Academic discourse: Dissociating standardized and conversational measures of language proficiency in bilingual kindergarteners

KATHLEEN F. PEETS
Ryerson University

ELLEN BIALYSTOK
York University

ADDRESS FOR CORRESPONDENCE
Kathleen F. Peets, Department of Early Childhood Studies, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada. E-mail: kpeets@ryerson.ca

ABSTRACT
This study examined the relationship between performance on standardized measures of language proficiency and conversational measures of the same features used in academic discourse among 24 monolingual and 25 bilingual kindergarteners. Academic discourse performance was considered for both its linguistic and its genre features in two discourse forms: narrative and explanation. Bilinguals performed more poorly than monolinguals on standardized measures of language proficiency, yet they performed similarly to monolinguals in the discourse-based linguistic and genre features. Moreover, genre features were more strongly related to linguistic features assessed through discourse than to standardized tests of these same features. These findings indicate that standardized measures of language proficiency underrepresent the abilities of bilingual children and that children's second language proficiency may be more accurately reflected in conversation.

Throughout North America, bilingual children speaking a variety of heritage first languages (L1s) are entering school with minimal experience in English as a second language (L2), yet it is through the English L2 that academic learning will occur. Therefore, children must acquire competence in *academic discourse* in order to succeed in English-language schooling. Academic discourse contrasts with everyday conversation in its formal conventions and relationships with the complex and often technical domains of education (e.g., literacy, science, mathematics, and history). Other related terms include *classroom discourse* (Cazden, 2001), *academic language* (Snow, 2010; Snow & Uccelli, 2009), and the *language of schooling* (Schleppegrell, 2001). Despite the importance of academic discourse

to learning, it is unclear what degree of proficiency is required to be able to effectively use language for academic purposes in the classroom. The present study examined bilingual and monolingual children's performance on both standardized and discourse-based measures of proficiency (vocabulary, morphology, and syntax) and two academic discourse genres that are related to academic success in the early years: narratives and explanations (Griffin, Hemphill, Camp, & Wolf, 2004; Miller et al., 2006; Tabors, Snow, & Dickinson, 2001). Understanding the academic discourse performance of bilingual children will provide a more accurate and ecologically valid perspective on their language use than do standardized tests allow. This understanding is especially critical to the question of academic achievement among L2 students for whom standardized assessments may not accurately represent ability levels.

English proficiency is typically measured through the use of standardized tests that involve the uniform administration of structured questions to children to assess language abilities, often with reference to the performances of a normative population based on monolingual speakers of English. The language demanded in these tests is highly decontextualized, in that children are typically required to perform such tasks as looking at a visual array or listening to a verbal prompt indicating or producing a correct response. Although these tests do provide important information about a child's language ability (targeting specific linguistic forms in isolation, comparison with peers), they do not evaluate children's actual use of language in conversational contexts such as the classroom. Although exclusive reliance on formal decontextualized tests can theoretically underrepresent the language abilities of all children, the problem is particularly crucial for bilingual children. The reason is that the kinds of errors in morphology and syntax that are often made by L2 learners or children becoming bilingual appear, on the surface, to be the same as those shown by children with language impairment. Consequently, standardized tests may misidentify a child with a language difference (as in home language is not English) as a child with a language disorder (Bedore & Peña, 2008; Genesee, Paradis, & Crago, 2004; Paradis, Emmerzael, & Duncan, 2010). Because of the lack of assessment tools available in many of the L1s spoken among North American children and the impracticality (and methodological concerns) of translating English tests, many researchers have argued for alternate means of language assessment for bilingual children. Among such alternatives, discourse analysis (Craig & Evans, 1992; Nettelbladt, Hannson, & Niholm, 2001; Thompson, Craig, & Washington, 2004) and parent interview (Paradis et al., 2010) have been suggested as more accurate, or ecologically valid, measures of language proficiency than standardized tests.

In a study of over 1,700 children between the ages of 3 and 10 years, bilingual children schooled in English as an L2 were shown to control a smaller vocabulary than their monolingual peers, a difference that was greater for home-based words than for school-based words (Bialystok, Luk, Peets, & Yang, 2010). These results are important because, among monolinguals, weak vocabulary is associated with weaker academic performance, particularly on verbal outcomes such as reading comprehension (Hemphill & Tivnan, 2008; Muter, Hulme, Snowling, & Stevenson, 2004; Ricketts, Nation, & Bishop, 2007). However, no research to date has investigated whether the documented difference in vocabulary development in the

language of schooling found for bilingual children has implications for their development of the use of English for academic purposes. In related research, core proficiency does not predict difficulty in narrative discourse (Bliss & McCabe, 2003; Fiestas & Peña, 2004), a genre relevant to the academic use of language in early childhood. An understanding of the relationship between proficiency and use of language for academic purposes is critical to informing not only our interpretation of standardized language assessment routinely administered to nonnative English speakers but also our understanding of what aspects of English ability are important to emphasize in L2 language acquisition to support academic success.

Early narrative skills in monolinguals have been associated with better linguistic (Bishop & Edmundson, 1987; Miranda, McCabe & Bliss, 1998) and academic (Bashir & Scavuzzo, 1992; Bishop & Adams, 1990; Fazio, Naremore, & Connell, 1996; O'Neill, Pearce, & Pick, 2004; Reese, Suggate, Long, & Schaughency, 2010; Westby, 1984) achievement. Personal and fictional narratives recount a series of connected events around a central theme, often with a conflict that is resolved by the end of the story (Labov & Waletzky, 1967; Peterson & McCabe, 1983). Both forms are relevant to the early childhood classroom context, with personal narratives being integrated into class events such as "Monday morning news" or "sharing time." In other words, although the personal narrative is part of conversational language use among adults, it is part of academic discourse in early childhood education, officially sanctioned in classroom activities. Moreover, appropriate topics allowable in the classroom context are typically made clear by the teacher (see Cazden, 2001). Fictional narratives may provide a more direct link than personal narratives to literacy-based forms such as the storybook and may therefore be more closely related to academics beyond early childhood education. Each of these story types, personal and fictional, present different challenges to a young child (Peets, 2008; Shiro, 2003).

Although difficulties in narration have been reported for children learning English as an L2 (Gutierrez-Clellan, 2002), bilingual children have produced equally strong fictional (Fiestas & Peña, 2004) and personal (Bliss & McCabe, 2003) narratives in their two languages. A study by Uccelli and Paez (2007) examining Spanish–English bilingual children found a relationship between Spanish vocabulary at age 4 and Spanish story quality at age 6. They also found that kindergarten Spanish narrative performance predicted first-grade English narratives, suggesting cross-linguistic transfer at the discourse level (Uccelli & Paez, 2007). However, none of these studies compared performance of bilingual children to matched monolinguals, so no developmental assessment or interpretation of the role of bilingualism in these outcomes can be made.

