

Manuscript Information

Journal name: Journal of psychoeducational assessment
NIHMS ID: NIHMS721074
Manuscript Title: Informative, compare and contrast, and persuasive essays of fifth and seventh graders: Not all essay writing the same.
Submitter: Virginia W. Berninger (vwb@u.washington.edu)

Manuscript Files

Type	Fig/Table #	Filename	Size	Uploaded
manuscript		Journal of Psychoeducational Assessment-2015-Davidson-0734282915604977.pdf	372550	2015-09-24 17:57:23

This PDF receipt will only be used as the basis for generating PubMed Central (PMC) documents. PMC documents will be made available for review after conversion. Any corrections that need to be made will be done at that time. No materials will be released to PMC without the approval of an author. Only the PMC documents will appear on PubMed Central -- this PDF Receipt will not appear on PubMed Central.

Informative, Compare and Contrast, and Persuasive Essay Composing of Fifth and Seventh Graders: Not All Essay Writing Is the Same

Matt Davidson¹ and Virginia Berninger¹

Abstract

Typically developing writers in fifth ($n = 110$, $M = 10$ years 8 months) or seventh ($n = 97$, $M = 12$ years 7 months) grade wrote informative, compare and contrast, and persuasive essays for which the content was held constant—two mountains with a history of volcanic eruption. Relevant background knowledge was provided by reading text and showing colorful illustrations to the students before writing each genre. Results showed considerable variability between genre pairs within and across individual writers in content quality, organization quality, and length. Results, which support multiple expository genres, are consistent with prior research showing multiple genres (narrative vs. expository or even within narrative). Results are discussed in reference to the importance of assessing multiple genres in inferring composing expertise as emphasized by Olinghouse and colleagues.

Keywords

measurement, education assessment, writing, disciplines and subjects, correlational, research methods

Sources of Variation in Written Composition

Generativity of Language

Any assessment study that wishes to draw conclusions about writing ability based on written products faces a host of confounding variables and numerous sources of variance. Chomsky (2006) called attention to the generativity of language—the same language units can be combined in multiple ways to express thought. Likewise, the same word can be used to express different ideas, and the same idea can be expressed with different words (Stahl & Nagy, 2005). Words can be combined in different ways within syntax to express comparable ideas, and the same idea can be expressed in multiple multi-word syntactic structures (Chomsky). Likewise, written composition is generative: Thought can be translated into a variety of genres or discourse

¹University of Washington, Seattle, USA

Corresponding Author:

Matt Davidson, University of Washington, 2012 Skagit Lane, Miller Hall, Seattle, WA 98195, USA.

Email: mattjd@uw.edu; vwb@u.washington.edu

structures, and discourse structures can be translated into a variety of kinds of thoughts (Fayol, Alamargot, & Berninger, 2012). Thus, generativity of language is one source of individual differences in composing.

Topics

Features of the writing prompt, which establishes the topic, may also be an important source of variance in essay ratings. Schoonen (2005) had sixth-grade students write four essays, each of which was analyzed both holistically and analytically by raters, in terms of their organization, content, and language use. Using a structural equation model to estimate the variance components in students' scores, he found, among other things, more variance due to different writing prompts than due to raters. Kobrin, Deng, and Shaw (2011), who studied prompt characteristics for SAT writing tasks, found that prompts which presented two sides of an argument led writers to produce slightly longer texts. On the other hand, the topic may not be the only source of variance in quality of writing. In their study, van den Bergh, De Maeyer, van Wiejen, and Tillema (2012) found that holistic ratings of essay quality were not so dependent on topics. Regardless of the prompt used, a single writing topic may not be adequate to assess composing ability, as explained in the next section, "Genre."

