“Im gonna tell you all about it”: Capturing Children’s Voices in Writing Assessment

Regan Humphrey

Department of Psychology
Rhodes College
Memphis, Tennessee

2012

Submitted in partial fulfillment of the requirements for the
Bachelor of Arts degree with Honors in Psychology

This Honors paper by Regan Humphrey has been read

and approved for Honors in Psychology.

Dr. Marsha Walton
Project Advisor

_____

Dr. Janet Panter
Second Reader

_____

Dr. Gail Murray
Extra-Departmental Reader

_____

Dr. Natalie Person
Department Chair

_____

Acknowledgements

First and foremost, I would like to thank Dr. Marsha Walton for her guidance throughout the years of work that went into this Honors project. I am so grateful to Dr. Walton for opening up the world of children's narratives to me, and for sharing her thoughtful advice about every aspect of the research process. Special thanks also go to the members of Dr. Walton's research team – particularly Annika Wuerfel, Jiawen Li, and Kelly Parry – for their insightful feedback and many hours spent coding stories. The members of Dr. Walton's Senior Seminar class, especially Ginny Clowers and Kyle Capstick, were also extremely helpful in the coding process.

I am particularly indebted to Dr. Alice Davidson of Rollins College for sharing her data with me, as well as for her role in the development of the cultural voice measure and her advice on the data analysis process. Additional thanks go to Caitlin Campbell from Rollins, who also helped to develop the measure of cultural voice. For his help with the data analysis and the intricacies of SPSS, I am very grateful to Dr. Chris Wetzel. I would also like to acknowledge Dr. Linda Mason and Dr. Elizabeth Benedek-Wood, of Pennsylvania State University, for welcoming me into their lab and sharing their knowledge of writing assessment with me, and the Hyde Foundation for supporting my trip there.

Finally, I am very grateful to the children and teachers who participated in our study. The students shared themselves so willingly with us, and their stories are incredibly compelling. It has been a privilege to spend so much time with the words of these young authors.

CONTENTS

List of Tables

Abstract

"Im gonna tell you all about it": Capturing Children's Voices in Writing Assessment
by
Regan Humphrey

In today's educational system, standardized tests play a more important role than ever before. Increasingly, these tests include direct assessments of writing skills; however, there is controversy regarding which facets of writing should be included in these assessments. A particular target of debate is the quality of authorial voice, a unique, personal style of writing. This study explores the relationship between authorial voice and the mechanics of writing, as well as academic performance. Participants – 180 third, fourth, and fifth graders – wrote stories about peer conflict in the fall and spring of one academic year. Stories were coded for two measures of mechanics (conventions and clarity) and four measures of authorial voice (expressiveness, metanarrative comments, cultural voice, and emphasis markers). Students and their teachers also completed the following measures of attitudes and behaviors related to academic performance in the fall and spring: academic self-concept; peer-nominated academic reputation; and teacher-rated academic skills, academic effort, writing ability, and attention problems. The results showed a moderate relationship between authorial voice and mechanics in the fall, but this relationship weakened in the spring. Of all the writing skills, conventions was the strongest predictor of academic performance, particularly for the teacher ratings, although the peer ratings were more sensitive to authorial voice. I suggest that the neglect of authorial voice stems from mechanics-focused state standards, and propose an alternative writing curriculum that uses cultural voice as a springboard to develop other writing skills.

"Im gonna tell you all about it": Capturing Children's Voices in Writing Assessment

What makes a good writer? Is it a mastery of grammar, facility with language, creativity, or some other quality? What is the best way to measure writing skills as they develop? These questions have become increasingly important in today's educational environment. With the implementation of No Child Left Behind, standardized tests play a more prominent role than ever. On an individual level, they may affect children's academic outcomes, in some cases influencing decisions about promotion to the next grade or graduation. On a broader level, they influence the selection of administrators and the services that are provided to students, such as tutoring or the option to transfer to a different school (No Child Left Behind Act, 2001). Furthermore, an increasing number of school districts rely on student test scores when making decisions about the retention and compensation of teachers. Many state-mandated standardized tests now include assessments of writing skills. Given the powerful influence of these tests, it is crucial they implement scoring procedures that provide a complete and accurate picture of students' writing skills.

There is controversy in the field regarding which aspects of writing must be addressed in order to provide such a complete picture (Mullis, 1984). In particular, there is disagreement about whether or not authorial voice should be taken into consideration. Authorial voice refers to a distinct writing style unique to a particular author; it has largely been neglected in previous research and is omitted from many writing assessment procedures, despite evidence suggesting it may be a critical facet of quality writing.

In this paper, I will examine various facets of writing included in assessments of writing skills and their relationship to academic performance. I begin by addressing the

role of writing skills, particularly in the genre of narrative writing, in children's academic and social success to demonstrate the importance of developing sensitive and accurate measures to assess these skills. Next I will consider the diverse methods of writing assessment implemented in both educational and research contexts. I will then look more closely at the construct of authorial voice and demonstrate the need for further research investigating the relationship between authorial voice and other facets of writing, as well as academic performance.

The Importance of Writing Skills in Middle Childhood

Writing skills exert tremendous influence on children's academic and social success. Writing plays a major part in the typical school environment, both in how children learn material and how they demonstrate what they have learned. Children whose writing skills are lacking usually struggle not only in language arts classes, but also in many other content areas (Antonacci & O'Callaghan, 2011; Mason, Benedek-Wood, & Valasa, 2009). For example, children are frequently required to submit written reports in science and social studies classes. In addition, even brief writing assignments in foreign language classes can be an immense challenge for students who are not confident writing in their native language. Students may become frustrated and discouraged if they feel they understand the material but cannot communicate their knowledge, decreasing their motivation for academic work. In addition, research has shown that literacy and writing skills are associated with social skills from a very early age: children who perform well on writing tasks and tests of early reading skills tend to display greater social competence and fewer problem behaviors (Benner, Beaudoin, Kinder, & Mooney, 2005; Dray, Selman, & Schultz, 2009).

The narrative genre, in particular, seems to hold a special importance in many areas of child development. Bruner (1990) argued that humans have an innate predisposition towards the narrative form, and that the push to produce narratives drives language acquisition. He contended that narrative is the key medium through which children make sense of the world around them. Paley (2004) supported these views, claiming, "Even before the spoken word, the pictures in the young child's mind assume a storylike quality" (p. 14). Children, she says, are natural storytellers, constantly rattling off a seemingly endless stream of narratives. These narratives can be powerful tools.

The same skills that contribute to coherent and engaging narrative writing – such as taking another person's perspective, making moral evaluations, and explaining one's own motives and goals – can facilitate conflict resolution and other interpersonal skills (Harris & Walton, 2009). In the process of sharing stories, children form social and emotional connections with one another (Engel, 1995; Paley, 1990). In fact, Bruner (1990) claimed that the sharing of stories is crucial for social and cultural stability: "[…] one of the principal forms of peacekeeping is the human gift for presenting, dramatizing, and explicating the mitigating circumstances surrounding conflict-threatening breaches in the ordinariness of life" (p. 95). Walton & Brewer (2002), building on Bruner's theory, contended that narrative plays a critical role in bringing children into the moral discourse of their culture.

On a more individual level, Engel (1995) posited that narrative is critical to the process of identity development. She wrote, "[…] who we are now is interlaced with the who we might be, the who we would like to be, the who we are afraid of being. Our possible selves contribute to our actual selves, and we construct those possible selves

through the stories we tell" (Engel, 1995, p. 187). Similarly, Bruner (1986) argued that the language of narrative is unique in its emphasis on subjunctivity, which allows for the creation of infinite "possible worlds" through storytelling. Finally, the narrative genre is pervasive in elementary school language arts classes, frequently occurring in both reading and writing assignments (McCabe, 1997). Consequently, skill in the narrative form can have a major impact on children's academic success.

Clearly, children's writing skills, especially in the genre of narrative, are of tremendous importance. Children with strong narrative writing skills are equipped for success in many academic fields, as well as the social arena and even in the development of their own identities. Conversely, children whose narrative writing skills are lacking will likely face challenges in all of these areas. For this reason, it is crucial that educators be able to assess students' narrative writing skills with accuracy, using measures that create a complete picture of students' strengths and weaknesses in this area.

Methods of Writing Assessment

Methods of assessing children's writing skills, both for research purposes and in the classroom, vary widely. To begin with, there are disagreements about the value of direct versus indirect assessments of writing. Direct writing assessments evaluate writing skills by examining actual writing samples; indirect assessments use other means, most commonly multiple choice questions about grammar, punctuation, and vocabulary (Heck & Crislip, 2001; Huot, 1990). Each method has weaknesses. Although direct assessments are much more difficult to score reliably, the validity of indirect assessments is questionable. Performance on multiple choice tests may bear little relationship to actual writing skills. There is also evidence that direct assessments are less likely to

result in differential validity according to socioeconomic status or ethnic group: Heck & Crislip (2001) found significantly greater disparities in performance among both socioeconomic groups and ethnic groups on indirect writing assessments than on direct assessments. Given these issues of validity, it seems the benefits of direct writing assessment may outweigh its scoring challenges. Consequently, direct assessment of writing has become increasingly popular in recent years.

This popularity has produced further debate about the best way to score direct assessments of children's writing. A variety of scoring techniques have emerged, focusing on many different aspects of writing. For the most part, these techniques can be grouped into three major methods: holistic scoring, primary trait scoring, and analytic scoring (also sometimes referred to as multiple-dimension ratings or trait-based scoring; Mullis, 1984). Holistic scoring is based on the theory that, when it comes to successful writing, the whole is greater than the sum of its parts. It also makes the assumption that experts know good writing when they see it. Thus this system simply requires raters to select the score that represents the overall quality of a piece of writing. This score may be based on a rubric that encourages raters to consider facets such as the organization of the piece, the content, or the correct use of grammar and punctuation. However, these traits are not rated independently; each piece of writing receives only one holistic score. Furthermore, there are usually no specific instructions about weighting the various traits mentioned in the rubric. It is up to each rater to determine how much each trait influences the overall quality of the writing. This can be problematic, as research has indicated that some traits influence holistic scores more than others. For example, content and organization have been shown to exert more influence over scores than

sentence structure or mechanics (Huot, 1990). Despite these concerns, holistic scoring procedures remain popular. In part because this method takes very little time, it is used in many standardized tests, including the Tennessee Comprehensive Assessment Program Writing Assessment (Tennessee Department of Education, n.d.)

Primary trait scoring also typically produces one score per writing sample. However, it attempts to address some of the concerns about holistic scoring by reducing the facets of writing taken into consideration (Mullis, 1984). In this method, each piece of writing is assigned a score based on how well it accomplishes its purpose. For example, when scoring a persuasive essay the ratings might be based on whether or not the author presented a convincing argument. This narrow focus alleviates one of the aforementioned concerns surrounding holistic scoring – the finding that certain aspects of writing, such as organization, have a greater influence on the score than other aspects that are also supposedly included in the assessment. In doing so, however, it dramatically decreases the number of skills that are assessed. Primary trait scoring produces a greatly simplified picture of students' writing skills, which may limit its usefulness.

Analytic scoring systems attempt to deliver a more well-rounded assessment of students' writing skills by assigning each writing sample multiple scores corresponding to a number of traits. A variety of scoring systems have been developed within this category, each addressing different facets of writing. Some provide a count of desired elements, usually specific to a particular genre (Lane et al., 2008; Lane et al., 2010; Mason, Kubina, & Taft, 2011; Novak, Herman, & Gearhart, 1996). For narratives, these elements might include establishing the setting or characters. For persuasive essays, the desired elements might include presenting evidence to support one's position or refuting

a counter-argument. These counts are not typically used alone, but rather as a supplement to holistic ratings. Scoring systems of this type are often used in research settings. They can be particularly useful in evaluating the effectiveness of a writing skills intervention, as they can help to track improvement in specific areas targeted by the intervention (Harris, Graham, & Mason, 2006; Mason, Kubina, Valasa, & Cramer, 2010; Mason & Shriner, 2008).

Other analytic rubrics focus solely on the mechanical aspects of writing. They may provide counts of correctly spelled words, correct punctuation marks, and so on, or they may use ratings of traits like grammar, semantics, and spelling (Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Wakely, Hooper, de Kruif, & Swartz, 2006). These rubrics – particularly those that employ counts rather than ratings – are typically easier to score reliably than other analytical scoring systems, as there is little subjectivity involved. Research has indicated that they correlate moderately with indirect writing assessments (Gansle et al., 2006). However, some question the validity of these mechanics-centered rubrics because of their narrow focus; though they may provide a variety of measures of mechanical correctness, they ignore a number of other important features of writing.

The analytical rubrics that are broadest in scope assess numerous facets of the writing process, often rating each writing sample on five or six different traits (Culham, 2003; Educational Records Bureau, 2006; Heck & Crislip, 2001). These might include, for example, ideas, organization, or word choice. By examining these traits separately, analytic systems aim to provide a clearer picture of students' strengths and weaknesses. Indeed, research indicates that multi-dimensional ratings do in fact assess a number of

different constructs and can provide valuable information that is lost in holistic ratings. Roid (1994) analyzed writing samples from over 10,000 students using a multi-dimensional rubric. Using a cluster analysis, she found that less than half of the students displayed patterns of all high or all low scores; the rest showed more complex patterns, with high scores on some traits and low scores on others. Thus, for 60% of the students, a holistic rating would have provided an incomplete picture of their writing skills.

Because they provide so much information, analytic rubrics of this sort are used in a variety of contexts. They are used to score direct writing assessments for research purposes, as described above, and on some standardized tests (ERB, 2006; Heck & Crislip, 2001). In addition, they are often implemented by teachers to grade regular writing assignments (Antonacci & O'Callaghan, 2011; National Council of Teachers of English, 2004). For example, the popular 6 + 1 TRAIT model, developed by the Northwest Regional Educational Laboratory (now called Education Northwest), is geared towards classroom teachers. It includes not only a multidimensional rubric for assessing student writing, but also suggestions for teaching each of the seven traits. This model contends that analytic writing assessment has benefits beyond the breadth of information it supplies to teachers: it increases the consistency of assessment across students, helps to ensure that students receive detailed and specific feedback, allows teachers to target instruction to align with students' strengths and weaknesses, and gives students a common vocabulary to talk about the qualities that contribute to good writing (Culham, 2003). As a result of these benefits, analytic rubrics are increasingly popular, though holistic and primary trait scoring continue to have their own proponents.

