman [1CL][SC]. My special powers are that I am very strong.
and I can beat up everyone. and I can fly [4CL][CC].
I will help the world by beating up evil villains and arresting people [3CL][CC].
I will help elderly people cross the road and give money to banks, homeless shelters, nursing homes, and orphans [2CL][CC]. That is who I am and how I will help the world [2CL][CC].

Note. CC = complex correct sentence; CL = clause for number of clauses; SC = simple correct sentence; SI = simple incorrect sentence.