Genre

Writing genre may also matter in assessing composing ability. Olinghouse, Santangelo, and Wilson (2012) examined the validity of "single-occasion, single-genre, holistically scored" pieces of writing (p. 55). They specifically investigated the validity of drawing an inference from a single score to the student's writing ability across both genres and curricular requirements. After having participants compose in three genres (story, informative, and persuasive), the authors computed correlations of individual compositions across genre pairs. They argued that strong correlations would indicate similar rank ordering across genre, while low or moderate would show different orderings for students. For their set of fifth-grade participants, Olinghouse et al. (2012) found moderate correlations across genre pairs for the holistic quality ratings (.37-.48), with lower correlations for genre elements (.12-.23).

Others have studied the genre-related linguistic skills needed for written composing. Some of this line of research on genre has been related to taking annual tests yoked to state standards (Troia & Olinghouse, 2013). For example, Beck and Jeffery (2007) investigated which genres students were producing for high-stakes state assessments in Texas, California, and New York. The authors provided a framework for detecting genre features related to specific linguistic structures such as particular phrases that are associated with a particular genre. Indeed, much of the current research on genre is grounded in prior research showing that oral as well as written discourse knowledge informs quality of genre writing (Gillespie, Olinghouse, & Graham, 2013; Olinghouse & Graham, 2009; Scott, 1994) and so does vocabulary knowledge (Olinghouse & Wilson, 2013).

Yet others have studied the contrasting cognitive processes involved in composing different genres, ranging from the hard work to the play involved, as illustrated by the two examples that follow. Olive, Favart, Beauvais, and Beauvais (2009) studied the cognitive effort required to compose narrative and argumentative texts using a reaction time task in which fifth and ninth graders pressed a button with their non-dominant hand each time they heard a beep during a writing task. Ninth graders appeared to exert less effort than fifth graders only for the argumentative texts. The researchers also measured the number and diversity of types of connectives, such as transition words, conjunctions, and logical connectors which tie text together, in each genre. They found that ninth graders used more, and more diverse, connective phrases in general than fifth graders, but that more connectives were used by both fifth and ninth graders in argumentative than narrative genre. Boscolo, Gelati, and Galvan (2012), in contrast, studied play with multiple narrative genres in written expression. For example, personal narratives and third person

narratives may have very different text content and organization and use different vocabulary choices.

Genre also influences writing through writing–reading relationships. Much of school writing involves writing summaries of what has been read or writing reports based on a variety of source material, which also varies in genres (Moore & MacArthur, 2012). So genre may influence both read and written texts.

Research Aims of the Current Study

In contrast to some studies that investigate only one of the many writing genres, this study investigated multiple writing genres within expository writing; although essay writing is sometimes assumed to be a homogeneous genre in contrast to narrative writing, there are actually multiple genres of essay writing. Little is known, however, about whether a single aptitude for expository writing underlies all essay writing or is specific to the kind of essay being written. Whereas some research on multiple writing genres has focused on the early grades (e.g., Kamberelis, 1999), the current study focused on multiple genres in the upper elementary (fifth) and middle of middle school (seventh) grades. A sizable body of research has investigated effective ways to teach developing writers to compose in a variety of genres—both typically developing and those who struggle for a variety of reasons, including but not restricted to specific learning disabilities (SLDs; e.g., Boscolo et al., 2012; Epstein-Jannai, 2004; Troia, 2009). In contrast, the current study was designed to be part of the larger line of research on improving assessment of composition across genres. Such assessment research could have important applications to (a) large-scale writing assessments of genre for all students in the Common Core Era including those with SLDs (e.g., Olinghouse & Colwell, 2013), and (b) individually administered clinical assessments for students with a variety of school-related learning problems or talents.

Both correlations and mean difference were examined for assessing three kinds of essays during middle childhood and early adolescence—informative, compare/contrast, and persuasive. Two kinds of correlations were examined—Pearson product moment correlations (r_s) and intra-class correlation coefficients (ICCs). The first is sensitive to intra-individual differences within students. The second is sensitive to inter-individual differences among students. The first hypothesis tested was that fifth graders and seventh graders would exhibit intra-individual differences across the three kinds of essays, as evidenced by low to moderate r_s across pairs of essays on quality ratings for content and organization. That is, each student may exhibit relative strengths or relative weaknesses in the kinds of essay writing that is easier for her or him. The second hypothesis tested was that the fifth and seventh graders would show significant and sizable ICCs, indicating sizable differences among participants in the different kinds of essay writing. Evidence of such differences within and among students would provide further evidence for the generativity of composing across three genres of essay writing.