Authorial Voice

There is considerable disagreement about whether or not analytical rubrics should include authorial voice among the traits they assess. As mentioned previously, authorial voice refers to a distinct narrative style unique to a certain author, or the ability to tell an engaging story with a sense of individuality. This can be a difficult concept to operationalize. The 6 + 1 TRAIT model attempts to do so, defining voice as "[…] the writer's music coming out through the words, the sense that a real person is speaking to you and cares about the message" (Culham, 2003, p. 102). The description continues, "When the writer is engaged personally with the topic, he or she imparts a personal flavor to the piece that is unmistakably his or hers alone. It is that individual something – different from the mark of all other writers – that we call 'voice'" (Culham, 2003, p. 102). In this model, voice is rated on a five-point scale ranging from "The writer seems indifferent, uninvolved, or distanced from the topic and/or the audience" to "The writer speaks directly to the reader in a way that is individual, compelling, and engaging" (Culham, 2003, p. 109). The 6 + 1 TRAIT model, however, is distinctly in the minority in this respect. Most writing assessment procedures in both research and educational contexts do not include voice at all, largely because it has proved challenging to score – perhaps as a result of the difficulty in producing a concrete definition of the construct. Indeed, some of the few rubrics that do include voice use a much narrower description of it; for example, in a rubric published by the National Council of Teachers of English (2004), voice is equated with "point of view" and focuses primarily on a "sense of audience."

Despite the challenges associated with measuring authorial voice, theory suggests that it is a very important construct in the context of development. It requires significant facility with language, allowing the author to employ it in a creative and evocative manner. Indeed, a strong authorial voice is one of the hallmarks of great writers (Engel, 1995). Beyond that, however, authorial voice requires significant psychological sophistication. The author must be able to modify his or her writing to suit the audience at hand, and to present a clear, compelling perspective on events with which the audience has no prior familiarity. He or she must possess the skill to craft exciting and engaging language while still including all of the necessary facts for the audience to understand what happened. All of this requires considerable metacognitive skill (Gombert, 1992).

Given this developmental significance of authorial voice, it is worth examining in more detail. There are many aspects of writing that could potentially influence authorial voice. Perhaps the broadest category is expressive language. Focusing more closely on the link between metacognitive skill and authorial voice, one might also consider the quality of metanarrative awareness. Another way in which an author can imbue his or her writing with a personal tone is by using language that reflects his or her cultural community. Finally, on a very local level, orthographic features and intensifiers can be used to add to the authorial voice of a piece of writing. I will consider each of these possible contributors to authorial voice in turn, then I will address the overall developmental trajectory of authorial voice.

Expressive Language. Expressive language is language that conveys meaning in a vivid, forceful manner. There are a variety of components that make language more expressive. One component is the use of figurative language devices, such as hyperbole,

irony, metaphor, and simile. Their use makes writing more interesting in many ways, from emphasizing certain points to adding humor (Roberts & Kreuz, 1994). In fact, any use of humor generally adds to expressiveness. The strategic use of reported speech, particularly dialogue, also contributes (Walton & Walton-Wetzel, 2008). Careful word choice, especially the use of advanced or unusual vocabulary, further increases the expressiveness of a piece of writing. All of these elements make language more interesting and engaging, thus enhancing authorial voice.

Metanarrative Awareness. Writers with a strong authorial voice are often conscious of and even explicitly remark upon the reader or the storytelling process. This is referred to as metanarrative awareness. As mentioned above, such cognizance of the writing process demonstrates considerable metacognitive sophistication (Gombert, 1992). Previous research has also indicated that metanarrative awareness in children's writing is linked to ratings of the perceived "literariness" of their stories, further supporting the idea that metanarrative awareness contributes to authorial voice (Walton & Walton-Wetzel, 2008).

Language of Cultural Communities. Language is one of the key ways in which people demonstrate membership in particular sociocultural groups (Bucholtz & Hall, 2004). A number of different aspects of language can be used for this purpose, but two of the most common are colloquialisms and idiomatic expressions. These devices contribute to engaging writing, helping to convey the person behind the story to the reader and adding to a unique writing style (Roberts & Kreuz, 1994). Children's mastery of these forms has implications even beyond their writing skills. Familiarity with certain slang terms can serve to associate a person with a particular peer group, which has a

considerable influence on social status. This is especially true of young people, who often have their own distinctive and rapidly evolving slang vocabularies (Labov, 1992). In addition, research suggests that children's comprehension of non-decomposable idioms (in which one cannot infer the meaning of the idiom by examining its parts) is associated with the development of theory of mind, a construct which has a variety of significant cognitive and social implications (Caillies & Le Sourn-Bissaoui, 2008). Research has also indicated that children who use more cultural language in their writing are more attentive to others' thoughts and emotions and more attuned to moral issues (Davidson, Walton, Humphrey, & Campbell, 2012). Finally, by identifying themselves with a certain cultural group through language, children are engaging in identity construction (Bamberg, 2004; Bucholtz & Hall, 2004). All of this evidence suggests that the ability to appropriate the language of one's cultural community is an important contributor to authorial voice as well as to numerous other aspects of development.

Despite its significance, no other researchers have attempted to measure this construct – which I have termed cultural voice – in children's narrative writing. In fact, little research has been done on children's production of idioms or slang terms at all, as previous studies have focused primarily on their comprehension of these forms. In part, this lack of research can be attributed to the difficulty in determining what qualifies as an idiom. Typically, when people think of idioms, they think of clausal idioms, such as "kick the bucket" or "spill the beans" (Kerbel & Grunwell, 1997). However, closer examination reveals that far more of the English language is idiomatic than one might initially realize. In addition to clausal idioms, there are also phrasal verb idioms, like "carry on," plus adjectival and adverbial phrase idioms, such as "ever so" and "straight

away" (Kerbel & Grunwell, 1997). Taking all of these forms into consideration, Kerbel and Grunwell (1997) found that elementary school teachers used an average of 1.73 idioms per minute, far more than the teachers themselves predicted. Lakoff and Johnson (1980) went so far as to argue that metaphor is a key part of the human conceptual system. They contend that a variety of "conventional metaphors" – such as "argument is war" and "ideas are objects" – structure our understanding of the world. Given this pervasiveness of figurative language, it can be challenging to determine which instances of idiom demonstrate a true integration into one's cultural community and the ability to appropriate the language of that community for one's own purposes. In the present study, I have endeavored to create a measure of cultural voice that achieves this goal.

Emphasis Markers. On a more "micro" level, writers also use certain textual features to emphasize particular parts of the story. One such feature is the intensifier: a word or phrase that expresses extremity or exceptionality, such as "I get *way too* mad." Intensifiers can either increase or decrease the intensity of a statement. Orthographic devices can also be used to achieve a similar purpose. For instance, exclamation points, repeated words, and capital letters are frequently employed to add emphasis. These emphasis markers serve authorial voice in a number of ways. They can make writing more engaging and interesting, and give readers the sense that they can hear the author telling the story. They also help the author to guide the reader's experience, drawing the reader's attention to certain parts of the story.

Development of Authorial Voice. Bruner (1990) argued that there are four key qualities of narrative to which humans are innately predisposed: agentivity, sequentiality, canonicality, and, most importantly for the present study, voice. From their earliest

exposure to narratives, he contended, children are aware that every story must be told from someone's perspective, and that perspective comes through in narrative voice. Similarly, Engel (1995) pointed out that most young children display a strong voice in oral storytelling, even though their narrative skills may be lacking in other areas. However, as children become more adept at telling detailed, sequential stories that present a clear picture of an event, they may tend to lose some of the vibrant voice that once colored their narratives.

The reasons for this change are unclear. Some have posited that schooling, and particularly writing instruction, could cause such a decline (Fitzgerald & Teasley, 1986). As children are pushed to conform to the structural and linguistic conventions of schooled narratives, their unique authorial voices may be suppressed (Fang, 2001). At the very least, research suggests that traditional writing instruction does not strengthen authorial voice. Fitzgerald and Teasley (1986) found that instruction on narrative structure improved the overall quality and organization of children's writing, but had no impact on creativity (based on subscales for uniqueness, idea production, language use, and originality, giving it many similarities to authorial voice). Clearly, despite the fact that it is a trademark characteristic of good writing, authorial voice is seriously neglected in our current educational system.

Nevertheless, research suggests that it is possible to reverse the downward trend in authorial voice during middle childhood. Harris (2009) conducted a study on the effects of KidsTalk, a story-sharing program at an urban community center. This program gave children the opportunity to tell stories about their personal experiences to their peers on a regular basis, without the constraints that might be placed on their

narratives in a school setting. The results indicated that the expression of authorial voice in the stories increased over the course of the program, and that children with high participation rates showed a strengthening of voice in their stories over time. Stagg (2007) also investigated the effects of KidsTalk. She found that, over the course of nine months, the children's stories displayed increasing psychological sophistication. Moreover, this was true not only for children who had participated in the program since its inception, but also for children who entered the program in the middle. This finding provided evidence for the development of a strong narrative culture beyond the KidsTalk sessions. These two studies indicate that children's authorial voices will grow stronger in an environment that celebrates storytelling and provides ample opportunities to share narratives of all kinds, not just schooled narratives. Further support for this idea can be found in the writings of master teacher Vivian Paley, who has long advocated for classrooms centered on narrative practice and placing a high value on each child's unique authorial voice (Paley, 1990, 1997, & 2004).

The Present Study

The first aim of the present study is to develop techniques for assessing authorial voice in children's narrative writing. I then use these techniques to determine the relationship between voice and other, more commonly used measures of writing skills. That is, I attempt to determine whether there is a correlation between proficiency in the mechanics of writing and authorial voice. I also examine the relationships between both types of measures and students' academic performance. Specifically, do mechanical proficiency and authorial voice differentially predict academic performance, as rated by students' peers and teachers?

Previous research demonstrated a link between mechanical skill and standardized measures of academic achievement, as well as metacognition, self-efficacy, and self-regulation with regards to writing (Gansle et al., 2006; Wakely, Hooper, de Kruif, & Swartz, 2006). Consequently, I predict that mechanical proficiency will be a positive predictor of academic performance. As discussed above, little research has been conducted on voice, perhaps because it is not included in many direct writing assessments. Nevertheless, given its developmental significance as well as the facility with written language required to achieve a strong authorial voice, I hypothesize it will also predict academic performance.

Methods

Participants

Participants included 180 third, fourth, and fifth graders at a public elementary school in the southern United States. Of these participants, 42.5% were girls and 55.8% were boys (for three participants, data on sex were missing). The sample was ethnically diverse, as 36.5% of participants were African American, 31.5% were European American, 18.8% were Hispanic, 2.8% were Asian, 0.6% were Indian, and 7.7% were multi-racial. The students were primarily of low socioeconomic status, with 82% of children at the school qualifying for free or reduced lunch (Florida Department of Education, 2009). Data were collected in the fall (N = 91) and spring (N = 100) of one academic year. Due to absences and mid-year transfers, some students participated only in the fall or the spring, while others participated in both semesters.

Procedures

The parents of all student participants signed informed consent forms. A decision

was made to obtain the students' assent as well, in order to explain the purposes of the study to them and give them the option to decline participation; thus students completed assent forms. Teachers also gave informed consent for their own participation. Researchers visited each classroom twice in the fall and twice in the spring to collect data. During the first visit of each semester, children wrote personal narratives about their experiences with peer conflict in response to the prompt "Write a story about a conflict with a classmate that really happened to you. Think about an argument, a fight, or some kind of misunderstanding. Write everything you can remember about what happened from the beginning to the end." The children were given 30 minutes to write. After these narratives were collected, they were transcribed and segmented into subject-predicate units (see Appendix A for transcription and segmentation procedures). All identifying information was removed from the transcribed stories, and names were changed to pseudonyms to preserve confidentiality.

During the second classroom visit of each semester, participants completed a number of survey measures regarding their own and others' social-emotional and academic behavior. Some of these measures required students to nominate peers who possessed certain characteristics; for these items, students were provided with a roster listing the names of all of the students in their class and were asked to list as many or as few of their classmates as they wished. Scores on these measures represent the number of nominations received, which can range from zero to the total number of students in a given participant's class. Teachers also answered survey questions about participants' social and academic behavior each semester. After the surveys were completed, students' names were removed and replaced with identification numbers.

Measures

Social Health Profile. For each student, teachers completed a series of items from the Social Health Profile, including measures of academic skills, academic effort, and attention problems (Conduct Problems Prevention Research Group, 1999). For all of the items in these measures, teachers assigned a rating on a five-point scale ranging from "strongly disagree" to "strongly agree." The measure of academic skills was calculated from four questions about proficiency in different content areas: "good at math," "good at reading," "good at science," and "good at writing." These items formed a reliable scale (Fall Cronbach's $\alpha = .90$; Spring Cronbach's $\alpha = .87$). See Table 1 for the means and standard deviations of all measures of academic performance in the fall and spring. I also examined the "good at writing" item individually, which I refer to as writing ability. Academic effort was calculated from four items: "works hard at school," "shows poor effort" (reverse scored), "completes assignments," and "does the best he/she can at schoolwork." These items formed a reliable scale (Fall Cronbach's $\alpha = .85$; Spring Cronbach's $\alpha = .89$). The measure of attention problems was calculated from five items: "easily distracted," "does not pay attention," "stays on task" (reverse scored), "thinks before acting" (reverse scored), and "concentrates" (reverse scored). These items formed a reliable scale (Fall Cronbach's $\alpha = .73$; Spring Cronbach's $\alpha = .75$). All of the teacher-reported measures were standardized within each class to control for differences in rating style between teachers.

Peer Academic Reputation. Peer academic reputation was calculated from four peer-nomination items taken from Gest, Domitrovich, and Welsh (2005): "very good at reading," "not very good at reading," "always know the right answer," "never know the

right answer." The negatively worded items were reverse scored. Together, these four items formed a reliable scale (Fall Cronbach's $\alpha = .78$; Spring Cronbach's $\alpha = .80$). To control for differences in class size, these scores were standardized within each class.

Self-Perception Profile for Children. Fourth- and fifth-grade students were also asked to complete a number of items from the Self-Perception Profile for Children, which asks children to select one of two self-statements that applies best to them and then indicate whether that statement is "sort of true" or "really true" of them (Harter, 1982). This measure included the academic self-concept scale, calculated from the following items: "feel they are very good at their school work," "feel like they are just as smart as other kids their age," "almost always figure out the answers," "do very well in their class work". The reliability for this scale was moderate (Fall Cronbach's $\alpha = .69$; Spring Cronbach's $\alpha = .60$).

Writing Skills. Narratives were coded for two measures of mechanics: clarity and conventions. They were also coded for four measures of authorial voice: expressiveness, metanarrative comments, cultural voice, and emphasis markers. All measures of writing skills were coded by teams of undergraduate researchers. These researchers engaged in practice coding using similar narratives from other data sets, then established reliability by independently coding sets of 50 stories from the data used in the present study. Once reliability was achieved (Pearson's $r = .77 – .96$; see Table 2), the data were divided between the coders, who consulted with each other if they encountered difficulties in coding. See Appendix B for the coding manual used for these measures and Appendix C for a more detailed discussion of reliability procedures.