Compared with narrative research, far less work has been carried out on the development of explanatory discourse in child language, and of these studies (Berman, 2008; Gillam, Peña, & Miller, 1999; Nippold, Hesketh, Duthie, & Mansfield, 2005), none has addressed the challenge of producing oral explanations in one's L2 in the early years. Research with monolinguals in expository language use tends to examine older children and focus on the written, as opposed to oral, modality (Schleppegrell, 2001; Snyder & Caccamise, 2010). However, the presence of expository discourse can be found in the early childhood classroom in much of basic instruction. Teachers offer explanations to children as they teach

in any of the content areas of the curriculum, and explanations by children are expected in the introductions to the sciences (e.g., how seeds grow) and in informal class conversations (e.g., when a teacher prompts a child for what happened on a recent field trip by the class).

Foundational work by Snow and colleagues provides an exception to this pattern by investigating two types of expository discourse: definitions among monolinguals in elementary school (Snow, Cancini, Gonzalez, & Schriberg, 1989) and explanations at home among monolingual children (Aukrust & Snow, 1998; Beals & Snow, 1994). Snow and colleagues (1989) found that definitional skill was related to literacy and that bilingual children whose L2 was English showed no differences from their monolingual peers in English definition scores. In a study of dinner-table conversation in low-income American families, Beals and Snow (1994) found that explanations constituted an equal amount of talk as did narratives; although the explanations were shorter, they were more frequent. In a cross-linguistic study of American and Norwegian dinner-table conversation, Aukrust and Snow (1998) found that Americans were more likely to produce spontaneous explanations, whereas Norwegian families were more likely to produce spontaneous narratives. Therefore, the degree to which young bilingual children from diverse linguistic backgrounds have been exposed to explanation in the context of the home varies and may be culture and language specific. Of these studies, only the Snow and colleagues (1989) investigation of definitional skill compared the performance of monolinguals to bilinguals, and they found no differences between the groups.

The present study examined the production of narratives and explanations by monolingual and bilingual children for their genre and linguistic features and compared these productions to performance on standardized tests of linguistic features. Because no previous study has explored this question, the current study is largely exploratory, with the goal of identifying the relationships between various means of measuring linguistic and discourse performance among young monolingual and bilingual children. Specifically, the primary research question was whether the reported disadvantage in standardized measures of English proficiency among bilinguals was also reflected in academic discourse performance. If so, bilingual children would be at risk for compromised academic achievement because of the relation between these variables found for monolingual children. English proficiency was assessed in both standardized tests and discourse-based measures of vocabulary, morphology, and syntax. Following previous research, we hypothesized that bilingual children would obtain lower scores than monolinguals in standardized test performance on all three measures. We had no hypothesis about discourse-based linguistic features, because this is the first study to compare these two types of measurement. Discourse was examined through both familiar (personal narrative or personal routine explanation) and unfamiliar (picture book narrative elicitation or magnet task explanation) topics. Based on previous reports of bilingual strength in narrative discourse, we predicted that the bilingual group would perform equally well in comparison to the monolinguals in the discourse measures and better in the unfamiliar tasks, based on the context of their L2 acquisition being primarily in school, an unfamiliar context. This last prediction is based on vocabulary findings showing a strength in academic versus home contexts among diverse bilinguals (Bialystok et al., 2010). The interpretation of these

findings was that bilinguals are typically acquiring English in a school context, resulting in the acquisition of academic English in unfamiliar contexts, in contrast with the use of a heritage language in familiar, home contexts.

METHOD

Participants

Twenty-four monolingual and 25 bilingual kindergarten children, aged 5 years, 2 months (5;2) to 6;3, were recruited through school districts and daycares in middle-class neighborhoods in a large Canadian city. Four of the 25 bilingual children were born outside of Canada and arrived between the ages of 1;5 and 3;7. At each data collection site, both monolingual and bilingual children were recruited to ensure comparable populations in each sample. Only children whose parents reported no use or knowledge of an L2 were included in the monolingual sample, and only children whose parents reported *equal* or *greater* use of an L1 other than English at home combined with English vocabulary within 1 *SD* of the mean were included in the bilingual sample. This information was elicited through the use of a Language Background Questionnaire, in which parents ranked the child's use and proficiency in English and a heritage language across a range of oral and written domains in the home, and through a standard assessment of receptive vocabulary. The rationale for the use and proficiency exclusion criteria is based in the goal of examining academic language use by balanced bilinguals and not examining challenges to this use emanating from lack of L2 proficiency. In other words, the current study was focused on the bilingual population, not the English language learner (ELL) population. There were 11 different languages represented in the bilingual group (Urdu, Gujarati, Mandarin, Tamil, Arabic, Italian, French, Greek, Farsi, Russian, and Pashto). In total, 56 children agreed to participate, and of those, 7 children were excluded for not meeting usage criteria described above or being more than 1 *SD* below the mean in English vocabulary.

Measures

Maternal education was used as a proxy for socioeconomic status (SES) and was measured by a 5-point scale included in the Language Background Questionnaire. The points on the scale indicated levels of education: 1 = *no high school*, 2 = *a high school diploma*, 3 = *some postsecondary education*, 4 = *bachelor's degree*, and 5 = *graduate or professional degree*. Nonverbal IQ was measured using Raven's Colored Progressive Matrices (Raven, Court, & Raven, 1998). The results for these measures are reported in Table 1. Participants in the two language groups did not differ in age, $F < 1$; SES, $F < 1$; or IQ, $F (1, 47) = 2.55$, $p = .12$.

Standardized assessments of language proficiency were made for vocabulary, morphology, and syntax by psychology students who were trained in psychometric measurement. Receptive vocabulary was measured using the Peabody Picture Vocabulary Test—Third Edition, Form A (PPVT-III; Dunn & Dunn, 1997), productive morphology using the Wug test (Berko, 1957), and productive syntax

Table 1. *Mean score (standard deviation) for background variables by language group*

	Monolingual	Bilingual	*p*
Age	5.6 (0.44)	5.5 (0.37)	*ns*
SES (maternal education)	3.9 (0.93)	3.8 (0.94)	*ns*
Nonverbal IQ (Raven's percentile)	69.4 (24.7)	56.8 (30.4)	*ns*

Note: SES, socioeconomic status.

using the formulated sentences subtest of the Clinical Evaluation of Language Fundamentals—Fourth Edition (Semel, Wiig, & Secord, 2003).

In the PPVT-III, children are shown a plate containing four pictures, the experimenter names one of them, and the child must indicate the correct target picture. Testing continues to a criterion level according to procedures described in the manual, and scores are converted to standard scores through normalized tables based on age.

The Wug test was originally designed to test productive morphology by providing children with nonsense words in the context of a visual illustration and a series of short sentences. The final sentence in the sequence provides the context for a morphological marking such as past tense, plural, or continuous present. The child is then required to add the correct English morphological ending onto the nonsense words. The score is the total (out of 33) of correct responses produced by the child.