Method

Participants

Both groups were in Year 5 of a 5-year longitudinal study of typical oral and written language development when the writing samples analyzed in the current study were collected. One group, which began in first grade, was in fifth grade ($n = 110$; 44% male, 56% female). Their mean age was 10 years 8 months (128.50 months, $SD = 3.64$). They represented the diversity present in the local school population where the study was conducted: European American (64.8%), Asian American (23.4%), African American (6.3%), Hispanic (1.6%), Native American (1.6%), and other (2.3%). The other group, which began in third grade, was in seventh grade ($n = 97$; 49.5% male, 50.5% female). Their mean age was 12 years 7 months (151.21 months, $SD = 3.71$). They

were similarly representative of the local diversity: European American (65.5%), Asian American (21.2%), African American (9.7%), Hispanic (0.9%), and other (2.7%). A variety of educational levels were also represented by parents, from no high school education to completion of a graduate degree, of students at both grade level, but the mode was college.

Procedures

Writing samples were collected during the 4-hr annual visit to the university where the research was conducted with frequent breaks in between activities. Each child wrote three essays, each representing a different genre of expository writing, with topic held constant (Mt. Rainier and Mt. St. Helens). All essays were written on the same day, with brief breaks in between, in a constant order, as is common in longitudinal psychoeducational assessment research. To control for potential differences in background knowledge, before each essay, the researcher gave each participant paper copies of background information about the mountains, and then read that information out loud as the student read along silently. Essays were written in the following constant order by all participating typically developing writers: an informative essay discussing the seasonal changes on Mt. Rainier, an essay comparing and contrasting Mt. St. Helens and Mt. Rainier, and a persuasive essay, stating and defending an opinion on some controversies about the mountains. Participants were given 5 min to write each composition, and if they stopped writing before 5 min elapsed, they could be prompted up to twice to keep writing. Specific instructions were as follows:

1. Please read silently while I read aloud this text about Mt. Rainier, which has the title, "The Many Seasons at Mt. Rainier National Park." Now look at these post cards that depict the different seasons at the mountain and how the mountain seems to change with the seasons. Now please write an informative essay that describes Mt. Rainier which has the title, "The Changing/Changeless Mt. Rainier," so that someone who has never visited the mountain can visualize what it looks like.
2. Please read silently while I read aloud this text that contains many facts about both Mt. St. Helens and Mt. Rainier. Now compare and contrast these mountains. Write a descriptive essay that tells how the mountains are alike and that tells how they are different. (Allow up to 5 min to write.)
3. Now read along silently while I read aloud another text about the controversies regarding these mountains. Controversies mean that different people have different opinions or points of view. Now I want you to write a persuasive essay in which you explain the different points of view about each controversy, give your opinion or point of view about each controversy, and defend your argument and convince the reader against the opposing opinion or point of view.

Measures

Content and organization coding scheme. All essays were coded for the quality of the content and the quality of the organization. A description of that coding, which was developed by the co-authors over a 3-month period based on much reading and rereading of the essays and related discussion, is presented next. This coding was constructed to be comparable across all three expository essay types—informative, compare and contrast, and persuasive. All coding was on a scale of 0 (*low*) to 5 (*high*).

Content quality

1. Content is not relevant.
2. Some content is relevant, but simply repeats facts that were given.

3. All content is relevant, but simply repeats facts that were given.
4. Content is relevant and elaborates on given facts.
5. Content is relevant and sophisticated, and qualifies, elaborates, and uses the information.

Organization quality

1. Ideas are presented in a list and are confusing.
2. Ideas are in sentences, but do not progress logically.
3. Organization is logical, but has little to no framing (e.g., topic sentences).
4. Some framing of ideas, as well as logical progression, is evident.
5. A lot of good framing, as well as logical progression, is evident.