Each narrative was rated on a three-point scale for clarity, which refers to the ease with which the writing can be understood. There are two primary contributors to clarity: sequentiality (clarity of plot structure) and issues of common ground (knowledge shared between the author and reader) (Clark, 1992). A story that is high in clarity is focused and coherent, without irrelevant digressions, and the events in the story follow a logical sequence; further, the author provides all of the information necessary for the reader to comprehend the story. On the other hand, a story that is low in clarity is difficult to follow, and the order of events may seem scrambled. The author may make mistaken assumptions about the reader's knowledge (for example, using pronouns without explaining to whom they refer). Clarity is a key concept in writing assessment and is present in some form in most rubrics, though it is often included in broader constructs such as "ideas" or "organization" (Culham, 2003; ERB, 2006; Gansle et al., 2006; Heck & Crislip, 2001; Novak et al., 1996; Roid, 1994). See Table 2 for inter-rater reliabilities as well as means and standard deviations in the fall and spring for all writing skills measures.

Coders rated narratives on a three-point scale for conventions, which assesses the correct use of conventions of writing, namely grammar, punctuation, capitalization and spelling. A story with a low score on conventions contains numerous errors in these areas, relative to the length and complexity of the story. In a story with a high score, on the other hand, there are few errors, and the author uses these conventions to guide the reader through the text. Measures of conventions, or its component parts, are very widely used in writing assessment (Culham, 2003; Dray et al., 2009; ERB, 2006; Gansle et al., 2006; Novak et al., 1996; Roid, 1994).

Coders rated each narrative on a three-point scale for expressiveness, which refers to an author's ability to tell an engaging story with individual flavor – that is, a distinct writing style unique to that person. A story that is low in expressiveness is straightforward, with minimal evidence of creativity. The author seems indifferent, uninvolved, or distanced from the audience. On the other hand, a story that is high in expressiveness gives the reader a sense of the person behind the words, and readers may feel as though they can hear the child speaking. In addition, it is often marked by creative use of language, such as similes and metaphors (e.g., "He ran at me like a bull, with fury in his eyes") or by the use of vocabulary that is advanced or atypical for the author's grade level (e.g., "I thought I would perish" rather than "I thought I would die").

Each story was coded for the number of lines including metanarrative comments. Metanarrative comments occur when a child shows awareness of the self as author. Children may demonstrate this awareness by making a reference to the reader (e.g., "I hope you like my story that I am going to tell you about," or, "So like I told you he kept on bothering me") or to the process of telling the story (e.g., "Well that's my story and it's all 100% true!").

Narratives were also coded for the number of lines that contained instances of cultural voice – that is, language that identifies a child as a member of a cultural community. This often takes the form of colloquialisms, slang terms, or idiomatic expressions. The present study counted any expressions that fit into the following categories as cultural voice: language that identifies the child as part of a school community (e.g., "I did not get wrote up"); language that identifies the child as part of a regional community, ethnic group or social class (e.g., "I ain't nothing of Chris!");

language that identifies the author with popular culture, street culture, an "in group" or a peer group (e.g., "tryed to do the matrix on my bike"); language that identifies the self in a broad cultural context by quoting literature, scripture, or other cultural authority (e.g., "thought she was the big bad wolf"); and common idioms that are not easily translatable (e.g., "I wone by aland slind"). After the narratives were coded, it became apparent that the first category described above (language that identifies the child as part of a school community) was masking the effects of the other categories of cultural voice. Likely because all of the children were in school and telling stories about school, the amount of school language used was nearly equal to that of all of the other categories combined. Consequently, it was subtracted from the cultural voice scores and analyses were run using the resulting measure.

Finally, narratives were coded for the number of lines that contained emphasis markers. Emphasis markers are literary devices that direct the reader to pay special attention to particular parts of the story, which fall into two categories: intensifiers and orthographic intentional devices. Intensifiers are words or phrases that express extremity or exceptionality. They occur when the author quantifies an adjective or adverb to increase or decrease the intensity of the statement (e.g., "it was **pretty** dumb," or "my eyes get watery because I get **way too** mad"). Orthographic attentional devices are orthographic features that the author uses to emphasize certain points. This includes the use of purposefully misspelled words (e.g., "I got **sooo** made"); exclamation points (e.g., "It is still going on today**!!**"); repeated words (e.g., "he talks about my mother **very very** wrong"); all capital letters (e.g., "**YOUR NOT MY FRIND ANYMORE**!"); and onomatopoeia (e.g., "**boom** as we drop to the flour").

Results

Relationships between Measures of Writing Skills

The first set of analyses addressed measurement concerns. With these analyses, I sought to determine the extent to which my six writing skills measures assessed distinct or overlapping features of children's writing, how they related to one another, and how they changed over time. Below, I report intercorrelations, repeated-measures ANOVAs, principal components factor analyses, and Cronbach's alpha values for the measures of writing skills.

A series of Pearson correlations examined the intercorrelations between the measures of writing skills in both the fall (see Table 3) and the spring (see Table 4). In the fall, there was a positive correlation between clarity and conventions ($r(87) = .417$, $p < .001$); clarity was not significantly correlated with any of the other writing skills measures. Conventions was also positively correlated with expressiveness ($r(87) = .296$, $p = .005$), metanarrative comments ($r(87) = .329$, $p = .002$), and emphasis markers ($r(87) = .334$, $p = .002$); it was not significantly correlated with cultural voice. There was a positive correlation between expressiveness and the metanarrative comments ($r(87) = .447$, $p < .001$), cultural voice ($r(87) = .433$, $p < .001$), and emphasis markers ($r(87) = .367$, $p < .001$). Metanarrative comments was positively correlated with emphasis markers ($r(87) = .357$, $p = .001$); it was not significantly correlated with cultural voice. There was a positive correlation between cultural voice and emphasis markers ($r(87) = .421$, $p < .001$).

In the spring, clarity was positively correlated with conventions ($r(96) = .259$, $p = .012$) and emphasis markers ($r(96) = .242$, $p = .018$); there was a marginally significant

positive correlation between clarity and expressiveness ($r(96) = .202$, $p = .05$). It was not significantly correlated with any of the other writing skills measures. Conventions was not significantly correlated with any of the writing skills measures other than clarity. There was a positive correlation between expressiveness and metanarrative comments ($r(96) = .293$, $p = .004$), cultural voice ($r(96) = .393$, $p < .001$), and emphasis markers ($r(96) = .446$, $p < .001$). Metanarrative comments was positively correlated with cultural voice ($r(96) = .210$, $p = .041$) and emphasis markers ($r(96) = .313$, $p = .002$). There was a positive correlation between cultural voice and emphasis markers ($r(96) = .333$, $p = .001$).

Another series of Pearson correlations examined the stability of the writing skills measures from fall to spring (See Table 5). There was moderate stability for conventions ($r(68) = .510$, $p < .001$) and modest, but significant stability for metanarrative comments ($r(68) = .322$, $p = .007$). There was marginally significant stability for emphasis markers ($r(68) = .231$, $p = .058$). There were non-significant relationships between fall and spring measures of clarity, expressiveness, and cultural voice. A series of repeated-measures ANOVAs showed no significant change from fall to spring on any of the writing skills measures.

Principal components factor analyses (PCA) were used to further investigate the relationships between the writing skills measures. With these analyses, I sought to determine which variables formed a cohesive unit, or factor, based on their correlations with one another. When the six fall writing skills measures were entered into a PCA, the analysis produced two factors (see Table 6 for factor loadings, Eigenvalues, and variance explained for both fall and spring). All six measures loaded on the first factor, while only

clarity and conventions loaded highly on the second factor. When the six spring writing skills measures were entered in a PCA, the analysis produced two factors. All of the measures except conventions loaded onto the first factor, with all of the voice measures loading more highly than the mechanics measures. Only clarity and conventions loaded on the second factor.

To determine if it would be reasonable to create a composite voice measure, Cronbach's alpha was calculated for the three count measures of voice: metanarrative comments, cultural voice, and emphasis markers. The alphas for these measures in the fall and the spring were .581 and .533, respectively. Based on these relatively low alphas, I decided not to combine the voice measures, but rather to examine them separately.

Grade and Sex Differences on Measures of Writing Skills

In these analyses, I examined differences on the writing skills measures according to grade and sex. This information would inform later analyses investigating predictive relationships between writing skills and academic performance. I used a series of MANOVAs to accomplish this. The first analysis was a 3 (Grade) by 2 (Sex of Author) MANOVA with the four fall voice measures as the dependent variables. Although none of the multivariate tests reached significance, there was a univariate main effect for Sex on the use of cultural voice ($F(1, 80) = 6.98$, $p = .01$), showing that girls used significantly more cultural voice ($M = 1.45$) than boys ($M = .61$). There were no significant main effects for sex on expressiveness, metanarrative comments, or emphasis markers.

In addition, there was a marginally significant univariate main effect of Grade on the use of metanarrative comments ($F(2, 80) = 3.08$, $p = .051$). Post-hoc tests using Fisher's LSD indicated that third graders used significantly fewer metanarrative comments ($M = .06$) than did fourth graders ($M = .66$) and fifth graders ($M = .61$); there was not a significant difference between fourth and fifth graders. There were no significant main effects for grade on expressiveness, cultural voice, or emphasis markers.

A 3 (Grade) by 2 (Sex) MANOVA with the two fall mechanics measures as the dependent variables yielded no significant multivariate or univariate effects.

A 3 (Grade) by 2 (Sex) MANOVA with the spring voice measures as the dependent variables revealed no main effects for Sex and no Grade by Sex interactions. There was a significant multivariate Grade main effect, using Wilks' Lambda ($F(8, 170) = 2.85$, $p = .005$). Tests of univariate effects showed a significant main effect for Grade on the use of metanarrative comments ($F(2, 88) = 8.56$, $p < .001$). Post-hoc tests using Fisher's LSD indicated that the third graders ($M = 0$) and the fifth graders ($M = .49$) used significantly fewer metanarrative comments than did the fourth graders ($M = 1.16$); there was not a significant difference between the third graders and the fifth graders.

There was also a significant univariate main effect of Grade on expressiveness ($F(2, 88) = 3.53$, $p = .03$). Post-hoc tests using Fisher's LSD indicated that the fifth graders ($M = 2.26$) scored significantly higher than did the third graders ($M = 1.79$); the fourth graders ($M = 2.11$) did not differ significantly from the other two grades.

A 3 (Grade) by 2 (Sex) MANOVA with the spring mechanics measures as the dependent variables yielded no significant main effects for Sex or Grade. Multivariate tests using Roy's Largest Root revealed a significant Sex by Grade interaction ($F(2, 88) =$

3.84, $p = .03$). Tests of univariate effects showed a significant Sex by Grade interaction on clarity ($F(2, 88) = 3.23$, $p = .04$). Simple effects tests showed that fourth-grade girls scored significantly higher on clarity ($M = 2.88$) than did third-grade girls ($M = 2.27$); fifth-grade girls ($M = 2.53$) did not differ significantly from girls in the other two grades. Among the boys, there were no significant differences between the three grades. There were no significant Sex by Grade interactions on the univariate test for conventions.

Relationships between Measures of Writing Skills and Academic Performance

In the final series of analyses, I investigated the relationship between the measures of writing skills and the measures of academic performance. With these analyses, I hoped to determine whether there was an association between children's writing skills and their self-reported, teacher-rated, and peer-nominated academic performance, as well as whether writing skills significantly predicted academic performance after controlling for grade and sex differences. Below, I report the results of correlations and hierarchical multiple regression analyses to answer these questions.

A series of Pearson correlations were used to examine the relationships between the measures of writing skills and the measures of academic performance in both the fall (see Table 7) and the spring (see Table 8). In the fall, conventions was positively correlated with academic self-concept ($r(59) = .381$, $p = .003$), teacher-rated academic skills ($r(79) = .692$, $p < .001$), teacher-rated academic effort ($r(79) = .265$, $p = .018$), teacher-rated writing ability ($r(79) = .589$, $p < .001$), and peer-nominated academic reputation ($r(87) = .505$, $p < .001$); it was negatively correlated with teacher-rated attention problems ($r(79) = -.322$, $p = .004$). Clarity was positively correlated with teacher-rated academic skills ($r(79) = .412$, $p < .001$), teacher-rated writing ability ($r(79)$

= .378, $p < .001$), and peer nominated academic reputation ($r(87) = .331$, $p = .002$), but not with academic self-concept, teacher-rated academic effort, or teacher-rated attention problems. Emphasis markers was positively correlated with teacher-rated academic skills ($r(79) = .299$, $p = .007$) and peer-nominated academic reputation ($r(87) = .365$, $p = .001$), but not with any of the other academic performance measures. Expressiveness was positively correlated with teacher-rated academic skills ($r(79) = .226$, $p = .045$), but not with any of the other academic performance measures. Cultural voice was negatively correlated with teacher-rated attention problems ($r(79) = -.238$, $p = .036$), but not with any of the other academic performance measures. Metanarrative comments was not significantly correlated with any of the academic performance measures.

In the spring, conventions was positively correlated with teacher-rated academic skills ($r(95) = .483$, $p < .001$), teacher-rated academic effort ($r(95) = .312$, $p = .002$), teacher-rated writing ability ($r(95) = .364$, $p < .001$), and peer-nominated academic reputation ($r(96) = .439$, $p < .001$); it was negatively correlated with teacher-rated attention problems ($r(95) = -.306$, $p = .003$). Conventions was not significantly correlated with academic self-concept. Clarity was positively correlated with teacher-rated academic skills ($r(95) = .259$, $p = .012$), teacher-rated writing ability ($r(95) = .244$, $p = .019$), and peer-nominated academic reputation ($r(96) = .251$, $p = .014$), but not with academic self-concept, teacher-rated academic effort, or teacher-rated attention problems. Emphasis markers was positively correlated with teacher-rated writing ability ($r(95) = .245$, $p = .019$) and peer-nominated academic reputation ($r(96) = .317$, $p = .002$), but not with any of the other academic performance measures. Expressiveness was positively correlated with peer-nominated academic reputation ($r(96) = .379$, $p <$,001), but not with

any of the other academic performance measures. Metanarrative comments and cultural voice were not significantly correlated with any of the academic performance measures.