The formulated sentences task of the Clinical Evaluation of Language Fundamentals presents children with a picture of an activity and a target word presented orally. The child is required to generate a sentence about the picture using the target word. Scoring is based on a 3-point system in which 2 indicates that the target word was used in a syntactically correct way with direct reference to the picture, 1 indicates a single syntactic error, and 0 means that there was more than one error or the sentence did not reference the picture. Standardized test scores were not used because only one subtest was administered.

The target linguistic features (vocabulary, morphology, and syntax) were identified in academic discourse samples (described below) using Computerized Language Analysis (MacWhinney, 2000). Vocabulary was assessed by means of counts of total number of words and total different words. Morphology was assessed through calculation of morphological errors that were coded for each transcript using the same categories as found in the Wugs test: plurality, third-person singular, past tense, comparative/superlative, present progressive, and possessive. A ratio comparing the number of morphological errors to the total number of morphological forms found in the transcript was used as the discourse-based measure of morphology. Syntax was assessed by computerized language analysis's mean length of utterance and syntactic error scoring. Syntactic errors were coded for each transcript in the following error categories: omission (auxiliary, conjunction, or preposition), commission (auxiliary, conjunction, or preposition), and word

order, and syntax error rate was calculated as a proportion of total *t* units. These linguistic measures were calculated as a mean score across all the language samples produced by each child to ensure a higher rate of ecological validity because short samples are typically not reflective of developmental measures such as mean length of utterance.

Error coding was carried out by the first author, and reliability was carried out independently on 20% of the transcripts with a second coder who is a PhD student in speech–language pathology. Both coders were blind to language group. The transcripts used for the reliability assessment were selected randomly from each discourse task (frog story, personal narrative, magnet task, and home explanation), so all were equally represented. Interrater reliability was .81 for morphology and .84 for syntax. Although this reliability rate is only moderate, it should also be noted that the two coders also had a high agreement rate for the absence of errors, which is relevant in a sample in which low rates of errors were a key finding (see discussion). Of the 40 files used for reliability coding, 17 were identified by both coders as having no morphological errors and 6 were identified by both coders as having no syntactic errors.

Discourse tasks

Discourse topics were created through a 2×2 design that manipulated familiarity and discourse type. Familiar topics were related to the child's personal experience outside of the school context and operationalized through the use of conversational prompts about experiences (i.e., narratives) and routines (i.e., explanations). Unfamiliar topics were concerned with school-based activities that involved new stories (a wordless picture book in the case of narrative) and novel materials (magnets in the case of explanations).

For unfamiliar narratives, children were asked to generate a story using the wordless picture book *Frog, Where Are You?* (Mayer, 1969), a widely used measure of narrative ability (Berman & Slobin, 1994; Strömqvist & Verhoeven, 2004). Standard instructions were given and are described in Appendix A. For familiar narratives, personal-experience stories were elicited with prompts modeled after those designed by Peterson and McCabe (1983). This style of prompt was a conversational approach in which the investigator reported on a short event that involved some conflict or "problem" to solve (bee sting, accident, etc.), followed by the question to the child "Has anything like that ever happened to you?" Five such prompts were used (see Appendix A), and when the child responded "yes" to this prompt, the investigator would then say "Tell me about that time." Some children produced more than one narrative in the course of the conversational elicitations, in which cases the strongest (typically the longest) narrative was selected for scoring. Five children did not produce one of the four discourse types, creating missing data for those conditions. Specifically, one participant did not provide a home routine, one child did not produce a frog story, and three children did not produce a personal narrative. In all of these cases, the remaining tasks were completed by the children and were included in analyses.

For unfamiliar explanations, children were given a set of objects (e.g., coin, metal paper clips, elastic band, and wooden block), an oblong magnet to be used

as a tool, and some colorful magnets in various shapes (heart, circle, etc.). The task was modeled after a magnet task used by Tabors, Roach, and Snow (2001) to examine parental styles in explaining magnetism to children, but here the focus was on child talk instead of adult input. The child was instructed to "touch each of the objects with the magnet, and tell me what happens." Children would then observe a lack of magnetism (wooden block, nickel, or popsicle stick), magnetism (other smaller magnets or paper clips), or repulsion (manipulated by having the child turn the magnet to reverse polarities) as they carried out the task. Children were given neutral back-channel responses by the investigator as they spoke (hmm, uhhuh, etc.) or prompts to continue ("Tell me more about that" or "What else") with no active contribution or suggestion of content to the child beyond the original prompt. For familiar explanations, conversation-based home routines were elicited on the premise that the investigator was having a 5-year-old child as a house guest and needed tips on such things as what to cook and how to get ready for school (see Appendix B for a list of all prompts).

Procedure

Testing was conducted individually in a single session lasting approximately 45 minutes. Each data collection session was carried out in the child's school, in a quiet corner of the classroom or a small room adjoining the main classroom. The testing was administered in the same order to each participant: PPVT-III, personal narrative, home routine, Raven's Colored Progressive Matrices, Wug test, magnet task, frog story, and formulated sentences task.

Transcription and discourse scoring and analysis

All audio data were transcribed into *t* units (clauses and their dependent clauses) using Codes for Human Analysis of Transcripts conventions of the Child Language Data Exchange System (MacWhinney, 2000) by trained research assistants. Each transcript was checked against the audio recording by a second transcriber, and any disputes were resolved between the two transcribers.

Each transcript was then scored using a checklist for the presence of genre-specific features of narratives and explanations. The coding manual for all four discourse tasks is shown in Appendix C. Frog stories and personal narratives were both scored on a scale with a maximum score of 11 for the following features (adapted from Willenberg & Kang, 2001): opening, orientation, character delineation, resolution, first mention of story characters, causal markers, emotional states, mental states, resolution, anchor tense, and reported speech. The magnet task was scored on a scale with a maximum score of 9 (some features were worth more than 1 point) using the following features: explanations, physical properties, hypotheses/predictions, generalizations, applied knowledge, and empirical reports. Home routines were scored on a scale with a maximum score of 6 for the following features: temporal markers, task-related vocabulary, avoidance of temporal order violations, and nonspecific nouns. It is important to note that these assessments were made on the conceptual and organizational qualities of the discourse and not on the accuracy of the linguistic features used to express

Table 2. *Mean score (standard deviation) for PPVT-III, Wug test of morphology
(out of 33), and CELF formulated sentences (out of 48) standardized tests
of language by language group*

	Monolingual	Bilingual	*p*
PPVT-III standardized (vocabulary)**	108.5 (12.8)	97.0 (10.8)	.002
Wugs (morphology)**	13.3 (7.4)	7.7 (5.3)	.003
CELF formulated sentences (syntax)**	23.2 (8.9)	15.2 (8.6)	<.0001

Note: PPVT-III, Peabody Picture Vocabulary Test, Third Edition; CELF, Clinical
Evaluation of Language Fundamentals.
**$p < .01$.

them. Because the score for each discourse sample was based on a different maximum, genre features were converted to proportions so they could be compared across tasks. In other words, a score of 7 on one of the narrative tasks would be transformed to a proportion of 7 out of 11, or .63. In this way, comparisons could be made between a score of 7 on the narrative task and a score of 4 out of 6 on home routine, with a proportion of .66. The similarity in these scores is otherwise obscured by the difference in metric, so the transformation to proportions was used.