Inter-rater reliability was calculated separately for Grades 5 and 7. For fifth grade, three essays from each of the 10 students were rated separately by both researchers. The initial correlation between those ratings was .66. After discussing all ratings where the researchers differed by two or more points, the coding scheme was updated to take into account text features that both researchers had attended to, as is customary in linguistic coding research. Once ratings were adjusted based on the discussion, the correlation for fifth grade was re-computed to be .89. For seventh grade, the same procedures were used. After coding the 10 essays separately, the correlation between raters was .68. All differences of two or more points were discussed, and changes made in the coding scheme based on those discussions. After the changes from discussion in the coding schemes, the essays were recoded by each rater, as is customary in linguistic coding; the correlation of inter-rater reliability across coding schemes was then re-computed to be .88 for seventh grade. Once inter-rater reliability was established at the standard level in discourse coding (.80 or above) to be .89 for fifth grade and .88 for seventh grade, the first author completed all the coding within a brief time. At the time the data analyses were completed, both coders (the co-authors) reviewed the coding procedures for a randomly selected subsample and found almost complete agreement even though a new coefficient was not computed.

Number of words written. For each essay, the number of complete words was also counted. Misspelled words were counted as words, but neither incomplete nor scratched out words were counted in the totals.

Data Analyses

Pearson product-moment correlations (*r*s) were computed for the content and organization ratings for each genre. ICCs were computed for content and organization ratings and number of words. Repeated measures ANOVA was used to test the differences in means across the three expository genres at each grade level.

Results

Descriptive Statistics and Correlations—Grade 5

Table 1 presents the means, standard deviations, and ICC values for all measures across the three expository essays for Grade 5. Correlations of these measures across genre pairs, per Olinghouse et al. (2012), were computed for Grade 5. For content quality, informative and compare/contrast essays were correlated .65, informative and persuasive essays were correlated .60, and compare/contrast and persuasive essays were correlated .55. For organization quality, informative and compare/contrast essays were correlated .65, informative and persuasive essays

Table 1. Descriptive Statistics for Expository Essays Grade 5.

Measure	Informative		Compare/contrast		Persuasive		ICC
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Content	3.29	(1.15)	3.21	(1.03)	2.97	(0.88)	.84
Organization	3.29	(1.24)	2.98	(1.18)	3.08	(1.04)	.85
Words written	49.42	(20.97)	39.02	(16.70)	55.19	(23.25)	.88

Note. ICC = intra-class correlation, obtained from repeated measures (RM) ANOVA analyses (see Table 3).

Table 2. Descriptive Statistics for Expository Essays Grade 7.

Measure	Informative		Compare/contrast		Persuasive		ICC
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Content	4.06	(0.86)	3.93	(0.94)	3.91	(0.94)	.84
Organization	4.22	(0.87)	4.01	(1.00)	3.90	(0.90)	.80
Words written	70.73	(21.65)	58.29	(19.63)	82.27	(25.04)	.87

Note. ICC = intra-class correlation, obtained from repeated measures (RM) ANOVA analyses (see Table 3).

were correlated .56, and compare/contrast and persuasive essays were correlated .60. All correlations were significant at the $p < .001$ level. Squaring the correlations shows that participants shared only modest variance across the genres. Likewise, the high ICC values for these typically developing writers suggest that much of the variation between the genres was due to participant differences.

Descriptive Statistics and Correlations—Grade 7

Table 2 presents the means, standard deviations, and ICC values for all measures across the three expository essays for Grade 7. Correlations of these measures across genre pairs, per Olinghouse et al. (2012), were also computed for Grade 7. For content quality, informative and compare/contrast essays were correlated .52, informative and persuasive essays were correlated .57, and compare/contrast and persuasive essays were correlated .67. For organization quality, informative and compare/contrast essays were correlated .56, informative and persuasive essays were correlated .56, and compare/contrast and persuasive essays were correlated .55. All correlations were significant at the $p < .001$ level. Squaring the correlations shows that participants shared only modest variance across the genres. Likewise, the high ICC values for these typically developing writers suggest that much of the variation between the genres was due to participant differences.