Writing Skills as Predictors of Academic Performance

Another series of Pearson correlations examined the relationships between writing skills in the fall and academic performance in the spring, to investigate the value of the writing skills measures as a longitudinal predictor of academic performance (see Table 9). Fall conventions was positively correlated with spring academic self-concept ($r(54) = .336$, $p = .013$), teacher-rated academic skills ($r(73) = .619$, $p < .001$), teacher-rated academic effort ($r(73) = .300$, $p = .010$), teacher-rated writing ability ($r(73) = .446$, $p < .001$), and peer-nominated academic reputation ($r(74) = .519$, $p < .001$); it was negatively correlated with teacher-rated attention problems ($r(73) = -.359$, $p = .002$). Fall clarity was positively correlated with spring teacher-rated academic skills ($r(73) = .384$, $p = .001$), teacher-rated writing ability ($r(73) = .246$, $p = .037$), and peer-nominated academic reputation ($r(74) = .251$, $p = .031$), but not with spring academic self-concept, teacher-rated academic effort, or teacher-rated attention problems. Fall emphasis markers was positively correlated with spring teacher-rated academic skills ($r(73) = .243$, $p = .038$) and peer-nominated academic reputation ($r(74) = .297$, $p = .010$), but not with any of the other spring academic performance measures. Fall expressiveness was positively correlated with spring teacher-rated academic skills ($r(73) = .275$, $p = .018$), but not with any of the other spring academic performance measures. Fall metanarrative comments and cultural voice were not significantly correlated with any of the spring academic performance measures.

Concurrent and Longitudinal Predictions of Academic Performance, Accounting for Grade and Sex Differences

I next conducted a series of hierarchical multiple regression analyses to examine the predictive relationships between the writing skills and academic performance measures, controlling for the effects of grade and sex. For each of these analyses, grade and sex were entered in the first block of the regression. The six writing skills measures – conventions, clarity, expressiveness, metanarrative comments, cultural voice, and emphasis markers – were entered in the second block. In the third block, I entered the following interaction terms: grade by sex, grade by conventions, grade by clarity, grade by expressiveness, grade by metanarrative comments, grade by cultural voice, grade by emphasis markers, sex by conventions, sex by clarity, sex by expressiveness, sex by metanarrative comments, sex by cultural voice, and sex by emphasis markers. The first series of regressions examined the fall writing skills measures and fall interaction terms as predictors of the fall academic performance measures. The second series examined the spring writing skills measures and spring interaction terms as predictors of the spring academic performance measures. The third series examined the fall writing skills measures and fall interaction terms as predictors of the spring academic performance measures.

Concurrent Prediction of Academic Performance in the Fall. In the first regression, with fall teacher-rated academic skills as the dependent variable, the first model (which contained grade and sex) was not significant. The second model (which contained grade, sex, and the six fall writing skills measures) was significant ($F(8, 77) = 10.07$, $p < .001$), and it accounted for over half of the variance in academic skills ($R^2 =$

.54; ΔR^2 = .54, p < .001). Only grade (β = -.171, p = .049) and conventions (β = .63, p < .001) were significant predictors of academic skills in this model. The third model (which added the interaction terms) did not significantly add to the variance accounted for.

In the second regression, with fall teacher-rated writing ability as the dependent variable, the first model was not significant. The second model was significant (F(8, 77) = 6.88, p < .001), and it accounted for a little less than half of the variance in writing ability (R^2 = .44; ΔR^2 = .39, p < .001). Only sex (β = -.20, p = .041), clarity (β = .21, p = .042), and conventions (β = .53, p < .001) were significant predictors of writing ability. The third model did not significantly add to the variance accounted for.

In the next regression, with fall peer academic reputation as the dependent variable, only the second model (which contained grade, sex, and the six fall writing skills measures) was significant (F(8, 58) = 3.45, p = .003). These predictors accounted for over one third of the variance in academic reputation (R^2 = .36; ΔR^2 = .35, p = .001). Only conventions was a significant predictor of academic reputation (β = .38, p = .007), although emphasis markers was marginally significant (β = .27, p = .054).

In a regression with fall teacher-rated attention problems as the dependent variable, the writing skills measures did not significantly add to the variance accounted for. Regressions with fall academic self-concept and fall teacher-rated academic effort as the dependent variables yielded no significant results.

Concurrent Prediction of Academic Performance in the Spring. In a regression with spring teacher-rated academic skills as the dependent variable, the first model was not significant. The second model was significant (F(8, 66) = 4.36, p < .001), and it

accounted for over one third of the variance in academic skills (R^2 = .38; ΔR^2 = .36, p < .001). Only conventions was a significant predictor of academic skills (β = .47, p < .001). The third model did not significantly add to the variance accounted for.

In a regression with spring teacher-rated writing ability as the dependent variable, the first model was significant (F(2, 66) = 7.65, p = .001), accounting for just under one fifth of the variance in writing ability (R^2 = .19). The second model was also significant, (F(8, 66) = 6.47, p < .001), accounting for nearly half of the variance in writing ability (R^2 = .47; ΔR^2 = .28, p < .001). In this model, only grade (β = -.46, p < .001), conventions (β = .39, p < .001), and expressiveness (β = .24, p = .034) were significant predictors of writing ability. The third model did not significantly add to the variance accounted for.

In a regression with spring peer-nominated academic reputation as the dependent variable, the first model was not significant. The second model was significant (F(8, 66) = 7.12, p < .001), and it accounted for about half of the variance in academic reputation (R^2 = .50; ΔR^2 = .48, p < .001). Only conventions (β = .38, p < .001) and expressiveness (β = .37, p = .001) were significant predictors of academic reputation.

In a regression with spring teacher-rated attention problems as the dependent variable, the writing skills measures did not significantly add to the variance accounted for. Regressions with spring academic self-concept and spring teacher-rated academic effort as the dependent variables did not yield any significant results.

Fall Writing Skills Predicting Spring Academic Performance. The final series of regressions examined the fall writing skills measures as longitudinal predictors of spring academic performance. In a regression with spring teacher-rated academic skills as the

dependent variable, the first model was not significant. The second model was significant ($F(8, 50) = 4.93$, $p < .001$), and it accounted for nearly half of the variance in academic skills ($R^2 = .48$; $\Delta R^2 = .48$, $p < .001$). Only conventions was a significant predictor of academic skills ($\beta = .54$, $p < .001$). The third model did not significantly add to the variance accounted for.

In a regression with spring teacher-rated attention problems as the dependent variable, only the second model (which contained grade, sex, and the fall writing skills measures) was significant ($F(8, 50) = 2.53$, $p = .02$). These predictors accounted for about one third of the variance in attention problems ratings ($R^2 = .33$; $\Delta R^2 = .24$, $p = .039$). Only sex ($\beta = .43$, $p = .005$) and conventions ($\beta = -.44$, $p = .004$) were significant predictors of attention problems ratings.

In a regression with spring teacher-rated writing ability as the dependent variable, the first model was significant ($F(2, 50) = 7.63$, $p = .001$), accounting for nearly a quarter of the variance in writing ability ($R^2 = .24$). The second model was also significant ($F(8, 50) = 4.79$, $p < .001$), accounting for nearly half of the variance in writing ability ($R^2 = .48$; $\Delta R^2 = .24$, $p = .012$). In this model, only sex ($\beta = -.28$, $p = .032$), grade ($\beta = -.42$, $p = .001$), and conventions ($\beta = .39$, $p = .004$) were significant predictors of writing ability. The third model did not significantly add to the variance accounted for.

In a regression with spring peer-nominated academic reputation as the dependent variable, the first model was not significant. The second model was significant ($F(8, 50) = 4.32$, $p = .001$), accounting for almost half of the variance in academic reputation ($R^2 = .45$; $\Delta R^2 = .42$, $p < .001$). The third model was also significant ($F(8, 50) = 3.58$, $p = .001$), accounting for nearly three quarters of the variance in academic reputation ($R^2 =$

.72; ΔR^2 = .27, p = .041). In this model, only sex (β = -4.37, p = .002), expressiveness (β = 3.45, p = .019), and the grade by expressiveness interaction term (β = -3.87, p = .019) were significant predictors of academic reputation.

Regressions with spring academic self-concept and spring teacher-rated academic effort as the dependent variables did not yield any significant results.

Discussion

Measuring Writing Skills

The first aim of the present study was to develop measures of assessing authorial voice in children's writing and to examine the relationships between those measures of voice and measures of mechanical proficiency in writing. I created four measures of authorial voice – expressiveness, metanarrative comments, cultural voice, and emphasis markers – and compared them with two measures of mechanics – conventions and clarity – at two time periods over the course of an academic year. At both time periods, the measures of authorial voice were moderately but significantly intercorrelated, as were the measures of mechanics. Principal components factor analyses supported a qualitative distinction between the voice and mechanics measures, with an even sharper division between the two in the spring than in the fall. This suggests that they are two different skill sets. However, the relatively weak Cronbach's alpha for the three count measures of authorial voice (metanarrative comments, cultural voice, and emphasis markers) indicates that these measures are assessing distinct, though related, constructs. That is, there are separate skills within the authorial voice skill set.

In the fall, conventions was significantly correlated with three of the four voice measures (expressiveness, metanarrative comments, and emphasis markers); clarity, on

the other hand, was not significantly associated with any of them. In the spring, conventions was no longer significantly related to any of the voice measures, but clarity was modestly correlated with two of them (expressiveness and emphasis markers). As I will discuss in detail below, conventions seems to be the writing skill most valued in school. These results suggest that, in the fall, children with strong conventions were strong writers overall, also scoring highly on clarity and measures of authorial voice. Over the course of the school year, however, conventions became divorced from voice. One possible explanation for this change is that the teachers targeted conventions, with the result that children who were weak in both voice and conventions in the fall improved their knowledge of conventions by the spring, while their use of authorial voice remained low. Another potential explanation is that the children who were the strongest writers in the fall recognized that authorial voice is not valued and thus dropped it from their writing, though they retained their mastery of conventions.

Grade and Sex Differences in Writing Skills

There were relatively few grade and sex differences on the writing skills measures. In the fall, girls used more cultural voice than boys; however, there was no significant difference between the two in the spring, as the girls' use of cultural voice declined and the boys' increased slightly. In the fall, the fourth and fifth graders used more metanarrative comments than did the third graders. In the spring, the third and fifth graders declined in their use of metanarrative comments while the fourth graders increased, with the result that the fourth graders used significantly more than did the students in the other two grades. Similarly, the fourth-grade girls scored higher on clarity than did those in the third or fifth grades, though there were no significant grade

differences among the boys. Finally, in the spring, the fifth graders were more expressive than the third or fourth graders. These grade and sex differences did not form a clear pattern. Consequently, I will refrain from interpreting them until they have been replicated.

Relationship between Writing Skills and Academic Performance

The second aim of this study was to investigate the relationship between writing skills and academic performance. To accomplish this, I compared my six measures of writing skills with six measures of academic performance from multiple sources. These included a self-reported measure of academic self-concept, a peer-nominated measure of academic reputation, and the following teacher-rated scales: academic skills, academic effort, and attention problems. I also examined individually one of the items included in teacher-rated academic skills, in which the teachers were asked to evaluate a student's writing ability. As with the writing skills measures, the academic performance measures were collected twice over the course of one academic year. I hypothesized that both mechanics and authorial voice would be associated with academic performance. This prediction was only partially confirmed, as the relationship between mechanics and academic performance was much stronger than the relationship between authorial voice and academic performance.

I first used a series of correlational analyses to examine the relationships between these measures concurrently in the fall and the spring, as well as longitudinally. Of all of the writing skills measures, in the fall concurrent analysis, conventions was by far the most strongly associated with academic performance, correlating significantly with every academic performance measure. It was followed by clarity, which correlated with half of

the academic performance measures (teacher-rated academic skills, teacher-rated writing ability, and peer-nominated academic reputation). There were relatively few correlations between the authorial voice measures and the academic performance measures. Out of the academic performance measures, teacher-rated academic skills was the most strongly related to writing skills, correlating significantly with conventions, clarity, expressiveness, and emphasis markers. Surprisingly, the component of this scale that directly assessed writing ability only correlated significantly with the mechanics measures. This suggests that, although the teachers recognized children with stronger authorial voices as more academically proficient, they did not identify these qualities as important writing skills.

In the spring concurrent analysis, conventions remained the writing skill most strongly associated with academic performance, though it was no longer significantly correlated with academic self-concept. Clarity again followed, correlating with the same three academic performance measures, and again there were comparatively few correlations between authorial voice and academic performance. Interestingly, at this time period teacher-rated academic skills no longer correlated with any of the voice measures, though it continued to be associated with the mechanics measures. This could be a result of the divorce between mechanics and authorial voice that occurred in the spring, as described above. There was a modest correlation between teacher-rated writing ability and emphasis markers, suggesting that the teachers were not completely ignoring authorial voice in the spring. Nevertheless, at this time period, peer-nominated academic reputation, rather than any of the teacher ratings, was the academic performance measure most strongly related to writing skills. It correlated significantly

with conventions, clarity, expressiveness, and emphasis markers. This suggests that the children continued to value authorial voice and recognize it as a strength throughout the school year, even as it became less associated with teacher perceptions of academic proficiency.

In an analysis examining fall writing skills as a longitudinal predictor of spring academic performance, the pattern of correlations closely resembled that of the fall concurrent analysis. Once again, conventions was significantly correlated with all of the academic performance measures; clarity was significantly associated with teacher-rated academic skills, teacher-rated writing ability, and peer-nominated academic reputation; and there were relatively few significant correlations between authorial voice and academic performance. Moreover, teacher-rated academic skills was once again correlated with expressiveness and emphasis markers in addition to the measures of mechanics, while its writing-specific component question only correlated with mechanics. These results suggest that children who had stronger authorial voices in the fall performed better academically in the spring, despite the fact that voice was not regarded as a key part of writing ability. This supports the idea that children who were strong academic performers in the fall recognized that authorial voice was not valued and used less of it over the course of the school year. Thus these students continued to perform well academically in the spring, when voice was less associated with teacher perceptions of academic performance.

Controlling for Grade and Sex Differences. In the final set of analyses, I used a series of hierarchical multiple regressions to examine the relationship between writing skills and academic performance after controlling for grade and sex differences on the

measures of writing skills. The results of these analyses confirmed that conventions was the most significant predictor of academic performance, particularly for teacher ratings. Conventions was the only one of the writing skills measures that emerged as a significant predictor of teacher-rated academic skills in the fall concurrent, spring concurrent, and longitudinal analyses. It was also the only writing skill that significantly predicted teacher-rating writing ability and attention problems in the longitudinal analyses. Both mechanics measures emerged as significant predictors of teacher-rated writing ability in the fall concurrent analysis. Expressiveness, in addition to conventions, did significantly predict teacher-rated writing ability in the spring concurrent analysis, suggesting that teachers did value authorial voice to some degree. However, it seems to rank far below mechanics from their perspective.

In contrast with the teacher ratings, a measure of authorial voice emerged as a significant predictor of peer-nominated academic reputation in all three analyses. Conventions and emphasis markers significantly predicted academic reputation in the fall concurrent analysis, conventions and expressiveness were significant in the spring concurrent analysis, and expressiveness and the grade by expressiveness interaction term were significant in the longitudinal analysis (although the grade differences were not statistically significant in the fall, the mean expressiveness scores steadily increased with grade). This confirms that the children seemingly placed a higher value on authorial voice than their teachers, and they regarded their peers who had mastered these qualities as more academically skilled than those who had not. It seems that the children's conceptions of academic ability were focused as much or more on compelling storytelling as mechanical correctness.

