Oral narrative retelling among emergent bilinguals in a dual language immersion program

Audrey Lucero

To cite this article: Audrey Lucero (2016): Oral narrative retelling among emergent bilinguals in a dual language immersion program, International Journal of Bilingual Education and Bilingualism, DOI: 10.1080/13670050.2016.1165181

To link to this article: http://dx.doi.org/10.1080/13670050.2016.1165181

Published online: 24 Mar 2016.
Oral narrative retelling among emergent bilinguals in a dual language immersion program

Audrey Lucero

College of Education, University of Oregon, Eugene, OR, USA

ABSTRACT
Oral narrative retelling is a complex linguistic and cognitive task that has been shown to map onto reading fluency and comprehension. Therefore, it is important to understand oral retelling skill, especially among emergent bilingual children – those who are learning two languages simultaneously. In this article, exploratory quantitative and qualitative findings from a study investigating the bilingual narrative retelling abilities of young heritage Spanish-speaking emergent bilingual children are reported. Kindergarten, first, and second grade children (N = 65) were assessed in each language separately using comparable wordless picture books, and their performance within and across languages was investigated. Two findings emerged: (1) children’s performance on retelling was significantly related across languages; but (2) individual children exhibited different patterns of bilingual strengths and challenges. These findings underscore the value of conducting academic oral assessments in both languages. In particular, the oral language proficiency of some children would be underestimated if their performance in only one language was considered. Implications for instruction and assessment are discussed.

ARTICLE HISTORY
Received 25 September 2015
Accepted 25 February 2016

KEYWORDS
Oral narrative retell; emergent bilinguals; language development; dual language education; primary grades

Introduction

Children who speak a language other than English at home are a growing population in US schools (Goldenberg 2008; Suárez-Orozco, Suárez-Orozco, and Todorova 2008). A majority of these children are Spanish speakers and a subset are enrolled in dual language immersion (DLI) programs, where they receive content instruction in both Spanish and English. Such children are considered emergent bilingual because of their potential to develop bilingualism and biliteracy (García and Kleifgen 2010). Despite a vast body of research on various aspects of DLI education (Valdés 1997; Lindholm-Leary 2004/2005; Palmer 2007; deJong and Howard 2009), little is currently known about the language and literacy patterns emergent bilinguals children exhibit in each of their languages (Gutiérrez-Clellen 2002; Uccelli and Paéz 2007). Therefore, researchers have begun to advocate the systematic assessment of oral language among emergent bilinguals, especially in the key areas of listening and retelling (Muñoz et al. 2003; Fiestas and Peña 2004; Miller et al. 2006). Oral vocabulary and narrative ability have been found to be important precursors to literacy for both monolingual and bilingual children (Bishop and Adams 1990; Dickinson and Tabors 2001; August and Shanahan 2006) and are therefore fruitful areas for further investigation (Gutiérrez-Clellen et al. 2000).
Oral narrative retelling

One promising language assessment is oral narrative retelling (ONR), a complex linguistic and cognitive task that has been used extensively with emergent bilinguals in recent years (Muñoz et al. 2003; Miller et al. 2006; Dubasik and Wilcox 2013; Rojas and Iglesias 2013). The ONR task requires children to understand a story read aloud and organize a coherent, detailed retell (Gutiérrez-Clellen 2002; Bedore et al. 2010; Hipfner-Boucher et al. 2015). The processing demands are similar to those needed to comprehend written text in that retelling involves interpreting and reconstructing a story using appropriate event sequencing and vocabulary (Roth et al. 1996). More than 20 years of scholarship has highlighted the relations between oral narrative skills and academic success (see Boudreau 2008 for a review). Specifically, narration may serve a bridging function between oral language and written literacy by enabling a young child to become familiar with the kind of extended contextualized discourse they will later encounter in written texts (Gagarina et al. 2015).

ONR for emergent bilinguals

There are a number of widely cited benefits to using ONRs to assess emergent bilinguals. First, the ONR provides children with a familiar, authentic context in which to use language (Spycher 2009); fictional stories are commonly told and read aloud to children at home and in school, and children are generally expected to be proficient in retelling at an early age (Boyd and Naucler 2001). Narratives are widely shared between parents and children across cultural groups (Melzi, Schick, and Bostwick 2013), and are common in US households. Therefore, by first grade, ‘most children are … aware of common organizational patterns in stories and can use this knowledge to support comprehension’ (Garner and Bochna 2004, 69). Second, the task is not dependent on decoding, so narratives can be readily compared across two languages (Snow et al. 1995). Third, retelling is considered less biased against bilingual children than norm-referenced tasks because it is not normed on monolingual children (Gutiérrez-Clellen 1998; Reese, Sparks, and Suggate 2012; Gagarina et al. 2015). Finally, narrative analysis allows teachers to assess both microstructural and macrostructural components of language in context using a single task (Heilmann et al. 2010; Heilmann, Miller, and Nockerts 2010). This is especially beneficial for emergent bilingual children. For example, we might expect less proficient vocabulary and morphosyntax (microstructure) in the second language of these children, but they likely show similar strengths in macrostructural skills, such as the overall ability to construct coherent event sequences (Berman 2001; Pearson 2002; Fiestas and Peña 2004). The ONR assessment allows children to show strengths at either level.