Scoring was carried out by the first author, blind to language group, and a qualified and trained second rater (also blind to language group) independently scored 20% of the data (i.e., 10 transcripts selected randomly within each condition). Interrater reliability was .86 for the frog story, .85 for the personal narrative, .90 for home routine, and .83 for the magnet task.

RESULTS

Table 2 presents the results from the formal assessments of language proficiency. Bilingual children obtained lower scores than did monolingual children on standardized tests for vocabulary (PPVT-III), $F (1, 44) = 11.35, p = .002$; morphology (Wug test), $F (1, 44) = 9.30, p = .003$; and syntax (formulated sentences task), $F (1, 41) = 8.91, p < .0001$.

The results for discourse-based linguistic features are reported in Table 3. For vocabulary, both number of words and number of different words in each discourse sample were calculated, with no difference between groups for either measure (Fs < 1). Syntax was assessed by both mean length of utterance and syntactic errors, and again there were no differences between groups in either measure (Fs < 1). For morphology, in contrast, bilinguals produced a higher proportion of errors than did monolinguals, $F (1, 47) = 4.95, p = .03$.

The mean proportion of genre features included in the discourse by language group are reported in Table 4. A three-way analysis of variance for discourse type (narrative or explanatory), familiarity of topic (familiar or unfamiliar), and language group was conducted on these scores. There were no language group differences on any of these measures, $F (1, 43) = 1.14$, *ns*, or interactions between

Table 3. *Mean score (standard deviation) for discourse-based linguistic features by language group*

	Monolingual	Bilingual	*p*
Vocabulary			
Total words	625.0 (364.0)	587.3 (224.3)	*ns*
Total different words	244.9 (102.5)	224 (70.2)	*ns*
Morphology			
Morphological errors*	4.3 (4.6)	7.4 (5.0)	.03
Syntax			
Syntactic errors	4.4 (3.4)	6.4 (5.1)	*ns*
Mean length of utterance	5.8 (0.8)	5.8 (0.9)	*ns*

*$p < .05$.

Table 4. *Mean (standard deviation) proportion of genre features included in discourse by discourse type by language group*

Discourse Condition	Monolingual	Bilingual	*p*
Narrative			
Personal story	0.42 (0.20)	0.38 (0.21)	*ns*
Frog story	0.44 (0.22)	0.38 (0.21)	*ns*
Explanatory			
Home routine	0.69 (0.19)	0.64 (0.24)	*ns*
Magnet task	0.57 (0.24)	0.54 (0.22)	*ns*

language and any other variable ($Fs < 1$). There was a difference between the two discourse types, $F (1, 43) = 6.40$, $p = .01$, and the two familiarity levels, $F (1, 43) = 5.14$, $p = .03$, that needs to be understood in terms of a significant two-way interaction of type and familiarity, $F (1, 43) = 45.85$, $p < .0001$. Paired-sample t tests were carried out on the two levels of familiarity with both the narrative and the expository discourse types. In narratives, there was no difference between the two levels of familiarity ($F < 1$), but there was a significant difference between the levels for explanations, $F (1, 47) = -2.3$, $p = .03$, with the home routine showing a higher score, on average, than the magnet task across the entire sample.

One of the goals of the study was to assess the impact of poor standardized test performance on academic discourse to determine the degree to which formal proficiency was related to the academic use of language. To this end, correlations were calculated to examine the relationship between standardized and discourse-based assessments of linguistic features and discourse scores. For this purpose, a composite discourse score indicating the average proportion of genre features across the four samples was computed for each child. There were no correlations between this composite discourse score and standard scores of syntax, $r (41) = .23$, *ns*, or morphology, $r (45) = .20$, *ns*, but there was a significant relation with vocabulary, $r (45) = .38$, $p = .01$. There were strong relations between discourse

Table 5. *Bivariate correlations of standardized test scores of morphology (Wugs) and syntax (formulated sentences) with discourse-based scores of morphological and syntactic errors (n = 49)*

	Errors in Discourse	
	Syntactic	Morphological
Formulated sentences Standard score	−.33*	−.36*
Wug test score	−.39**	−.63***

*$p < .05$. **$p < .01$. ***$p < .001$.

scores and total number of words, $r (45) = .43$, $p = .003$; total different words, $r (45) = .50$, $p = .0004$; and mean length of utterance, $r (45) = .46$, $p = .001$. There was no relation between the composite discourse score and the discourse-based measures of morphological error rate, $r (45) = −.19$, *ns*, or syntactic error rate, $r (45) = −.19$, *ns*. Thus, the composite discourse score was correlated with both standard and discourse measures of vocabulary but not with any measures of syntax or morphology. In order to compare the standardized test performance on language domains with their respective discourse-based scores, correlations were run on these two types of measures of morphology and syntax, and these are reported in Table 5. Because the test scores are positive numbers that reflect a level of "correctness" or accuracy by the children, the correlations reported in Table 5 are negative. That is, the fewer morphological and syntactic errors that the children showed in the discourse, the higher their Wug test and formulated sentences scores tended to be, on average. This pattern persists across linguistic domains, so the same negative relationship is seen when looking at the Wug test scores and syntactic errors, and the formulated sentences standard scores and morphological errors.

DISCUSSION

The present study examined four samples of academic discourse comprising two genres (narrative and expository) produced by monolingual and bilingual children, and related children's productions to both standardized and discourse-based assessments of linguistic features. The children in the two language groups did not differ in age, SES, or nonverbal IQ, and attended the same schools and lived in the same middle-class communities. In spite of producing significant group differences on standardized assessments of language, all the children performed similarly on four genre-specific measures of academic discourse. Although the measures of each linguistic domain, such as morphology and syntax, are related to one another, as evidenced by the negative correlations reported above, the children are nonetheless able to accomplish academic discourse goals despite core language difficulties.

A closer look at the range of performances in both groups on the tasks enables a better understanding of the dissociation between discourse and core language. In the frog story, the strong bilingual performance is nonetheless accompanied by grammatical errors (underlined):

Child: the dog looked in at the frog's bottle.
Child: and it was gone.
Child: and then . . . then the dog fell with the bottle on his head.
Investigator: oh.
Child: and that little guy weared his dad's shoes.
Child: and then # he saved the dog.
Child: and then the bottle broke.
Investigator: oh [laughs].
Child: then the little boy . . . he called out for . . . um the frog.
Child: and he didn't answer or come to him.
Investigator: oh.
Child: the little boy look in the hole.
Child: and the dog looked in the beehive.