Comparing Three Genres for Fifth Graders on Multiple Outcome Measures

Repeated measures ANOVAs were performed for 110 fifth graders to examine mean differences between essay genres for each of three outcome measures: rating of content quality, rating of organization quality, and number of words written. Where Mauchley's test was significant, indicating a violation of sphericity, Greenhouse-Geisser-adjusted F tests are reported. These ANOVA results, which are reported in Table 3, demonstrate that at least one of the essay types significantly differed from the others for each outcome measure. Follow-up pairwise t tests were conducted using a Dunn-Sidak adjustment to see which essay types were significantly different from each other. Each model will be discussed in turn.

Table 3. Descriptives and Results for RM ANOVA Grade 5.

Outcome measure	<i>M</i>	<i>(SD)</i>	<i>F</i> test			ω^2
			<i>F</i>	<i>(df)</i>	<i>p</i>	
Content rating			7.02	(2, 218)	.001	.05
Informative	3.29	(1.15)				
Compare/contrast	3.21	(1.03)				
Persuasive	2.97	(0.88)				
Organization rating			5.09	(2, 218)	.007	.04
Informative	3.29	(1.24)				
Compare/contrast	2.98	(1.18)				
Persuasive	3.08	(1.04)				
Number of words			57.11	(1.86, 203.03)	<.001	.34
Informative	49.42	(20.97)				
Compare/contrast	39.02	(16.70)				
Persuasive	55.19	(23.25)				

Note. *df* = degree of freedom.

Content quality ratings. Follow-up tests found a significant decrease in content ratings between the compare/contrast essay and the persuasive essay, as well as between the informative essay and persuasive essay ($p = .024$ and $p = .002$, respectively). Consistently, content ratings were lower for persuasive essays even though they were written last after more practice in writing about the mountains. Trend contrasts for essay type showed that the change in content scores had significant linear components (as illustrated by the mean values in Table 1), with the contrast *F* test p value $< .01$. As expected, participants varied significantly from each other on content ratings, $Var = 2.31$, $F(1, 109) = 1420.48$, $p < .001$, $ICC = .84$.

Organization quality ratings. Follow-up *t* tests only found a significant difference in scores between the informative essay and the compare/contrast essay ($p = .006$). Trend contrasts showed that the change in organization scores had significant linear and quadratic components, with contrast *F* test p value $< .05$. Ratings decreased between the informative essay and compare/contrast essays, and then increased between the compare/contrast essays and persuasive essays, as shown by the mean values presented in Table 3. Understandably, organizational demands may be greater for compare/contrast than informative essays. Of interest, even though content ratings were lower for persuasive essays than compare/contrast essays, organizational ratings were higher for persuasive essays than compare/contrast essays. So the lower content ratings for persuasive essays cannot be attributed to their being written at the end of the series of essay genres. As expected, participants varied significantly from each other on organization ratings, $Var = 2.95$, $F(1, 109) = 1087.08$, $p < .001$, $ICC = .85$.

Number of words written (length). Follow-up *t* tests found significant differences in the number of words written between each pair of essays (all p values $< .01$). Trend contrasts showed a significant linear and quadratic effect for number of words written across the essays, with both contrast *F* test p values $< .001$. Although the number of words written decreased between the informative essay to the compare/contrast essay ($M = 10.40$), there was a large increase between the compare/contrast and the persuasive essays ($M = 16.17$), which surpassed the mean number of words written in the informative essay, as shown in Table 3. Of interest, this pattern of results paralleled those for the organizational ratings, but not the content ratings. As expected, participants varied significantly from each other on the number of words written, $Var = 1000.62$, $F(1, 109) = 755.92$, $p < .001$, $ICC = .88$.