The Importance of Authorial Voice

Writing Quality. It is not surprising that children should identify authorial voice as a contributor to academic prowess. Despite its weak relationship with teacher ratings of academic performance in the present study, authorial voice is clearly a critical aspect of writing. It is one of the primary characteristics of great writers, giving them a style that is all their own. It is what gives us a sense of having personal relationships with our favorite authors, what makes us want to read everything they have written. To put it plainly, a story with a strong authorial voice is just more fun to read. For example, compare the two stories below.

Example Story 1 (4th-grade girl)

> One day when I was playing inside with my sister and my friend my friend said I did something I didn't really do. When we started to fight my little sister went upstairs to my mom's room to wake her up and tell her what was happening. Soon she came out with my mom and my mom asked what happened so I told her. Finally she asked if I really did what I didn't do so I said "no," and she sent my old friend home. She told her mom what happened and I wasn't alowd to play with her for the rest of the summer.

Example Story 2 (5th-grade girl)

> One time in 4th grade My friend Angel and I, Were sitting at our own private desk doing our work. Now We are like best friends So we both wanted to sit together like every one does. So We asked are teacher Ms. lewis to get a big table and let us sit together. When Ms. lewis finaly got our desk we wilingiy Moved are stuff over to our new desk. After We were all settled and ready and rollen we were

> Just so happy and overjoyed because we got to sit by eachother. That wonderful satisfaction went one for about 2 ½ Months before…… [turn of page] After that 2 ½ months The battle began. Angel and I were just bellowing back & fourth. Then finally Ms. lewis finally got Fed-up with it and Just Moved us to separate sides of the classroom. Well that's My story and it's all a 100% true!

Both of these stories were written in the fall, and both describe adult intervention in a conflict between two friends. The first story is very strong in terms of mechanics – it is high in clarity, and almost completely free of errors in conventions. However, the use of authorial voice is minimal. The second story is quite clear, but it is marked by more errors of conventions, particularly in the area of capitalization. Nevertheless, Story 2 is strong in all four areas of authorial voice. The author uses vivid, expressive language (e.g., "bellowing back & fourth") and emphasis markers (e.g., "so happy") to make the story more engaging. Metanarrative comments (e.g., "Well that's My story and it's all a 100% true!") demonstrate the author's awareness of herself as an author, and cultural voice (e.g., "got Fed-up with it") shows her ability to appropriate the language of her cultural community and use it for her own purposes. Together, these elements create a story that is uniquely personal. Unlike Story 1, Story 2 gives the reader a vivid sense of who the author is as a person. Despite her mechanical errors, the second author comes across as a stronger storyteller because she goes beyond simply reporting what happened and imbues her work with an individual style.

Social Development. In addition to its contributions to writing quality, authorial voice has important implications for children's social development. Indeed, research has shown significant links between authorial voice and peer adjustment in middle childhood.

Humphrey, Walton, and Davidson (2012) found that children who were more expressive in their writing were less likely to be victimized by their peers, and children who showed greater metanarrative awareness reported feeling less lonely and engaged in more prosocial behaviors. That study also showed links between authorial voice and measures of psychological sophistication in narrative writing: children who were more expressive and those who showed greater metanarrative awareness demonstrated higher levels of psychological mindedness (references to thoughts, emotions, or psychological traits), and children with a stronger cultural voice made more references to others' minds and discussed more moral concerns. This suggests that the ability to create a compelling narrative could serve children not only in their writing, but also in conflict resolution and other areas of peer interaction. Taking all of this into consideration, it seems evident that educators should be working to support the development of authorial voice, rather than neglecting it.

Implications for Curriculum Development

Given the importance of authorial voice for multiple spheres of children's development, one might wonder why teachers appear to be largely ignoring it. One possible explanation is today's culture of high-stakes standardized testing. With the immense influence of standardized test scores on all areas of the educational system, teachers are strongly encouraged to adhere to the state-mandated learning objectives to be assessed on the tests. Unfortunately, these standards typically place a heavy emphasis on conventions but rarely include any expectations related to authorial voice. In order to create an educational environment that fosters a strong authorial voice, the standards that

shape our curricula – and the tests used to assess students' mastery of those standards – must be expanded to include the full range of writing skills, including authorial voice.

Indeed, if given the freedom to do so, there are numerous strategies that teachers could employ to encourage the growth of authorial voice. As was described earlier in this paper, numerous researchers have pointed out a decline in authorial voice as children learn to tell more coherent, schooled narratives; however, studies have demonstrated that this trend can be reversed. Harris (2009) and Stagg's (2007) research on the story-sharing program KidsTalk showed that the creation of a narrative culture strengthens authorial voice and other narrative skills. Similarly, Paley (1990, 1997, & 2004) has long documented the positive effects of her own story-sharing program, which creates a classroom culture that prizes narrative and authorial voice. This evidence shows that simply giving children an opportunity to share their stories with one another, and communicating the message that narratives are highly valued, is enough to foster the development of authorial voice.

That being said, I would argue that explicit instruction in the use of authorial voice could be even more effective. By addressing this skill directly and sending the message that it is valuable, teachers could not only improve the quality of their students' writing, but also provide them with a tool that could be useful in their peer interactions. Moreover, I believe that teachers could use children's unique voices as a springboard to build up other narrative skills, including mechanics. Below, I present a sample curriculum framework designed to build on a strong cultural voice, using it as the foundation for a host of other writing skills. I selected cultural voice as a starting point for two reasons. First, it is one of the facets of voice that is most neglected in schools,

and as a result there are very few existing lesson materials designed to address it. Second, it is likely that children who come from communities with rich narrative traditions – though they are often of low socioeconomic status – arrive at school already equipped with a strong cultural voice. Thus, incorporating cultural voice in the writing curriculum is an ideal way to target a group of children who typically struggle with academics.

The Author's Toolkit. The goal of "The Author's Toolkit" is to strengthen children's literacy skills by helping them to identify various tools that authors use to create a clear and compelling narrative and by encouraging the children to use those tools in their own writing. The unit is best suited for children who already have a strong cultural voice – an important authorial tool that is often neglected in formal instruction – and is designed to build on this strength.

The unit will begin with an age-appropriate reading that features strong cultural voice, and the teacher will help students to identify instances of cultural voice in the text. After that, students will brainstorm examples of cultural voice from their own cultures, write a story that features cultural voice, and share their stories with a classmate. In the next part of the unit, the teacher will introduce the idea that cultural voice is just one of many tools that author's use to tell a story. Together, the class will identify a variety of other tools, such as dialogue, plot structure, morals, and so on. Focusing on a few of these tools at a time, the class will repeat the process of identifying the tools in the literature, incorporating them in their own writing, and sharing their writing with peers. To ensure that they retain mastery of the complete toolkit, the teacher will encourage

students to make use of previously taught tools in oral story-sharing. See Appendix D for a more detailed description of "The Author's Toolkit."

Limitations and Future Research

One limitation of this study is that the measures of academic performance all derived from survey data, and thus did not include official grades or test scores. It is likely that grades would closely resemble teacher ratings of academic performance. However, standardized test scores could vary substantially from teacher perceptions. Future research might examine data of this sort to determine if the same relationships with writing skills hold true.

Another potential limitation relates to the cultural groups included in the sample. In the present study, participants were ethnically quite diverse, but the majority of the students were of low socioeconomic status. The relative values placed on authorial voice and mechanics could be different in communities of higher socioeconomic status. For example, in a private school environment that is not so strictly defined by state-mandated learning objectives, teachers might not have the same narrow focus on conventions. Furthermore, attitudes towards cultural voice, in particular, could be different in a school environment dominated by a single ethnic group. For instance, in a school predominantly populated by African American or Hispanic American students, the language traditions of those groups could be more highly valued. For these reasons, future research should explore the relationship between writing skills and academic performance in a variety of distinct cultural environments.

Finally, future research should investigate the effects of curriculum differences on authorial voice. For example, does authorial voice strengthen over the course of the

school year when teachers provide explicit instruction in the use of voice? In contexts where authorial voice is part of the formal curriculum, is it more strongly associated with academic performance? Such research could help shape the development of lesson materials targeting authorial voice, and potentially inform the revision of state standards and assessment practices to include this important facet of writing.

Conclusion

With the increasing popularity of direct writing assessment components on high-stakes standardized tests, it is critical that these writing samples are scored in a way that provides a complete and accurate picture of students' writing skills. Previous research in the area of writing assessment has largely neglected the construct of authorial voice, which is omitted from many scoring rubrics despite indications that it could play a major role in both academic and social development. The results of the present study indicate that authorial voice is also neglected in the classroom, as teachers focus on the mechanics of writing – although peers are more sensitive to the contributions of voice. I suggest that teachers' narrow focus on mechanics may result from state-mandated curriculum standards, which dictate both what is taught in the classroom and what is assessed on high-stakes tests. These standards should be amended to encourage teachers to address all of the facets of quality writing. Under these revised standards, teachers could not only facilitate authorial voice, but also use it as a springboard to build up other narrative skills. By prioritizing authorial voice, educators could improve students' writing skills in all areas.

Tables

Table 1

Measures of academic performance

	Fall			Spring		
	α	*M*	*SD*	α	*M*	*SD*
Peer academic reputation	.78	.08	.80	.80	.07	.79
Academic skills	.90	3.40	1.14	.87	3.54	1.02
Writing Ability	n/a	2.93	1.29	n/a	3.22	1.32
Academic effort	.85	3.92	.93	.89	3.86	.97
Attention problems	.73	2.72	1.17	.75	2.64	1.12
Academic self-concept	.69	2.91	.84	.60	3.03	.72

Table 2

Measures of writing skills

	Inter-rater Pearson *r*	Fall *M*	Fall *SD*	Spring *M*	Spring *SD*
Clarity	.87	2.34	.71	2.56	.68
Conventions	.93	2.08	.77	2.06	.77
Expressiveness	.82	2.02	.76	2.07	.67
Metanarrative comments	.88	.49	.96	.70	1.27
Cultural voice	.77-.96*	1.02	1.50	.81	1.19
Emphasis markers	.86	1.32	1.76	1.16	1.42

Note: Reliability for cultural voice was established between 4 coders; for all other measures, reliability was established between 2 coders

Table 3

Intercorrelations between Writing Skills Measures: Fall

	Conventions	Clarity	Expressiveness	Meta-narrative comments	Cultural voice	Emphasis markers
Conventions	-	.417***	.296**	.329**	.151	.334**
Clarity		-	.092	.153	-.007	.198
Expressiveness			-	.447***	.433***	.367***
Metanarrative comments				-	.189	.357***
Cultural voice					-	.421***
Emphasis markers						-

Note: N = 87
**p < .01, ***p ≤ .001

Table 4

Intercorrelations between Writing Skills Measures: Spring

	Conventions	Clarity	Expressiveness	Metanarrative comments	Cultural voice	Emphasis markers
Conventions	-	.259*	.094	.106	-.010	.155
Clarity		-	.202	.154	-.047	.242*
Expressiveness			-	.293**	.393***	.446***
Metanarrative comments				-	.210*	.313**
Cultural voice					-	.333***
Emphasis markers						-

Note: N = 96
*p < .05, **p < .01, ***p ≤ .001

Table 5

Stability of Writing Skills Measures from Fall to Spring

	Stability: Pearson's *r*
Conventions	.510***
Clarity	-.002
Expressiveness	.121
Metanarrative comments	.322**
Cultural voice	.090
Emphasis markers	.231

Note: N = 68
**p < .01, ***p ≤ .001

Table 6

Principal Components Factor Analysis of Writing Skills Measures

	Fall		Spring	
	Factor 1	Factor 2	Factor 1	Factor 2
Conventions	.651	.493	.332	.670
Clarity	.412	.749	.424	.661
Expressiveness	.731	-.291	.749	-.190
Metanarrative comments	.672	-.009	.575	-.015
Cultural voice	.583	-.547	.587	-.551
Emphasis markers	.728	-.126	.760	-.037
Eigenvalue	2.449	1.205	2.103	1.226
Variance explained (%)	40.81	20.08	35.04	20.43
Cumulative variance explained (%)	40.81	60.89	35.04	55.48

Table 7

Correlations between Fall Writing Skills and Fall Academic Performance Measures

	Academic Self-Concept	Academic Skills	Academic Effort	Writing Ability	Attention Problems	Academic Reputation
Conventions	.381**	.692***	.265*	.589***	-.322**	.505***
Clarity	.153	.412***	.147	.378***	-.127	.331**
Expressiveness	.181	.226*	.067	.165	-.182	.208
Metanarrative comments	.043	.166	.020	.204	-.179	.148
Cultural voice	.144	.133	.034	.167	-.238*	.085
Emphasis markers	.192	.299**	.062	.198	-.213	.365***

Note: N = 59 for Academic Self-Concept, N = 79 for teacher ratings, N = 87 for Academic Reputation
*p < .05, **p < .01, ***p ≤ .001

Table 8

Correlations between Spring Writing Skills and Spring Academic Performance Measures

	Academic Self-Concept	Academic Skills	Academic Effort	Writing Ability	Attention Problems	Academic Reputation
Conventions	.224	.483***	.312**	.364***	-.306**	.439***
Clarity	.017	.259*	.125	.244*	-.157	.251*
Expressiveness	.035	.182	.000	.197	-.068	.379***
Metanarrative comments	.096	.043	-.032	.199	.087	.194
Cultural voice	-.025	.067	-.028	.077	.100	.123
Emphasis markers	-.035	.183	.190	.245*	-.150	.317**

Note: N = 72 for Academic Self-Concept, N = 95 for teacher ratings, N = 96 for Academic Reputation
*p < .05, **p < .01, ***p ≤ .001

Table 9

Correlations between Fall Writing Skills and Spring Academic Performance Measures

	Academic Self-Concept	Academic Skills	Academic Effort	Writing Ability	Attention Problems	Academic Reputation
Conventions	.336*	.619***	.300**	.446***	-.359**	.519***
Clarity	.154	.384***	.204	.246*	-.110	.251*
Expressiveness	.176	.275*	.018	.096	-.030	.221
Metanarrative comments	.073	.110	.007	.233	.038	.209
Cultural voice	.150	.107	.032	.016	-.015	.145
Emphasis markers	.140	.243*	.057	.116	-.119	.297**

Note: N = 54 for Academic Self-Concept, N = 73 for teacher ratings, N = 74 for Academic Reputation
*p < .05, **p ≤ .01, ***p ≤ .001

Appendix A: Transcription and Segmentation Procedures

1. Open the folder "Narrative Data" in My Documents. Click on the Word document with the name of the teacher whose class' stories you are transcribing.
2. Before beginning to type, it is important to remember to type only what the student has written. Do not add extra punctuation or correct punctuation, capitalization or spelling. Make sure you turn off the autocorrect feature before you begin (click on the Office button in the top left corner, click on the Word Options button at the bottom of the menu, select Proofing from the menu on the left, click on the Autocorrect Options button, and un-check the appropriate boxes).
3. Single spacing should be used and type should be done in Times New Roman 12 point font. You do not need to insert a page break between stories; just leave a blank line.
4. The first thing to be typed is the identification number of the student. Space one tab, type in the grade level (in the file cabinet, there will be a list of the homeroom teachers and their corresponding grade levels). Space one tab and then type in the teacher's name. Space one tab and type either Spring or Fall and the year the story was written. Finally, add a blank line.