In contrast, a weak bilingual performance showed no such grammatical difficulties:

Child: there was a frog.
Child: and a dog.
Investigator: mmhm.
Investigator: okay.
Investigator: and then what happens here?
Child: they're sleeping.
Investigator: mmhm.
Child: and they fell in the stream.
Child: at first . . . everyone [unintelligible].
Investigator: and what . . . what happens next?
Child: there's a dog on his head.
Investigator: mmhm.
Investigator: and what happens here?
Child: he . . . he took a shoe.
Child: and the dog's in the . . . thing.

In the magnet task, strong monolingual performance also shows some grammatical errors as well as some scientific inaccuracies, yet the task reveals many of the features of effective discourse structure that identify it is an explanation:

Child: this one doesn't stick.
Investigator: why?
Child: so does this one because it is not magneted.
Investigator: how does that work?
Child: because if it's not like this.
Child: it won't stick.

Child: it has to be like this plastic to stick like these.
Investigator: uhhuh.
Investigator: what's happening?
Child: they're sticking.
Investigator: how does that work?
Child: because if you just move one over to make some more space for it.
Child: and move it one more.

Another purpose of the study was to distinguish between effects of type of discourse (explanatory or narrative) and level of familiarity in the quality of children's productions and the role of bilingualism on those results. The groups performed similarly across the different discourse types, with comparable means and distributions of scores. As a general pattern across the entire sample, the children did not perform as strongly in the two narrative tasks as they did in the two explanatory tasks, with mean scores below 50% of the possible 11 points. The scoring system that was used adheres closely to formal narrative structure and has been used with many children older than those in the current sample, so the age of the current sample may explain their lower performance in these measures. The interaction of discourse type with familiarity level showed that the less familiar magnet task was more difficult for both groups, but this did not hold for the narratives. This potentially negative impact of topic unfamiliarity may reflect the overall difficulty level of talking about something as formal and school based as magnetism, but the finding also does not support the hypothesis that the bilinguals would perform more strongly in unfamiliar tasks than in familiar ones based on the contexts in which each language was acquired. Despite the fact that the school provides the context for much of the English language acquisition that bilingual children undergo, the difficulty level of scientific language may simply be a more powerful factor than context of L2 acquisition.

Consistent with previous research in vocabulary level, bilingual children obtained lower scores than monolinguals in all three of the standardized assessments. This profile is consistent with the morphosyntactic difficulties of language impairment (Rice & Wexler, 1996), demonstrating the risk of misidentification of a language difference as a language disorder. For vocabulary, the standardized PPVT-III scores were similar to those obtained from a sample of 1,738 children between the ages of 3 and 10 years drawn from the same population (Bialystok et al., 2010). In that study, the mean score was 106.8 (present study $M = 108.5$) for monolinguals and 96.3 (present study $M = 97.0$) for bilinguals. However, the results were different when proficiency measures were derived from children's use of these same features in producing discourse. In this case, children in the two language groups obtained similar scores for vocabulary and syntax, although bilingual children continued to obtain lower scores than did monolinguals on measures of morphology. It might be predicted that morphology errors would parallel syntax errors because they work together to form the grammatical system. Syntax, however, is more easily simplified than morphology, because complex forms can be avoided. In the case of morphology, conjugating a verb or using the correct case of a pronoun is compulsory for well-formed sentences. The finding may also be explained by the relationship between L1 and L2 morphology, but

that question is best addressed with a homogenous bilingual group who share the same L1. What is important about the current results is that two levels of language (vocabulary and syntax) showed no group differences in the context of academic discourse use. These results suggest that bilingual children are effectively using the proficiency that they have in an economical way.

The finding that bilingual children demonstrated equivalent control over vocabulary as compared to monolinguals in the discourse samples in spite of poorer performance on the standardized test is different from that reported by Yan and Nicoladis (2009). Their study examined older children, between the ages of 7 and 10 years, and reported comparable scores for the two groups on receptive vocabulary but lower scores for bilinguals on expressive vocabulary, opposite to the pattern in our data. However, differences between the studies in children's age and tasks used, as well as population differences, need to be considered to evaluate this difference. That is, we examined *younger* children's receptive (expressive not formally measured) vocabulary scores and found them to be lower in comparison with monolingual peers. Although we found no differences in the discourse-based measure of expressive vocabulary, the picture-naming measure used by Yan and Nicoladis, in spite of being an experimental task, resembles a standardized test in its administration and structure. We expect that such a task in our sample would also produce lower scores among bilinguals. Despite this possibility, it is critical to note that the Yan and Nicoladis sample is based on a homogenous French–English bilingual sample in a context of broad support for the use of the French L1 in the English-learning context. This support does not exist in the same way for the heterogeneous bilinguals represented in Canada's immigrant population and the children of immigrants, the population examined in the present study.

This difference in population is critical when considering the findings of the current study alongside those carried out among Spanish–English ELLs and bilinguals in the United States. For example, the current study found that the bilinguals scored within the average range on receptive vocabulary, whereas most studies of Spanish–English bilinguals and ELLs find the opposite (e.g., Fernandez, Pearson, Umbel, Oller, & Molinet-Molina, 1992; Uchikoshi, 2006). Although we excluded a small number of children who scored outside the average range in the PPVT-III to provide a more accurate and conservative test of our hypotheses, it is critical to note that the vast majority of children whose L1 is not English in this population do score within the average range in English receptive vocabulary. In the study of PPVT-III scores of more than 1,700 children by Bialystok et al. (2010), no children were excluded and the results showed average vocabulary score in the normal range for both language groups, with a significant shift toward higher scores for monolinguals.

Another distinguishing feature of the sample in the present study is its diversity. This heterogeneous sample reflects the neighborhoods in which the children live and is reflected in other research in Canadian urban centers (e.g., Jean & Geva, 2009; Lesaux, Rupp, & Siegel, 2007). This is in stark contrast to the strong Hispanic communities in Miami (Fernandez et al., 1992), in which many children are able to function in Spanish within their communities, and even in the American northeast, in which Uchikoshi reports that many children live in Spanish-speaking homes and communities. Increased English exposure has been associated with

higher English receptive vocabulary scores: Fernandez et al. report higher receptive vocabulary scores among bilinguals (71 standard score) as compared to Spanish-speaking ELL children (53 standard score), and Uchikoshi reports a dramatic increase among her sample of Spanish-speaking ELLs of 17 points in receptive vocabulary within one academic year. The results of the current study, therefore, indicate the developmental outcomes of bilingual children being raised in a context in which home and school languages are developing in different contexts in some isolation of each other.