Table 4. Descriptives and Results for RM ANOVA Grade 7.

Outcome measure	<i>M</i>	<i>(SD)</i>	<i>F</i> test			
			<i>F</i>	<i>(df)</i>	<i>p</i>	ω^2
Content rating			1.99	(1.94, 186.65)	.141	NA
Informative	4.06	(0.86)				
Compare/contrast	3.93	(0.94)				
Persuasive	3.91	(0.94)				
Organization rating			6.09	(2, 192)	.003	.05
Informative	4.22	(0.87)				
Compare/contrast	4.01	(1.00)				
Persuasive	3.90	(0.90)				
Number of words			81.57	(2, 192)	<.001	.45
Informative	70.73	(21.65)				
Compare/contrast	58.29	(19.63)				
Persuasive	82.27	(25.04)				

Note. *df* = degree of freedom.

Comparing Three Genres for Seventh Graders on Multiple Outcome Measures

Repeated measures ANOVAs were performed for 97 seventh graders to examine mean differences of the three kinds of essays on content quality rating, organization quality rating, and number of words written (length). Greenhouse-Geisser adjusted *F* tests are reported as needed, as in the fifth-grade analysis above. Results for these ANOVAs are reported in Table 4.

Content quality ratings. No significant effect of essay type was found on content ratings at Grade 7, in contrast to what was found at Grade 5. However, participants accounted for significant variance in content scores, $Var = 1.82$, $F(1, 96) = 2520.20$, $p < .001$, $ICC = .84$.

Organization quality ratings. A significant effect was found for essay type on organization ratings, and a trend analysis found significant linear components ($p = .001$). Follow-up *t* tests using a Dunn–Sidak adjustment found that the mean differences were only significant between the informative essays and persuasive essays. In contrast to the fifth graders, the seventh graders showed lower organizational ratings for persuasive than informative essays. As expected, participants accounted for significant variance in organization scores, $Var = 1.71$, $F(1, 9) = 2779.04$, $p < .001$, $ICC = .80$.

Number of words written (length). A significant effect was found for essay type on the number of words written. Trend analysis indicated significant linear and quadratic components to the trend of words written (both p values $< .001$). Follow-up *t* tests with a Dunn–Sidak adjustment found a significant decrease in words written between informative essays and compare/contrast essays ($M = 12.44$), as well as a significant increase between the compare/contrast essays and persuasive essays ($M = 23.98$), and the informative essays and persuasive essays ($M = 11.54$), all p values $< .001$. Although the first two findings replicated those for fifth grade, the third pattern of results occurred only in the seventh grade. As expected, participants accounted for significant variance in the number of words written, $Var = 1139.47$, $F(1, 96) = 1266.77$, $p < .001$, $ICC = .87$.

Discussion

First Tested Hypothesis

The first hypothesis was confirmed. Results for fifth and seventh graders are consistent with findings of Olinghouse et al. (2012) and Boscolo et al. (2012), who have reported evidence for

intra-individual differences across genres in developing writers during middle childhood and early childhood, respectively. These differences occur not only between narrative and expository writing (Olinghouse et al., 2012) and within narrative writing (Boscolo et al., 2012) but also across different genres of expository writing—informative, compare and contrast, and persuasive essays, as shown in the current study for fifth and seventh graders. These findings have important implications for high-stakes assessments of writing, classroom assessments given by teachers, and individual assessments given by psychologists and speech and language specialists because they demonstrate that multiple written compositions representative of different genres must be used to draw conclusions about a student's written composing ability.

Second Tested Hypothesis

For fifth graders, this hypothesis was confirmed. Differences were found among the participants across the genres. However, for seventh graders, the hypothesis was only partly confirmed—for organization ratings but not content, and for length, one of the patterns for contrasts between essay genres for seventh graders was different from the fifth graders, although two were the same. Thus, the transition from upper elementary in fifth grade to middle of middle school in seventh grade in writing multiple essay genres is characterized by some constants and some variables.