   Example: 1007 4 Smith Spring 2010

5. If the student has given the essay a title, type it on the first line, but do not give it a line number.
6. Number each sentence in the student's story. The first sentence is always one (1., 2., 3., 4., etc.).

7. How to determine sentence boundaries:

   a. Each sentence of the student's essay is entered on a separate numbered line.

      Example: 1. Suzy and I went to the park.

      2. We went down the slide.

   b. If the student includes period, exclamation mark or question mark we recognize that as the end of a sentence, even if the punctuation is incorrectly placed.

      Example: 1. Have you ever?

      2. Seen a whale before!

      Do not divide just because a comma is used.

      Example: 1. Have you ever, Seen a whale before!

   c. If the student does not use punctuation, split their units according to subject-predicate units. Put each unit on a separate line, but do not add any punctuation that the student did not use.

      Example: a student may write -

      "I live in a house I have a brother

      he is six."

      1. I live in a house

      2. I have a brother

      3. he is six.

   d. If a student has a compound sentence that is combined with the conjunction "and," "but," or "so," separate the compound sentences. Put the first subject/predicate on one line (keep comma on this line if one is used to separate the two parts of the compound sentence) and skip to the next line

for the additional subject and predicate (words such as “and” and “but” that combine the two parts of the sentence are to be put on this second line). Separate lines this way ONLY if both clauses contain both a subject and a predicate. If two clauses are separated by “because,” “although,” or “when,” DO NOT separate them.

Example: a student may write-

“I had a birthday yesterday, and we celebrated it at school.”

1. I had a birthday yesterday,
2. and we celebrated it at school.

e. If the student is documenting dialogue include the speaker and the statement on the same line.

Example: A student may write:

“And she said, ‘You need to quit.’”

1. And she said, “You need to quit.”

8. “The End”: include it at the end, but do not number the line.
9. Brackets **[ ]** should be used to indicate any of the transcriber’s own comments. It should also be used to describe any pictures or orthographic features (such as using large or especially bold print).
10. Any word that cannot be read should be referred to in the transcription with a bracketed dash **[-]**. The length of the dash depends on the length of the word.
11. Any letter than cannot be read should be referred to in the transcription with an *****.
12. Use a question mark surrounded with brackets **[?]** to indicate you have guessed at the spelling of the preceding word.

13. For any story that includes a person's last name or family name, change the last name but keep the first name the same. Change the name minimally to preserve any ethnic information. Keep teachers' and staff's names the same. For example, you might change Garcia to LaRosa or Nguyen to Huynh. However, if a student mentions the name of another child, click the highlighter button and highlight the name; do NOT change it – we will need to match these names with their corresponding ID numbers later.
14. When you finish transcribing a story, write TRANSCRIBED BY: plus your name and the date at the end of the story on the electronic file. When you finish checking a story, add CHECKED BY: plus your name and the date. In both cases, write the same thing at the top of the paper copy of the story.

Appendix B: Coding Manual

Read the entire story through once before beginning to code. If you are coding a rating variable, assign a rating for the entire story. If you are coding a count variable, after reading through the whole story once, read through it again coding line-by-line, keeping count of the number of lines that include at least one instance of each variable. If a line includes more than one instance of the same variable, we will count it only once. If there are no instances of a variable in a particular story, enter a zero. If the child did not provide a story (or declined to participate by writing something along the lines of "I don't know" or "I can't think of anything"), leave the variable blank for that ID number – do not enter a zero in these cases. Code all stories for only one variable at a time (for example, do not code one story for emphasis markers and then expressiveness; instead, code the entire corpus for emphasis markers then the entire corpus for expressiveness).

It is imperative that you have these coding instructions in front of you the whole time you are coding, and that you refer to them often. It is surprisingly easy to miss whole categories of events, especially those that are relatively rare. When they have not shown up in several stories, coders will sometimes stop looking for them. The only way to avoid this is to be regularly looking at the coding manual. As a guideline, never code more than ten stories without re-reading the category descriptions.

Coding is not an activity that you can do when you are tired or distracted. Take regular breaks, and monitor your ability to stay alert. Thank you for being conscientious and thoughtful about this part of the research effort!

Mechanics

Clarity: Clarity refers to the ease with which the writing can be understood. There are two primary contributors to clarity: sequentiality (clarity of plot structure) and issues of common ground (knowledge shared between the author and reader). A story that is high in clarity is focused and coherent, without irrelevant digressions, and the events in the story follow a logical sequence (note that this may not be the same as a chronological sequence; the author may begin the story in the middle of the action and then move back to the beginning, but the sequence in which the events occurred remains clear). The author provides all of the information necessary for you to comprehend the story. After reading the story once, you understand what happened. On the other hand, a story that is low in clarity is difficult to follow, and the order of events may seem scrambled. The author may make mistaken assumptions about the reader's knowledge (for example, using pronouns without explaining to whom they refer). You may have to make several passes through the story to grasp the events described.

1 = This author did not express the story clearly. The story does not include enough information; that is, there are serious problems of common ground. Connections between events or ideas may be confusing or absent altogether. Though you may be able to understand the overall gist of the story, parts of it remain unclear after multiple readings.

2 = This author did an adequate job of expressing the story clearly. There may be a few digressions, and occasional problems of common ground. However, you can easily understand the majority of the story.

3 = This author expressed the story very clearly. The connections between ideas/events make sense, and all of the necessary information is included (there are no major problems of common ground). After reading the story once, you have a very clear understanding of what happened.

Conventions: This rating assesses the correct use of conventions of writing, namely grammar, punctuation, capitalization and spelling. When assigning a rating, consider the number of errors relative to the length and complexity of the story. If the author appears to flout conventions for stylistic effect (e.g., using all capital letters for emphasis or using incomplete sentences to add drama) do not count these as errors.

1 = This story contains numerous errors in spelling, grammar, punctuation, and capitalization relative to the length and complexity of the story. There are errors in almost every line.

2 = This story contains some errors in spelling, grammar, punctuation, and capitalization relative to the length and complexity of the story. Several lines in the story may contain errors. The grammar and punctuation is generally simplistic.

3 = This story contains few errors in spelling, grammar, punctuation, and capitalization relative to the length and complexity of the story, and the author uses more sophisticated forms of grammar and punctuation. The author uses these conventions to guide the reader through the text, and they contribute to the clarity and style of the writing.

Voice

Expressiveness: Expressiveness refers to an author's ability to tell an engaging story with individual flavor, a distinct writing style unique to that person. As you read a story with a strong voice, you get a sense of the person behind the words and feel as if you can almost hear the child speaking to you. Elements of tone, such as humor or irony, may contribute to expressiveness. In addition, it is often marked by creative use of language, such as similes and metaphors (e.g., "He ran at me like a bull, with fury in his eyes") or by the use of vocabulary that is advanced or atypical for the author's grade level (e.g., "I thought I would perish" rather than "I thought I would die").

1 = This story is straightforward, with minimal evidence of creativity. The author seems indifferent, uninvolved, or distanced from the audience. The writing comes across as monotone rather than expressive.

2 = This story is mostly straightforward, but there are some instances of creativity and expressiveness. The writer's voice may seem to come through for a few lines, then fade away.

3 = This story is expressive, and the author's voice comes through clearly throughout the story. The writing displays a unique personal style, and the reader feels a strong interaction with the writer.

Metanarrative Comments: Metanarrative comments occur when the author makes a reference to the reader (e.g., "I hope you like my story that I am going to tell you about," or, "So like I told you he kept on bothering me) or to the process of telling the story (e.g., "Well that's my story and it's all 100% true!"). Do count rhetorical questions by the

author as metanarrative comments (e.g., "Did I mention we where in 2nd grade"). Also count "The End" or "that's it".

Record the total number of lines that include at least one metanarrative comment.

If the child did not provide a story, leave the space blank (do not code as a 0).

Emphasis markers: Emphasis markers are literary devices that direct the reader to pay special attention to particular parts of the story. We will count two types of emphasis markers: intensifiers and orthographic attentional devices.

Intensifiers are words or phrases that express extremity or exceptionality. They occur when the author quantifies an adjective or adverb to increase or decrease the intensity of the statement. For example, "it was **pretty** dumb," or "my eyes get watery because I get **way too** mad" (although children may write "to" or "two" instead of "too", these words will be counted if they quantify an adjective or adverb in the context of the story).

Orthographic attentional devices are orthographic features that the author uses to emphasize their thoughts, actions, or emotions. This includes the use of purposefully misspelled words (e.g., "I got **sooo** made"); one or more exclamation points (e.g., "It is still going on today**!!**"); repeated words (e.g., "he talks about my mother **very very** wrong"); all capital letters (e.g., "**YOUR NOT MY FRIND ANYMORE**!"); and onomatopoeia (e.g., "i kicked him in the leg he said **ow**!").

Record the total number of lines that include at least one emphasis marker. If the child did not provide a story leave the space blank (do not code as a 0).

Cultural Voice: With this coding we hope to capture a facet of narrative voice by counting the colloquial or idiomatic expressions in each story. These are instances in

which the author incorporates linguistic expressions or turns of phrase whose meaning is not captured in a literal translation. The use of such expressions is evidence of the child's integration in a cultural community, and of the child's skill at appropriating language for his or her own purposes.

What makes this coding difficult is the fact that, pushed to the limits of our logic, we see that language is virtually never literal, that abstraction, by its very nature, is metaphorical. Any use of language entails an appropriation of a cultural toolkit and is evidence of the integration into a cultural community. Our task is to distinguish a subset of non-literal language we want to designate as idiomatic or colloquial.

We will approach this task by asking whether the expression the child uses could be used to identify him or her as part of a particular cultural community. These fall into several categories:

1. <u>Speech that identifies the child as a member of a school or educational community.</u>
    a. This will include all expressions that have special meaning in a school setting, such as being sent to the principal or sent to the office, getting suspended, telling the teacher, copying off someone's paper, taking names. Note that some of these expressions may be quite literal (e.g., going to the principal's office, putting my name on the board), but the intended meaning is not captured in the literal understanding, and it is the school context that gives the expressions the extra meanings associated with punishment or with reward.
    b. We will also include expressions that are especially common in a school setting, so as to be a regular part of school life, such as 'getting in or standing in line,' (noting that in other places a person would stand 'on line' or 'in a

queue'). Putting something in ABC order or 'raising our hands' are other examples of expressions that describe school-identified (if not school-exclusive) behaviors.

c. We are especially eager to include occasions in which the child appropriates the voice of the teacher or school authority by using specific and identifiable vocabulary of curriculum. For example, we know that 'she gave me an I-message' is part of a violence-prevention curriculum. If a child says, 'I used my walking feet,' and we know that there is a sign on the auditorium door that says 'Please use your walking feet," we would count this.

2. Language that identifies the child as part of a regional language community (e.g., 'y'all' or 'yous guys,' or as part of an ethnic or social class group (e.g., 'honey chile' or 'down yonder.')

3. Language that places the child in a 'generation' or that identifies the author with popular culture, street culture, an 'in group' or peer group.

a. Expressions that identify the author as a child or that would usually only be used by a child or spoken to a child should count here because they identify the author with an age group. These might include 'saying a bad word,' 'calling names,' 'getting a whooping,' or 'being a baby.'

b. We want to count expressions that could be read as the author's claim to be 'street-savvy' or 'cool.' Many of these would be recognized as 'slang' because it would not be used by someone trying to be proper or by someone who was not an insider. Examples of this kind of language are "She's too cool for school,' 'he thinks he's all that,' 'he got a beat down,' 'she got an

attitude with me,' 'checking his family' or 'dissing me.' The use of the verb 'to like' for quotation or for emphasis would count here, as in "I was like, really mad," or "I was like no way."

c. Count any expression that you recognize as a quotation of popular culture (TV, movies, lyrics). Examples of these are "so that's my final answer," or "he got three thumbs up" or "she's bad to the bone."

4. Language that claims cultural authority or identifies the author with a broader cultural context. This would include quotations from classical culture (e.g., the Bible, Shakespeare, Aesop) or the use of the language of jurisprudence or liturgy, (e.g., 'the truth and nothing but the truth,' or 'till death do us part.')

5. Common idioms and metaphors that are recognizable by most members of the larger culture, but would not be easily translatable. Examples of these are, 'kicked the bucket,' 'got on his high horse.' We believe that these show the child's willingness and ability to appropriate culture in a way that the use of more literal language does not. They differ from the above items (1 to 4) in that they do not seem to belong to a specific cultural community or source of cultural authority. The do seem to show, however, that the child is claiming a cultural/literary/discursive tradition in a notable way.

Our task is to count those items while NOT counting the following.

1. We will not count conventional metaphors that are so culturally common that most people do not recognize them as metaphorical at all. We will exclude most orientational metaphors (making up, working it out, getting through it) and possession or acquisition metaphors (getting sick of it, taking a break, getting back at, having a go at). These are certainly metaphorical and

colloquial, but they are so common in the language that they do not serve to identify the child with a cultural community.

2. Similarly, we will exclude common colloquial expressions that are easy to translate, and do not have culturally specific roots. For example, 'here we go again,' 'once in a while,' 'the rest of the day.' The difference between colloquialisms we count in item 5 above and those we do not count may be discerned by imagining a second-language learner trying to make sense of it. You can see that 'once in a while,' or 'ever so often' would cause less puzzlement than 'once in a blue moon.' The latter would count for us; the two former expressions would not.

For each line, code as a 1 if a colloquial or idiomatic expression is present, or as a 0 if it is not present. If a line appears to contain more than one colloquialism or idiomatic expression, we will count it only once. However, if an author uses the same expression multiple times on different lines of the story, we will count it each time it occurs. If the child did not write a story, leave the variable blank for that ID number rather than entering a 0.

Here are coding procedures that will support reliability.