There are two key implications of the dissociation between discourse measures and standardized language scores for bilingual children reported here. The first relates to the previous finding of a relationship between academic discourse and future academic achievement (Griffin et al., 2004; Miller et al., 2006; Tabors, Snow, et al., 2001). Given strengths in narrative and explanation skills, young bilingual children may *not* be at risk for academic difficulties that might otherwise have been predicted by their standardized test performance alone. We acknowledge, though, that this study examined very young children and that the same pattern of results may not hold in later years. The demands of academic discourse change throughout development, with an increasing emphasis on explanatory talk and more complex and literate language forms (Nippold, 2010; Snow, 2010). Extending this line of research throughout development into the elementary and secondary years would help not only to clarify English acquisition in the context of schooling for young bilinguals but also to inform the relationship between this development and academic achievement.

The second implication is for assessment and the potential for misidentification of language impairment. The limitations of standardized tests have been identified in the case of bilingual children with language impairment (Bedore & Peña, 2008; Peña & Gillam, 2000), and the present study of typically developing bilinguals has revealed similar limitations. The present results support the notion that discourse analysis is a more ecologically valid measure of bilingual children's language ability than are standardized tests alone. Moreover, the validity of monolingual norms that are used in standardized tests may also be problematic. For example, in the case of vocabulary, the bilingual children were performing, on average, within 1 *SD* below the mean, which is a pattern reported across ages among bilinguals (Bialystok et al., 2010). This pattern suggests that a more ecologically valid standardized measure of vocabulary would involve comparing bilinguals to bilinguals, rather than monolinguals; that is, it would involve developing bilingual norms for many of our widely used standardized tests of language. Naturally, there is heterogeneity of bilinguals in North America, and subgroup differences as a function of language would not be possible in developing such broadly defined "bilingual" norms. Nonetheless, as long as bilinguals show homogeneity in aspects of standard score performance, such as has been reported here, then comparing them to one another will minimally offer a more equitable solution than comparing them to monolinguals.

An educational implication of these results is that full L2 proficiency may not be necessary in order to succeed in school-related language tasks or perhaps in school achievement. The dissociation between standardized versus discourse-based measures of vocabulary and syntax suggest that the abilities of bilinguals in

compensating for disadvantages in linguistic features of language extend beyond skills in the social use of language. Instead, this dissociation, although not complete (in the case of morphology), suggests that bilinguals may be efficient and goal-driven language learners who are effectively doing "more with less."

APPENDIX A: NARRATIVE ELICITATIONS
Instructions used for fictional narratives

> Here I've got a book that tells a story, but that doesn't have any words. You don't have to know how to read to be able to read this book. The story told in this book is told all in pictures (show the child the book, turning slowly through the pages).

I would like you to tell me the story that you see in the pictures.

Standard prompts/feedback

- Neutral backchannel responses such as "uhhuh," "oh," and "mmhm."
- If the child has difficulty telling the story: "What else happens in the story?" "Tell me more . . . ," or "What happens on this page?"

If the child describes the picture instead of telling the stories: "Yes, that's what's in the pictures. Can you tell me the story you see in the pictures?"

Prompts used in the elicitation of personal narratives

1. "I took my cat to the vet last week, and when she was about to have her shots, she ran all over the room and the vet had to chase her. Do you have a cat or another pet?"
2. "One time I was on a trip with my family, and we were driving along and then all of a sudden our tire blew up and we went into a ditch."
3. "When I was at the dentist's office, there were these twin boys and they were about 5 years old. One of them was looking at a book and the other one wanted the same book, so they fought over it. Their mother had to separate them. Do you have any brothers or sisters?"
4. "One time when I was at the park, I was playing with my friends when all of a sudden a bee stung me! Has anything like that every happened to you?"
5. "Last summer I went on a holiday to the beach. There were lots of waves and one of them knocked me over. We had a lot of fun. Have you gone on a summer vacation before?" [Where? Can you tell me about that time?]

APPENDIX B: EXPLANATION PROMPTS
Prompts used for elicitation of home routines

> First elicitation: "My niece/nephew is coming to visit me! I don't know what to make him/her for dinner. Do you have any ideas?"

> If the child does not respond, follow up with: "What is your favorite dinner that you eat with your family?"

If/when the child responds: "That sounds yummy! How do you make that, do you think?"

If child starts in the middle of the process, you can redirect to provide more detail with: "What do you do first?"

If the child does not know how to make it: "Well how do you think—what things do you need?"

If the child says "My mother makes it": "Ok, what does she do? How does she start?"

Continue with backup questions such as "What else do you think you need to do?" or "What does he/she do next?"

If the child does not respond or responds only minimally, move to the next elicitation strategy.

Second elicitation (use only if first strategy is unsuccessful): "I also have to wake my niece/nephew up in the morning and bring him/her to school. How do you do that every day?"

If child does not respond, follow up with: "What do you do first?"

If/when child responds, follow with questions to elicit elaboration: "How do you do that?" "Why do you do that?" "What happens next?"

Prompts used for magnet task

"I have some objects here—can you tell me what they are?" [Point to each object and have child name it, if unable to name it, label the item for him/her.]

"Here is a magnet." [Hand child green rectangular magnet.]

"Touch the magnet to these different things, and tell me what happens."

As the child is experimenting with the magnet, ask him to explain what is happening if he is not already doing so, with prompts such as

"What happened when you touched the star?"

"Which ones do not stick?" [Wait for child response.] "Why?"

"What does the red/yellow/star etc. one do? What happens if you turn it the other way around?"

"How does that work?"

"What else can you do with the magnets?" "Tell me how you do that."

Engage the child in this way for approximately 5 minutes.

APPENDIX C: CODING MANUAL WITH EXAMPLES

This manual is to be used in conjunction with scoring sheets developed for the academic discourse study. The categories for each score sheet are included in this document and explained in more detail than is possible on the score sheet itself.

It is important to note that when scoring the home routine and the magnet task, accuracy of information is not the goal. Rather, the use of language in a scientific structure, providing the relevant types of response, is more important than whether the child knows anything

about magnetism or cooking. (The idea here is that it is challenging to explain these things, so children will have to rely on fairly complex language to meet this challenge.)

1. Frog story and personal narrative (see Willenberg & Kang, 2001):
   Opening: "It's about" "Once upon a time" "One day"

- A formal opening to a story should let the listener know that the story is about to begin and does not have to be conventional.

Orientation: characters, timeframe, or setting "One day last summer"

- Contextual information related to the characters, such as "Once there was a boy, a dog, and a frog . . ." or "I have this neighbor . . .". What is not considered orientation is if characters are partially mentioned (e.g., two out of three) or if they are mentioned using a definite article (this is not appropriate orientation, because we do not know the specific characters they are referring to).
- Contextual information related to when the story occurred, such as a day of the week, time of year, season, time of day—anything at all related to timeframe
- Contextual information related to setting, such as where the story took place, as in "the boy was in his room" or "we were at summer camp."

Character delineation: unambiguously presenting characters

- To assess lack of ambiguity, you must look at the pronouns to see if there is any lack of clarity.
- This judgment is not subjective. If there is a "he" and the child has just referred to two possible characters, then it is ambiguous, for example, "The boy and the frog looked out the window. Then he fell," which contrasts with "The frog looked out the window, then he fell and the jar broke."