Educational Applications

Of educational significance, the current findings show that not all expository essays are created equally—different types of expository essays are likely to produce different results for the same student (intra-individual differences) and different students (inter-individual differences). Specific kinds of individual differences depend, to some extent, on criteria employed to evaluate the quality of composing—content, organization, and length—and grade level. A single annual assessment of a single written composition may not be adequate to conclude whether students meet common core or other high-stakes standards in writing, or to capture an individual student's ability to write different kinds of expository essays, or to create a plan to translate assessment findings into educational practice. Both high-stakes tests and individually administered psychometric tests should assess multiple grade-appropriate genres for determining composition ability/abilities. Both profiles for composing across genres and an overall score based on indicators of writing in multiple genres, similar to that used in assessing scholastic aptitude(s), would contribute to psychoeducational assessment of composing.

Clearly, future translation of research into educational practice should include development of assessment tools for writing that are evidence-based. Annual tests that are based only on a “pass–fail criterion” are not evidence-based. Currently, normed tests exist for assessing various aspects of handwriting, spelling, and sentence-level composing. Regrettably lacking are standardized, psychometric, normed measures of text level composing for a *variety of text genres within the same instrument*, so that relative strengths and weaknesses for specific genres as well as writing ability across genres can be identified with an instrument standardized on the same population. Traditional approaches to reliability of assessment attribute variation in performance to unreliability of the test instrument, but in the case of a process such as written composing, which is inherently generative (Chomsky, 2006), the variation may be fundamental to the process being assessed. Given the generative nature of composing written language, which is fundamentally an open-ended process and challenging to assess in a standardized way, innovative approaches to assess composing for multiple genres and to link assessment findings to writing instruction in developmentally appropriate ways remain to be developed.

Moreover, the writing genres assessed with normed measures should reflect the kinds of composing required for school writing assessments. Although much writing for pleasure may involve narrative writing, narrative composing is not sufficient to complete the kinds of writing and

integrated reading–writing required for successful completion of written assignments at school and for homework during middle childhood and adolescence. Hopefully, prior research on multiple genres reviewed in the introduction and the new findings reported in the current study will contribute to development of such standardized, normed assessment tools for multiple genres that have ecological validity for the kinds of writing students are expected to do at school.

In addition, writing portfolios might be used to collect and periodically review classroom writing assignments across the school year. Both first drafts and revisions can be included in the portfolios as well as extended writing that may be co-constructed with classmates and/or the teacher. Assessment tools might be developed for periodic review of the writing portfolios at specific grade levels. Such periodic reviews should include teachers, students, and parents.

Limitations and Future Directions

The topics chosen for each of the genres were constrained by a desire to keep the topics as constant as possible—the mountains familiar to the participating students in the study—and provide constant background knowledge to all participants before engaging in essay genre composing. Future research might investigate the interaction of topics and genres by comparing multiple genres on a variety of topics common to each. Also, the nature of the longitudinal assessment did not permit linking assessment results with classroom instruction other than sharing assessment results for normed measures (not researcher-designed measures) in an annual assessment report that parents were encouraged to share with schools to use as teachers found useful. Future research should address the dynamic interplay of assessment-instruction links throughout the school year at specific grade levels.

The current research on multiple genres was restricted to two time points in writing development and schooling—middle childhood during the upper elementary grades and early adolescence during middle school. Given both the constants and variations in essay writing observed between upper elementary and middle school, future research should explore development of composing multiple genres from P to 12 and beyond in postsecondary education and the world of work. Moreover, the current study involved a one-time assessment in a larger project, much as the annual testing linked to high-stakes standards involves a one-time assessment in a given school year. Future research on multiple genres of written composition should involve multiple assessments within a grade level and the relationships of assessment findings to ongoing writing instruction in the classroom—both designing it and evaluating responses to it.