1. Code only when rested and alert.
2. Re-read the coding manual each time you sit down to code, and re-read it after each 15 stories coded.
3. Read the stories out loud. Colloquialisms (and other indicators of 'voice') are more easily recognized when voiced than when read silently.
4. Ask yourself the following questions about each line of the story:

a. Is this ‘school language’?

b. Is this regional language?

c. Does the language suggest an ethnic group or any clear cultural community?

d. Is this ‘street talk’ or ‘jive talking’ or language of ‘the hood’?

e. Does the use of this expression communicate “I’m cool” or “I’m an insider”?

f. Does the use of this expression constitute a claim to cultural authority?

g. Can you identify this expression as a quotation or a reasonable effort at quotation of any popular or classical source?

5. If all of the above answers are ‘no,’ but the expression still seems as though it may be countable, then do the following:

a. Parse the phrase. If its meaning is perfectly obvious to anyone who knows the vocabulary of the component parts, it will not usually count as colloquial. The expressions we want to identify are those that would generally trip up a second language learner. They would have to learn them separately as expressions, or they would have to figure them out from context because the words themselves would not reveal the meaning or would suggest a different meaning than that intended.

b. Ask yourself who would use this expression. If you can identify a particular cultural community in which this expression is common or with which this expression is identified, then we do want to count it. If almost

any speaker of English is likely to say this, it is less likely to be one of the phrases we want to count.

Appendix C: Reliability Report

Clarity

Phase I

In Phase I, Regan Humphrey and Jiawen Li jointly coded 30 stories from a previous data set, collected in 1998 and 1999, discussing ambiguities as we came across them. Our primary challenge in this phase was determining how strict or lenient to be. However, we did not make any major changes to the coding manual in this phase.

Phase II

In Phase II, Humphrey and Li independently coded 2 sets of 30 stories from the 1998/1999 data set, meeting after each set to compare our codes and discuss misses. In these practice sets, we found that we had very few ratings of 1. Originally, in order to receive a rating of 1, the coding manual required that the reader be unable to comprehend a story even after multiple readings. In an attempt to achieve a more even distribution of ratings, we revised the coding manual. The revised criteria for a story rated 1 stated, "Though you may be able to understand the overall gist of the story, parts of it remain unclear after multiple readings." In the course of this coding, we also found that there were two main contributors to clarity: sequentiality (clarity of plot structure) and issues of common ground (knowledge shared between the author and reader). We revised the coding manual again to include more specific guidelines regarding both of these aspects of clarity for each rating.

Next, Li and Humphrey independently coded a set of 10 stories from the Florida Fall 2009 data and a set of 10 stories from the Florida Spring 2010 data. The ID numbers from the Florida data used in this phase were: 25909F, 26687F, 34411F, 11448F,

46911F, 10889F, 81388F, 97727F, 52738F, 55076F, 23582S, 26687S, 53010S, 61346S, 87293S, 57044S, 60503S, 39742S, 60163S, 95324S (F = fall, S = spring).

Phase III

In this phase, Humphrey and Li independently coded 50 randomly selected stories from the Florida Fall 2009 and Spring 2010 data. The inter-rater reliability was .36 (Pearson *r*). The ID numbers included in this set were: 44420F, 54214F, 58569F, 60163F, 56806F, 25455F, 39232F, 74315F, 44311F, 90965F, 45585F, 87293F, 16778F, 31602F, 17802F, 42546F, 60503F, 31926F, 39226F, 41472F, 39376F, 86883F, 38660F, 41818F, 99731F, 15681S, 54214S, 70101S, 48156S, 67743S, 53738S, 63081S, 36691S, 89482S, 26131S, 34411S, 13117S, 60832S, 45585S, 28484S, 29794S, 97727S, 25375S, 26525S, 33232S, 73902S, 21487S, 44420S, 31751S, 56582S (F = fall, S = spring).

In an effort to achieve a higher reliability, Humphrey and Li conducted further discussion and practice then independently coded another set of 50 randomly selected stories from the Florida Fall 2009 and Spring 2010 data. We achieved an inter-rater reliability of .87 (Pearson *r*). The ID numbers included in this set were: 87293F, 13117F, 99650F, 80388F, 70098F, 70101F, 31751F, 43761F, 25802F, 26057F, 99566F, 70888F, 76837F, 68757F, 56582F, 86317F, 63081F, 24653F, 25375F, 58719F, 60289F, 39350F, 77933F, 68042F, 61870F, 26131S, 54214S, 67743S, 28484S, 86883S, 23407S, 27173S, 68757S, 29643S, 39376S, 41472S, 12915S, 31602S, 52738S, 39232S, 80002S, 57920S, 34972S, 44311S, 41818S, 58569S, 70098S, 90995S, 46911S, 16288S (F = fall, S = spring).

Conventions

Phase I

In Phase I, Regan Humphrey and Annika Wuerfel jointly coded 5 stories from the 1998/1999 data set, discussing ambiguities as we came across them. At this point, we found that the most difficult aspect of the coding was balancing the various facets of conventions (grammar, punctuation, capitalization, and spelling), as many children were strong in some areas and weak in others. Nevertheless, we did not make any substantial changes to the coding manual in this phase.

Phase II

In Phase II, Humphrey and Wuerfel independently coded a set of 25 stories and a set of 30 stories from the 1998/1999 data, as well as a set of 10 stories from the Florida Fall 2009 data and a set of 10 stories from the Florida Spring 2010 data. After each set we met to compare codes and discuss our misses. In the two practice sets using Florida data, we had 3 Type III errors and no Type I or Type II errors. We did not find it necessary to make any changes to the coding manual during this phase. The ID numbers from the Florida data used in this phase were: 25909F, 26687F, 34411F, 11448F, 46911F, 10889F, 81388F, 97727F, 52738F, 55076F, 23582S, 26687S, 53010S, 61346S, 87293S, 57044S, 60503S, 39742S, 60163S, 95324S (F = fall, S = spring).

Phase III

In Phase III, Humphrey and Wuerfel independently coded 50 randomly selected stories from the Florida Fall 2009 and Florida Spring 2010 data. The inter-rater reliability for this set was .69 (Pearson *r*). The ID numbers included in this set were: 44420F, 54214F, 58569F, 60163F, 56806F, 25455F, 39232F, 74315F, 44311F, 90965F,

45585F, 87293F, 16778F, 31602F, 17802F, 42546F, 60503F, 31926F, 39226F, 41472F, 39376F, 86883F, 38660F, 41818F, 99731F, 15681S, 54214S, 70101S, 18156S, 67743S, 53738S, 63081S, 36691S, 89482S, 26131S, 34411S, 13117S, 60832S, 45585S, 28484S, 29794S, 97727S, 25375S, 26525S, 33232S, 73902S, 21487S, 44420S, 31751S, 56582S (F = fall, S = spring).

In an effort to achieve a higher reliability, Humphrey and Wuerfel conducted further discussion and practice then independently coded another set of 50 stories randomly selected from the Florida Fall 2009 and Spring 2010 data. We achieved a reliability of .93 (Pearson *r*). The ID numbers included in this set were: 87293F, 13117F, 99650F, 80388F, 70098F, 70101F, 31751F, 43761F, 25802F, 26057F, 99566F, 70888F, 76837F, 68757F, 56582F, 86317F, 63081F, 24653F, 25375F, 58719F, 60289F, 39350F, 77933F, 68042F, 61870F, 26131S, 54214S, 67743S, 28484S, 86883S, 23407S, 27173S, 68757S, 29643S, 39376S, 41472S, 12915S, 31602S, 52738S, 39232S, 80002S, 57920S, 34972S, 44311S, 41818S, 58569S, 70098S, 90995S, 46911S, 16288S (F = fall, S = spring).

Emphasis Markers

Phase I

In Phase I, Ginny Brady and Regan Humphrey coded 53 stories from the Memphis Spring 2010 data set, discussing ambiguities as we came across them together. Because the Memphis data was already coded for this variable we did not make substantial changes to the coding manual. However, we did contact previous coders to determine if they included certain features, such as superlatives, underlining, and ellipses.

They did not include these features, so we decided not to count underlining and ellipses and to add superlatives using the search feature in Microsoft Word.

Phase II

In Phase II, Brady and Humphrey independently coded two sets of 20 stories from the Memphis Spring 2010 data set, discussing our misses after each set. As in Phase I, we did not make any major changes to the coding manual.

Phase III

In Phase III, Brady and Humphrey independently coded a set of 50 randomly selected stories from the Florida Fall 2009 and Spring 2010 data sets. For this set, we had 3 Type I errors, no Type II errors, and 19 Type III errors. We achieved a reliability of .86 (Pearson *r*). The ID numbers used in this set were: 87293F, 13117F, 99650F, 13117F, 99650F, 80388F, 70098F, 70101F, 31751F, 43761F, 25802F, 26057F, 99566F, 70888F, 76837F, 68757F, 56582F, 86317F, 63081F, 24653F, 10889F, 58719F, 60289F, 26687F, 77933F, 68042F, 61870F, 26131S, 54214S, 67743S, 28484S, 86883S, 23407S, 61346S, 68757S, 29643S, 29643S, 41472S, 12915S, 31602S, 52738S, 39232S, 80002S, 80002S, 34972S, 44311S, 41818S, 58569S, 70098S, 90995S, 95324S, 95324S (F = Fall, S = Spring).

In an effort to achieve a higher reliability, Brady and Humphrey coded another set of 50 randomly selected stories from the Florida Fall 2009 and Spring 2010 data. We achieved a reliability of .859 (Pearson *r*). In this set, we had 8 Type I errors, no Type II errors, and 27 Type III errors. The ID numbers included in this set were: 44420F, 54214F, 58569F, 60163F, 56806F, 25455F, 39232F, 74315F, 44311F, 90965F, 45585F, 97727F, 16778F, 31602F, 17802F, 42546F, 60503F, 31926F,39226F, 41472F, 39376F,

86883F, 38660F, 41818F, 26887F, 15681S, 53010S, 70101S, 18156S, 60503S, 53738S, 63081S, 36691S, 89482S, 26687S, 34411S, 13117S, 60832S, 45585S, 39742S, 29794S, 97727S, 25375S, 26525S, 33232S, 73902S, 21487S, 44420S, 31751S, and 56582S (F = Fall, S = Spring).

Expressiveness

Phase I

In Phase I, Ginny Brady, Kyle Capstick, and Regan Humphrey coded 10 stories from the Memphis Spring 2010 data, rating each story together. Our disagreements primarily resulted from uncertainties about how pervasive the evidence of expressiveness needed to be, relative to the length of the story, to achieve a rating of a 2 versus a 1 or a 3 versus a 2. We found that we were able to clarify this issue with practice, so we did not make any major changes to the coding manual. The ID numbers used during this phase were: 1891C, 2180C, 2137C, 2004C, 2013C, 2135C, 1833C, 2020C, 1906C, and 1891P (C = conflict prompt, P = playground prompt).

Phase II

In Phase II, Brady, Capstick, and Humphrey independently coded 3 sets of 10 stories from the Memphis Spring 2010, Florida Fall 2009, and Florida Spring 2010 data. We discussed our misses after each set. As in Phase I, we did not make any substantial changes to the coding manual. In this phase, we had 9 Type III errors, but no Type I or Type II errors. The ID numbers from the Memphis data used in this phase were: 1891F, 2180P, 2137P, 2004P, 2016C, 2135P, 1833P, 2020P, 1906P, 1906F, 2079C, 2180F, 2171C, 2137F, 2004F, 2048C, 2135F, 2121P, 2020F, 1906P (C = conflict prompt, P = playground prompt, F = friendship prompt). The ID numbers from the Florida data used

in this phase were: 25909F, 26687F, 34411F, 11448F, 46911F, 10889F, 81388F, 97727F, 52738F, 55076F, 23582S, 26687S, 53010S, 61346S, 87293S, 57044S, 60503S, 39742S, 60163S, and 95324S (F = Fall, S = Spring).

Phase III

In Phase III, Brady, Capstick, and Humphrey independently coded 50 randomly selected stories from the Memphis Spring 2010, Florida Fall 2009, and Florida Spring 2010 data. . For this set, we had one Type I errors, no Type II errors, and 22 Type III errors (we recorded an error if all three coders were not in agreement). Using Pearson correlations, we achieved a reliability of .72 (between Brady and Humphrey), .704 (between Capstick and Humphrey), and .596 (between Brady and Capstick). The ID numbers from the Florida data included in this set were: 87293F, 13117F, 99650F, 80388F, 70098F, 70101F, 31751F, 43761F, 25802F, 26131S, 54214S, 67743S, 28484S, 86883S, 23407S, 68757S, 29643S, and 39376S (F=Fall, S=Spring). The ID numbers from the Memphis data included in this set were: 2178F, 2183C, 2232C, 2015P, 2147C, 1969F, 1956C, 2029C, 2247C, 1950C, 2138P, 1964P, 1997C, 2200P, 2035F, 2153F, 2186C, 2161C, 2228P, 2243F, 2040C, 2062P, 1863F, 2210F, 2179C, 2129P, 2140P, 2219P, 2164C, 2163P, 2170F, and 2231C (C= conflict prompt, F= friendship prompt, P= playground prompt).

In an attempt to achieve a higher reliability, Brady, Capstick, and Humphrey independently coded another set of 50 stories from the Florida Fall 2009 data, Florida Spring 2010 data, and Memphis Spring 2010 data. For this set, we had 26 Type III errors but no Type I or Type II errors (we recorded an error if all three coders were not in agreement). Using Pearson correlations, we achieved a reliability of .82 (Brady and

Capstick), .666 (Capstick and Humphrey), and .648 (Brady and Humphrey). Given these reliabilities, we decided that only Brady and Capstick would conduct the expressiveness coding. The ID numbers from the Florida data included in this data set were: 26057F, 99566F, 70888F, 76837F, 68757F, 56582F, 86317F, 63081F, 24653F, 41472S, 12915S, 31602S, 52738S, 39232S, 80002S, 57920S, 34972S, and 44311S (F = fall, S = spring). The ID numbers from the Memphis data included in this set were: 1900C, 2166F, 2260F, 2018F, 2227P, 2177C, 1965F, 2254F, 2039C, 2185C, 2154F, 1987F, 2162C, 1929C, 2212F, 2248F, 2165P, 2130P, 2015F, 2222P, 2149P, 2008C, 1929P, 1940P, 2230P, 2229C, 2241P, 2163P, 1990P, 2054C, 2177P, 2226C (C = conflict, P = playground, F = friendship).

Cultural Voice

Phase I

In this phase, Caitlin Campbell, Alice Davidson, Regan Humphrey, and Marsha Walton coded 3 sets of 30 stories from the 1998-1999 data, comparing codes via Skype. At this point, we attempted to code all instances of colloquialisms and idiomatic language. However, over the course of coding it became clear that almost all language can be viewed as idiomatic, and many instances of idiomatic language did not seem to indicate that the child was really appropriating the language of his or her culture. Consequently, we revised the coding manual to include only those idioms that identified an author as a member of a particular cultural group, as well as idioms whose meanings could not be inferred simply by knowing the meanings of their component parts.