First mention of story characters: not assuming prior knowledge

- Assuming prior knowledge is when a child uses "the" in the first mention, as in "the boy was just waking up."
- In a personal narrative, this would refer to if a child simply starts talking about someone by name who they have not identified to you, as in "David came over to my house . . ." (when you don't know who David is).

Causal markers (structural/linguistic consideration): because, therefore, but also causal "and" and "so"

- Although one can do a simple search for "because," it is more difficult to assess "and" as a causal connective. To do so, determine if there is a causal relationship between the clauses, as in "the frog leaned over the edge of the window too far and then he fell." "So" is also problematic because it could be used as an additive connective, much like, "and," "then," and "and then."
- "So" is considered to be causal if it connects two events in a situation where one event is the result of a preceding event; for example, "The bear climbed the tree so he could get the kite." However, in the following example, "so" is not a causal connective: "Some bears

were playing in the field, so they decided to fly their kite." In this case, "so" functions like "and then," indicating the temporal ordering of events.

Emotional states: including "fun"

- Any mention of any emotion will count in this category, which also extends to the use of the word "fun" and "scared," "frightened," or "angry."
- Any mention of the expression of emotion, including verbs such as "cry," "frown," "smile," and so forth.

Mental states: thought, wondered, planned, wanted

- Any mention of a word that refers to something going on inside the head of a character, as in the examples above, is counted as a mental state word. This could also include things like "remembered" and "imagined."

Defeat of expectations: negative assertion, as in "contrary to what you'd expect. . . ."

- Although usually signaled by a negative construction, a defeat of expectations could also be indicated by the adversative connective, but, for example, "He wanted to climb the tree, but he got stuck."
- In addition to using "but," children can also defeat expectations with an "even though" construction, as in "Even though the boy was tired, he kept looking for his frog." The meaning here is that you expect one thing to happen, yet another does, much to your surprise. This is a tool in effective storytelling.

Resolution: "they lived happily ever after"

- Events that constitute a resolution are those that involve action taken to remedy the situation or attempt to restore the natural order of things.
- Again, as with the opening, the resolution need not be a fairytale conventional one. It could simply be something like "then the cat was happy because it got to go home" or "they waved goodbye to the frog family."

Anchor tense: strict use of past tense, with exception of quotations

- As long as there is no present or future tense used (except if a character is being quoted), then the child gets a point for using the anchor past tense.

Reported speech

- Direct quotations of story characters, as in "she said 'don't do that!'" but NOT "she said not to do that." In the frog story, there may be many variations on this, such as "he was like 'whooooo!'" These are perfectly acceptable.

2. Magnet task

Explanations

- These are at the core of what this genre is about and involve the child giving information about the why or how things work.

- Responses will typically involve causatives, but these will also be counted separately, below.
- A contribution is not an explanation if it involves restating what was observed, as in "it sticks because it sticks," but "it sticks because it's a magnet" would be scored as an explanation.

Observation of physical properties

- When a child describes any of the objects based on how they look as opposed to how they function, then it counts as a physical property.
- These include both adjectives, such as color, texture, weight, and so on, and comparisons between the object and something else that it might resemble, as in "this magnet looks like a key," "this one is blue," and so on, and "materials," such as wood, plastic, and metal.
- It does not include the labels of the objects, which have previously been provided by the investigator: star, circle, ring, paperclips, block, elastic band, popsicle stick, heart, stick (the magnet they use to touch objects with is in the shape of a short stick or a long rectangle).

Hypotheses/predictions

- As in regular science talk, people often say what they think will happen before it actually does, and these predictions get credited here.
- For example, "the paper clips will stick," "those will not stick," or "the block won't work."

Generalizations

- Another feature of science talk, generalizations state things that apply "across the board" or in many situations, regardless of specifics. They are similar to rules or summaries of how things work.
- For example, "magnets stick to things," "magnets don't stick to wood," or "magnets only stick to magnets" (regardless of accuracy).
- Be careful to ensure that there is a statement made about something beyond the "here-and-now," that is, that a child isn't simply saying "it sticks because it's X" (this would be an explanation). An example of a generalization would be "metal sticks to magnets" (as in all metal).

Theory of magnetism

- General statement about what either sticks or does not stick to magnets and why (regardless of accuracy, as long as logic is not circular).
- Theories go beyond generalizations because they also involve an explanation of why things are the way they are. A generalization might be "magnets stick to hard things," but it becomes a theory if the child says "because hard things are metal."
- Here the critical difference between this category and generalizations is that information is offered as to why the generalization exists. To build on the previous example to make it into a theory, an example would be "Metal sticks to magnets because it is heavy."

Applied knowledge

- This information often follows the prompt "is there anything else you can do with the magnets?"

- Applied knowledge is going beyond the magnets in front of them to any use outside of the current interaction (e.g., they're used in machines, you can put them on the fridge, you can make things with them, etc.).
- For example, "I have magnets on my fridge."
- Note that this does not necessarily involve scientific content, but it shows an understanding of objects beyond the here-and-now and beyond their current application/use (it is a form of generalizing knowledge). Anything like "I can make this into X," where X = a Christmas tree, a car, and so on, counts as applied knowledge.

Empirical report

- Sometimes children may offer a commentary on their "research process," starting from what they are doing and finishing with what the results are.
- The child carries out observations in a systematic way (i.e., as a demonstration of contrasting effects) and reports "X things stick, but X things do not."

3. Home routine

Temporal markers and markers of habituation

- Any indication of time, for example, "first," "and then," or "after," as well as habituation such as "usually," "every day," or "always."

0 = absence of temporal markers
1 = temporal markers limited to "then" and "and then"
2 = temporal markers beyond (but could include) "then" and "and then"

Task-related vocabulary

- Credit is given for providing specific vocabulary that relates to the topic of explanation, and this is different for each type of explanation (dinner vs. routine)

0 = no mention of ingredients (e.g., milk, marshmallows) or tools (bowl, blender, oven, etc.), or excluding general name of dinner, as in "chicken," "macaroni," "pizza," and so on.
1 = mention of either ingredients or tools
2 = mention of both ingredients and tools

- Morning routine

0 = no mention of either task-related objects (toothbrush, cereal, boots, bath, etc.) or manner of travel (car, bus, walk, etc.)
1 = mention of either task-related objects (toothbrush, cereal, boots, etc.) or manner of travel (car, bus, walk, etc.)
2 = mention of both task-related objects (toothbrush, cereal, boots, etc.) and manner of travel (car, bus, walk, etc.)

No temporal order violations

- This category and the last involve the absence of features that violate the genre of explanations. In these cases, if you do not find evidence of them, you give full marks.

- Where a child has independently provided a temporal sequence that is logically correct, score 1. Where the timeline has not been clearly established, score 0.
- Examples of permissible violations of temporal order, where they are intentional by the child, and not simply errors.