Although the current study expanded the study of genres beyond the usual narrative/expository divide to three kinds of expository essays, not all possible or relevant genres, expository or otherwise, were investigated, which hopefully future assessment and instructional research will do. Consistent with the generative nature of writing genres, our research group continues to be amazed at the creativity of the genres observed in developing writers, even for the same written assignment (manuscripts in preparation). Despite the limitations of the current research and need for much future research on multiple genres of composing across development and schooling, the current study does provide converging evidence for the reported findings of Olinghouse and colleagues and Boscolo and colleagues. Hopefully, these and other relevant studies will inspire others—both practitioners and researchers—to approach writing assessment from the perspective of multiple genres in writing given the generativity of language for thought expression.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by HD P50HD071764 and HD25858 from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) at the National Institutes of Health (NIH), USA.

References

- Beck, S. W., & Jeffery, J. V. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing, 12*, 60-79.
- Boscolo, P., Gelati, C., & Galvan, N. (2012). Teaching elementary school students to play with meanings and genre. *Reading & Writing Quarterly, 28*, 29-50.
- Chomsky, N. (2006). *Language and mind* (3rd ed.). New York, NY: Cambridge University Press.
- Epstein-Jannai, M. (2004). Writing to learn: Constructing the concepts of genre in a writing workshop. In G. Rijlaarsdam, H. Van den Bergh, & M. Cousijn (Eds.), *Studies in writing: Vol. 14. Effective learning and teaching of writing* (2nd ed., pp. 349-365). New York, NY: Kluwer Academic Publishers.
- Fayol, M., Alamargot, D., & Berninger, V. (Eds.). (2012). *Translation of thought to written text while composing: Advancing theory, knowledge, methods, and applications*. New York, NY: Psychology Press.
- Gillespie, A., Olinghouse, N. G., & Graham, S. (2013). Fifth-grade students' knowledge about writing process and writing genres. *Elementary School Journal, 113*, 565-585.
- Kamberelis, G. (1999). Genre development and learning: Children stories, science reports, and poems. *Research on the Teaching of English, 33*, 131-135.
- Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing, 16*, 154-169.
- Moore, N. S., & MacArthur, C. A. (2012). Effects of being a reader and observing readers on argumentative writing. In M. Torrance, et al. (Eds.), *Learning to write effectively: Current trends in European research. Studies in writing* (Vol. 25, pp. 157-159). Bingley, UK: Emerald.
- Olinghouse, N. G., & Colwell, R. (2013). Preparing students with learning disabilities for large-scale writing assessments. *Intervention in School and Clinic, 49*, 67-76.
- Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology, 101*, 37-50.
- Olinghouse, N. G., Santangelo, T., & Wilson, J. (2012). Examining the validity of single-occasion, single-genre, holistically scored writing assessments. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices* (pp. 55-82). Leiden, The Netherlands: Brill.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing: An Interdisciplinary Journal, 26*, 45-65.
- Olive, T., Favart, M., Beauvais, C., & Beauvais, L. (2009). Children's cognitive effort and fluency in writing: Effects of genre and of handwriting automatization. *Learning and Instruction, 19*, 299-308.
- Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing, 22*, 1-30.
- Scott, C. (1994). A discourse continuum for school-age students: Impact of modality and genre. In G. Wallach & K. Butler (Eds.), *Language learning disabilities in school-age children and adolescents: Some underlying principles and applications* (2nd ed., pp. 219-252). Columbus, OH: Macmillan.
- Stahl, S., & Nagy, W. (2005). *Teaching word meaning*. Mahwah, NJ: Lawrence Erlbaum.
- Troia, G. A. (Ed.). (2009). *Instruction and assessment for struggling writers: Evidence-based practices*. New York, NY: Guilford Press.
- Troia, G. A., & Olinghouse, N. G. (2013). The Common Core State Standards and evidence-based educational practices: The case of writing. *School Psychology Review, 42*, 343-357.
- van den Bergh, H., De Maeyer, S., van Wiejen, D., & Tillema, M. (2012). Generalizability of text quality scores. In E. Van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices* (pp. 23-32). Leiden, The Netherlands: Brill.