Phase II

In this phase, Campbell, Davidson, Humphrey, and Walton independently coded 2 sets of 28 stories from the Florida Fall 2009 and Memphis Spring 2010 conflict stories. After each set, we discussed differences via Skype and further refined the coding manual to reflect the specific instances of colloquial or idiomatic language that we wished to include. The ID numbers from the Florida data used during this phase were: 57920, 99650, 69228, 53010, 15700, 63081, 31751, 39376, 11448, 54214, 53670, 91474, 34972, 25909, 41472, 39262, 77933, 22432, 90965, 70098, 99731, 26687, 38660, 44635, 31602, and 34411. The ID numbers from the Memphis data used during this phase were: 2253, 2255, 2248, 2251, 2249, 2252, 2140, 2156, 2183, 2130, 2135, 2168, 2161, 2164, 2150, 2138, 2166, 2179, 2162, 2176, 2148, 2147, 2175, 2137, 2155, 2157, 2153, 2159, 2139, and 2185.

Phase III

In this phase, Campbell, Davidson, Humphrey, and Walton independently coded 38 randomly selected stories from the Florida Fall 2009 and Memphis Spring 2010 data. We achieved the following inter-rater reliabilities (Pearson *r*): .77 (Campbell and Davidson), .82 (Campbell and Walton), .86 (Davidson and Humphrey), .86 (Campbell and Humphrey), .91 (Humphrey and Walton), and .96 (Davidson and Walton). The ID numbers from the Florida data included in this set were: 23582, 17690, 26687, 54831, 53010, 57920, 61346, 25802, 87293, 16778, 57044, 11448, 60503, 56806, 39742, 43761, 60163, 39350, 95324, and 56582. The ID numbers from the Memphis data included in this set were: 1891 C, 2079 P, 2180 C, 2171 C, 2170 P, 2254 P, 2004 C, 2007 F, 2013 C,

2016 P, 2144 F, 2174 C, 1833 C, 2121 F, 2020 C, 2027 C, 1906 C, 1905 C (C = conflict, P = playground, F = friendship).

Metanarrative Comments

Metanarrative comments were coded by a pair of previous students in Dr. Walton's lab, using similar procedures to those described above. They achieved an inter-rater reliability of .88 (Pearson *r*).

Appendix D: The Author's Toolkit

The goal of this unit is to strengthen children's literacy skills by helping them to identify various tools that authors use to create a clear and compelling narrative and by encouraging the children to use those tools in their own writing. The unit is best suited for children who already have a strong cultural voice – an important authorial tool that is often neglected in formal instruction. The activities described below are designed to build on this strength, using it as the starting point in the exploration of the author's toolkit.

The unit will begin with an age-appropriate reading that features strong cultural voice. To help students identify the instances of cultural voice in the story, the teacher should preface the reading with an introduction to the particular cultural group it features. The teacher may choose to tailor the reading selection to the student population or to recent topics of study in other subject areas, such as social studies. Possible selections for younger children and the cultures highlighted include: *The Talking Eggs* by Robert D. San Souci (Creole); *An Orange for Frankie* (rural Midwest), *Pink and Say* (African American and European American Union soldiers in the Civil War), *The Trees of the Dancing Goats* (Russian and Jewish), *Mrs. Katz and Tush* (Polish and Jewish), all by Patricia Polacco; and *Mirandy and Brother Wind*, *Precious and the Boo Hag*, and *The Honest-to-Goodness Truth*, all by Patricia McKissack and featuring rural African American culture. Possible selections for older children include *The House on Mango Street* by Sandra Cisneros (Hispanic American) and *Tom Sawyer* by Mark Twain (African Americans and European Americans in the South), either excerpted or in their entirety.

The teacher will read the story aloud to the class (in the case of older children reading a longer work, the teacher should choose relevant excerpts to read aloud). Afterwards, the teacher will ask the students to think about how the characters in the book sound when they speak. In the course of this discussion, the teacher should draw students' attention to instances of dialect, colloquialisms, and idioms, asking the students how these elements serve the story. The teacher can identify these elements as instances of cultural voice (with older students, the teacher may encourage the students to come up with a label for this type of language on their own before introducing the term "cultural voice").

Next, the class will brainstorm examples of cultural voice from the students' own cultures (this can include regional culture, peer group, popular culture, ethnic group, and so on). The students will then write a story that features cultural voice. Afterwards, the students will break up into pairs and share their stories with a partner. The partners will try to identify instances of cultural voice in each others' stories.

The next portion of the unit will branch out to include other authorial tools. The teacher will introduce the idea that authors use a variety of tools to tell a story; cultural voice is just one of many such tools. Some tools help the reader to understand what is happening in the story, while others serve to make the story more interesting and fun to read. With the teacher's guidance, and referring to the previously read book as necessary, the class will brainstorm examples of these tools (e.g., dialogue, plot structure, punctuation, character development, figurative language, tone, morals, etc.). Next the teacher will read a new book aloud to the class (for older students who previously read a longer work, the teacher may choose to focus on a different excerpt instead of a

completely new book). Afterwards, the teacher will ask the class to identify the tools that this author used to tell the story. How did this author's use of tools differ from the first author? How were they similar? At this point, students will write another story using some of the tools that they have identified. Again, students will share their stories in pairs, and they will try to identify the tools that their partners have used in their stories.

For this last part of the unit, the teacher may choose to examine only two or three tools at a time, depending on the age and literacy skills of the students. After the students have mastered those tools, the teacher can repeat the process with a new set of tools, asking the children to identify them in literature and then incorporate them into their own writing. An ideal way to ensure that students maintain mastery of previously learned tools is to incorporate them into oral storytelling. Students should have the opportunity to share stories with their peers on a regular basis. During this story-sharing time, the teacher may point out when students make use of an authorial tool. For example, "John used dialogue in his story." Alternatively, the teacher may ask students to point out the tools used in their peers' stories. Either way, this will help keep the full spectrum of the author's toolkit fresh in the students' minds and give them regular practice using the tools in a more relaxed environment.

References

Antonacci, P.A. & O'Callaghan, C.M. (2011). *Developing content area literacy*. Los Angeles, CA: SAGE Publications.

Bamberg, M. (2004). Forms and functions of 'slut bashing' in male identity constructions in 15-year-olds. *Human Development, 47*(6), 331 – 353.

Benner, G.J., Beaudoin, K., Kinder, D., & Mooney, P. (2005). The relationship between the beginning reading skills and social adjustment of a general sample of elementary aged children. *Education and Treatment of Children 28*(3), 250 – 264.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bucholtz, M. & Hall, K. (2004). Language and identity. In A. Duranti (ed.), *A companion to linguistic anthropology*. Malden, MA: Blackwell, 369 – 394.

Caillies, S. & Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. *Developmental Science, 11*(5), 703 – 711.

Cisneros, S. (1984). *The House on Mango Street*. New York, NY: Vintage Books.

Clark, H.H. (1992). *Arenas of Language Use*. Chicago, IL: University of Chicago Press.

Conduct Problems Prevention Research Group (1999). Initial impact of the fast track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology, 67*(5), 631 – 647.

Culham, R. (2003). *6+1 traits of writing*. New York, NY: Scholastic Professional Books.

Davidson, A.J., Walton, M.D., Humphrey, R.C., Campbell, C. (2012, March). *"She Would act Like a Queen at Ressice and Think she Would Rule the Fort": Cultural Voice and Early Identity Development*. Poster presented at the Society for Research on Adolescence Biennial Meeting, Vancouver, BC, Canada.

Dray, A.J., Selman, R.L., & Schultz, L.H. (2009). Communicating with intent: A study of social awareness and children's writing. *Journal of Applied Developmental Psychology 30*(2), 116 – 128.

Educational Records Bureau (2006). WrAP analytic scoring guide. Retrieved April 27, 2010 from the Educational Records Bureau's website: http://www.erbtest.org/assets/2007/2/8/wrapscoringguide.pdf.

Engel, S.E. (1995). *The stories children tell*. New York, NY: W.H. Freeman and Company.

Fang, Z. (2001). The development of schooled narrative competence among second graders. *Reading Psychology 22*, 205 – 223.

Florida Department of Education (2009). Florida Differentiated Accountability Program 2009-2010 School Improvement Plan. Retrieved November 14, 2011 from the Orange County Public Schools' website: https://www.ocps.net/cs/services/accountability/improvement/SIP/Fern%20Creek%20E%2009-10%20DOE.pdf.

Gansle, K.A., VanDerHeyden, A.M., Noell, G.H., Resetar, J.L., & Williams, K.L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review 35*(4), 435 – 450.

Gest, S.D., Domitrovich, C.E., & Welsh, J.A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology, 97*(3), 337 – 346.

Gombert, J.É.. (1992). *Metalinguistic development* (T. Pownall, Trans.). Chicago: University of Chicago Press. (Original work published 1990).

Harris, A. (2009). "It was hard for me to say": A qualitative study of KidsTalk, a story-sharing program in an urban community organization. Master's Thesis, University of Memphis.

Harris, A. R. & Walton, M. D. (2009). "Thank you for making me write this." Narrative skills and the management of conflict in urban schools. *The Urban Review: Issues and Ideas in Public Education, 41(4),* 287 – 311.

Harris, K.R., Graham, S., & Mason, L.H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*(2), 295 – 340.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*(1), 87 – 97.

Heck, R.H. & Crislip, M. (2001). Direct and indirect writing assessments: Examining issues of equity and utility. *Educational Evaluation and Policy Analysis 23*(3), 275 – 292.

Humphrey, R.C., Walton, M.D., Davidson, A.J. (2012, May). *"She could have said it nicer": Authorial Voice and Peer Adjustment*. Paper presented at the Narrative Matters Conference, Paris, France.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60*(2), 237 – 263.

Kerbel, D. & Grunwell, P. (1997). Idioms in the classroom: An investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching and Therapy, 13*, 113 – 123.

Labov, T. (1992). Social and language boundaries among adolescents. *American Speech, 67*, 339-366.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lane, K.L., Graham, S., Harris, K.R., Little, M.A., Sandmel, K., & Brindle, M. (2010). Story writing: The effects of self-regulated strategy development for second-grade students with writing and behavioral difficulties. *The Journal of Special Education, 44(2)*, 107 – 128.

Lane, K.L., Harris, K.R., Graham, S., Weisenbach, J.L., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second-grade students with behavioral and writing difficulties. *The Journal of Special Education, 41(4)*, 234 – 253.

Mason, L.H., Benedek-Wood, E., & Valasa, L. (2009). Teaching low-achieving students to self- regulate persuasive quick write responses. *Journal of Adolescent and Adult Literacy, 53*(4), 303 – 312.

Mason, L.H., Kubina, R.M., & Taft, R.J. (2011). Developing quick writing skills of middle school students with disabilities. *Journal of Special Education, 44*(4), 205 – 220.

Mason, L.H., Kubina, R.M, Valasa, L.L., & Cramer, A.M. (2010). Evaluating effective writing instruction for adolescent students in an emotional and behavior support setting. *Behavioral Disorders, 35*(2), 140 – 156.

Mason, L.H. & Shriner, J.G. (2008). Self-regulated strategy development instruction for writing an opinion essay: Effects for six students with emotional/behavioral disorders. *Reading and Writing, 21*, 71 – 93.

McCabe, A. (1997). Developmental and cross-cultural aspects of children's narration. In M. Bamberg (ed.), *Narrative Development: Six Approaches*. Mahwah, NJ: Lawrence Erlbaum Associates, 137 – 174.

McKissack, P. (1997). *Mirandy and Brother Wind*. New York, NY: Dragonfly Books.

McKissack, P. (2003). *The Honest-to-Goodness Truth*. New York, NY: Aladdin.

McKissack, P. (2005). *Precious and the Boo Hag*. New York, NY: Atheneum Books.

Mullis, I.V.S. (1984). Scoring direct writing assessments: What are the alternatives? *Educational Measurement: Issues and Practice, 3*(1), 16 – 18.

National Council of Teachers of English (2004). Writing Rubric. Retrieved October 9, 2011 from the NCTE website: http://www.readwritethink.org/files/resources/lesson_images/lesson401/Rubric.pdf.

No Child Left Behind Act of 2001, Pub. L. No. 107-110. Retrieved October 26, 2011from the Department of Education's website: http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Novak, J.R., Herman, J.L., & Gearhart, M. (1996). Establishing validity for performance-based assessments: An illustration for collections of student writing. *Journal of Educational Research, 89*(4), 220 – 233.

Paley, V.G. (1990). *The boy who would be a helicopter: The uses of storytelling in the classroom*. Cambridge, MA: Harvard University Press.

Paley, V.G. (1997). *The Girl with the Brown Crayon*. Cambridge, MA: Harvard University Press.

Paley, V.G. (2004). *A child's work: The importance of fantasy play*. Chicago, IL: University of Chicago Press.

Polacco, P. (1994). *Pink and Say*. New York, NY: Philomel Books.

Polacco, P. (2000). *The Trees of the Dancing Goats*. New York, NY: Aladdin.

Polacco, P. (2004). *An Orange for Frankie*. New York, NY: Philomel Books.

Polacco, P. (2009). *Mrs. Katz and Tush*. New York, NY: Doubleday Books.

Roberts, R.M. & Kreuz, R.J. (1994). Why do people use figurative language? *Psychological Science, 5*(3), 159 – 163.

Roid, G.H. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education 7*(2), 159 – 170.

San Souci, R.D. (1989). *The Talking Eggs*. New York, NY: Dial Books.

Stagg, A. (2007). "She acted like she wanted to get an attitude with me": A narrative analysis of children's understanding of interpersonal conflict. Senior Honors Thesis, Rhodes College.

Tennessee Department of Education (n.d.). Writing Assessment scoring information. Retrieved June 7, 2011 from the Tennessee Department of Education's website: http://www.tn.gov/education/assessment/writing_scoring.shtml.

Twain, M. (1982). *The Adventures of Tom Sawyer*. New York, NY: J. Messner.

Wakely, M.B., Hooper, S.R., de Kruif, R.E.L., & Swartz, C. (2006). Subtypes of written expression in elementary school children: A linguistic-based model. *Developmental Neuropsychology*, *29*(1), 125 – 159.

Walton, M.D. & Brewer, C.L. (2002). The role of personal narrative in bringing children into the moral discourse of their culture. *Narrative Inquiry*, *11*(2), 307 – 334.

Walton, M. D., & Walton-Wetzel, J. A. (2008). *Reported speech and the development of narrative voice in middle childhood.* Georgetown University Round Table in Linguistics. Washington, D. C.

I give permission for public access to my Honors paper and for any copying or digitization to be done at the discretion of the College Archivist and/or the College Librarian.

Signed_____

Regan Humphrey

Date _____