An interjection of an exception would be "usually I get up and do *X*, but yesterday I did *Y*," followed by a continuation of the "usual routine."

An interjection of a personal anecdote would be "one time my uncle took me to school and he got lost," followed by a continuation of routine.

- Examples of a violation that must be scored as an error: "I get up, I get dressed. Then I go downstairs and eat my breakfast and put my lunch in my knapsack. I have a bath."

No use of nonspecific nouns

- Nonspecific nouns are typically used by people when they cannot recall the name of an object; instead, they substitute a generic pronoun in its place. Any instance of this type of generic pronoun in favor of a specific noun will result in 0 points.
- For example, "stuff" and "thing."

ACKNOWLEDGMENTS

REFERENCES

Aukrust, V. G., & Snow, C. E. (1998). Narratives and explanations during mealtime conversations in Norway and the U.S. *Language in Society, 27*, 221–246.

Bashir, A. S., & Scavuzzo, A. (1992). Children with language disorders: Natural history and academic success. *Journal of Learning Disabilities, 25*, 53–65.

Beals, D. E., & Snow, C. E. (1994). "Thunder is when the angels are upstairs bowling": Narratives and explanations at the dinner table. *Journal of Narrative and Life History, 4*, 331–352.

Bedore, L. M., & Peňa, E. D. (2008). Assessment of bilingual children for identification of language impairment. *International Journal of Bilingual Education and Bilingualism, 11*, 1–29.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Berman, R. A. (2008). The psycholinguistics of developing text construction. *Journal of Child Language, 35*, 735–771.

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*, 525–531.

Bishop, D., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31*, 1027–1050.

Bishop, D. V., & Edmundson, A. (1987). Language-impaired 4-year olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156–173.

Bliss, L. S., & McCabe, A. (2004). Narratives from Spanish-speaking children with impaired and typical language development. *Imagination, Cognition and Personality, 24*, 331–346.

Cazden, C. (2001). *Classroom discourse: The Language of teaching and learning*. Portsmouth, NH: Heinemann.

Craig, H. K., & Evans, J. L. (1992). Pragmatics and SLI: Within-group variations in discourse behaviors. *Journal of Speech and Hearing Research, 36*, 777–789.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Fazio, B. B., Naremore, R. C., & Connell, P. J. (1996). Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech and Hearing Research, 39*, 611–624.

Fernandez, M. C., Pearson, B. Z., Umbel, V. M., Oller, D. K., & Molinet-Molina, M. (1992). Bilingual receptive vocabulary in Hispanic preschool children. *Hispanic Journal of Behavioral Sciences, 14*, 268–276.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*, 155–168.

Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore, MD: Brookes.

Gillam, R. B., Peña, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders, 20*, 33–47.

Griffin, T. M., Hemphill, L., Camp, L. & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language, 24*, 123–147.

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

Hemphill, L., & Tivnan, T. (2008). The importance of early vocabulary for literacy achievement in high-poverty schools. *Journal of Education for Students Placed at Risk, 13*, 426–451.

Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics, 30*, 153–185.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Jehlm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington Press.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*, 821–834.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs* (3rd ed.). Mahwah, NJ: Erlbaum.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Press.

Miller, J. F., Heilman, J., Nockerts, A., Iglesias, A., Fabiano, L., & Frances, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*, 30–43.

Miranda, A. E., McCabe, A., & Bliss, L. S. (1998). Jumping around and leaving things out: A profile of the narrative abilities of children with specific language impairment. *Applied Psycholinguistics, 19*, 647–667.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665–681.

Nettelbladt, U., Hansson, K., & Nilholm, C. (2001). Why ask questions? Contextual effects on grammatical structure in the language production of children with specific language impairment. *Child Language Teaching and Therapy, 17*, 89–106.

Nippold, M. A. (2010). Explaining complex matters: How knowledge of a domain drives language. In M. A. Nippold & C. M. Scott (Eds.), *Expository discourse in children, adolescents, and adults: Development and disorders* (pp. 41–61). New York: Psychology Press.

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents and adults. *Journal of Speech, Language, and Hearing Research, 48*, 1048–1064.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test—Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language, 24*, 149–183.

Paradis, J., Emmerzael, K., & Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders, 43*, 474–497.

Peets, K. F. (2008). *Narrative elicitation styles: Differences between picture-prompted narratives and personal anecdotes.* Poster presented at the Georgetown University Round Table on Linguistics, Georgetown University.

Peña, E. D., & Gillam, R. B. (2000). Dynamic assessment of children referred for speech and language evaluations. In C. Lidz & J. G. Elliott (Eds.), *Advances in cognition and educational practice* (Vol. 6, pp. 543–575). New York: Elsevier Science.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum Press.

Raven, J. C., Court, J. H., & Raven, J. (1998). *Colored Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing, 23*, 627–644.

Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment. *Journal of Speech, Language, and Hearing Research, 39*, 1236–1257.

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*, 235–257.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*, 431–459.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). Toronto: Pearson.

Shiro, M. (2003). Genre and evaluation in narrative development. *Journal of Child Language, 30*, 165–195.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*, 450–452.

Snow, C. E., Cancini, H., Gonzalez, R., & Shriber, E. (1989). Giving formal definitions: An oral language correlate of school literacy. In D. Bloome (Ed.), *Classrooms and literacy* (pp. 223–249). Norwood, NJ: Ablex.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York: Cambridge University Press.

Snyder, L., & Caccamise, D. (2010). Comprehension processes for expository text: Building meaning and making sense. In M. A. Nippold & C. M. Scott (Eds.), *Expository discourse in children, adolescents and adults: Development and disorders* (pp. 13–39). New York: Psychology Press.

Strömqvist, S., & Verhoeven, L. (Eds). (2004). *Relating events in narrative: Vol. 2. Typological and contextual perspectives*. Mahwah, NJ: Erlbaum.

Tabors, P., Roach, K., & Snow, C. E. (2001). Home language and literacy environment: Final results. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and at school* (pp. 111–138). Baltimore, MD: Brookes.

Tabors, P., Snow, C. E., & Dickinson, D. K. (2001). Homes and schools together: Supporting language and literacy development. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and at school* (pp. 313–334). Baltimore, MD: Brookes.

Thompson, C. A., Craig, K., & Washington, J. A. (2004). Variable production of African American English across oracy and literacy contexts. *Language, Speech, and Hearing Services in Schools, 35*, 269–282.

Uccelli, P., & Paez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools, 38*, 225–236.

Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergartners: What are the best predictors? *Bilingualism: Language and Cognition, 9*, 33–49.

Westby, C. (1984). Development of narrative language abilities. In G. Wallach & K. Butler (Eds.), *Language and learning disabilities in school-age children* (pp. 103–127). Baltimore, MD: Williams & Wilkins.

Willenberg, I., & Kang, J. (2001). *Bear story coding manual*. Unpublished manuscript, Harvard University.

Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism, Language, and Cognition, 12*, 323–335.