Some large-scale research has found that bilingual children score similarly on macrostructure across languages even when one language is considered weaker than the other (Gagarina et al. 2015). In addition, at least one study has shown a robust relationship between ONR performance and other reading skills among heritage Spanish-speaking emergent bilinguals (Miller et al. 2006). In a study of 1531 kindergarten through third graders in transitional bilingual classrooms, Miller et al. (2006) investigated relations between retelling, reading comprehension, and decoding in Spanish and English. They found children’s microstructure and macrostructure
retelling abilities to be significantly predictive of decoding and reading comprehension both within and across languages. This focus on reading outcomes is noteworthy, but it does not provide insight into emergent bilingual children’s patterns of oral language proficiency in both languages.

Therefore, there remains a need to understand, ‘the typical developmental trajectories and individual differences of children learning two languages … in order to develop educational programs that can most effectively meet their needs’ (Dubasik and Wilcox 2013, 163). This is especially true when considering the different communicative contexts and discourse genres of school in which children must exhibit proficiency (Berman 2001; Barnes, Kim, and Phillips 2014). A number of studies have investigated the performance of individual bilingual children on spontaneous oral narratives (Montanari 2004; LoFranco, Peña, and Bedore 2006; Iluz-Cohen and Walters 2012), but fewer have examined the retelling abilities of individual children (Gutiérrez-Clellen 2002; Westerveld 2014). Retelling places different cognitive and linguistic demands on children than spontaneous narratives because of the need to process extended discourse and re-formulate it into a coherent story (Ripich and Griffith 1988; Lever and Sénéchal 2011). One prominent study that examined individual performance on Spanish–English bilingual narrative retelling found that second graders differed in terms of which language was stronger and how closely related their retelling abilities were across languages (Gutiérrez-Clellen 2002). The other study, conducted with Samoan–English four-year-old bilinguals, found that considering a child’s narrative proficiency in only one language significantly underestimated their overall narrative abilities, as performance on narrative macrostructure was not significantly correlated across languages (Westerveld 2014). The goal of the current study was to build on previous research by elucidating patterns of bilingual performance among early school-age Spanish–English bilingual children enrolled in DLI education.

The present study

The purpose of this article is to report findings from an exploratory study examining the bilingual ONR abilities of heritage Spanish-speaking emergent bilingual children enrolled in a 50:50 DLI program. Scholars have previously noted the importance of attending to oral language and literacy skills in both of the languages a bilingual child speaks (Manis, Lindsey, and Bailey 2004; Miller et al. 2006; Gottardo and Mueller 2009). Doing so helps identify the academic oral language skills that are potentially related across languages and those that may not be (Goodrich, Lonigan, and Farver 2013) and improves our theoretical understanding of the contributions that first (L1) and second (L2) language skills make to the development of literacy (Droop and Verhoeven 2003; Simon-Cereijido and Gutiérrez-Clellen 2009). The practical goal of this research is to inform instructional and assessment decisions in DLI programs. The specific research questions guiding this study were:

1. How were microstructural and macrostructural components of emergent bilingual children’s narratives related across languages?
2. What patterns of bilingualism did children exhibit on oral narrative retells?

Given the literature base on which this research was premised, two hypotheses were developed. First, it was expected that there would be significant cross-linguistic correlations on ONR components overall (Pearson 2002; Bedore et al. 2010; Lucero 2015). However, it was also expected that individual children would exhibit different patterns of bilingual strength and challenge, with some children showing high (or low) levels of proficiency in both languages, and others showing discrepancies in performance on the two languages (Gutiérrez-Clellen 2002).
Method

Participants

Participants were 65 kindergarten, first, and second grade children attending a Spanish–English 50:50 DLI program in a mid-sized city in the Northwestern USA. The school enrolled 300 children in grades K-5, approximately 52% of whom were white and 40% of whom were Latino (Oregon, Department of Education 2013). Seventy-one percent of students were considered economically disadvantaged and approximately 20% were eligible for English as a second language (ESL) services. The school enrolled a considerably higher percentage of Latino students and those eligible for free and reduced lunch than the district averages (14% and 43%, respectively). It was located in a working class area, and most of the children lived within the school’s neighborhood boundaries. The DLI program followed a 50:50 model, such that all children received 50% of their daily instruction in each language. In K-2 classrooms, there were approximately an even number of heritage Spanish and English speakers. Literacy instruction was conducted in both Spanish and English at all grades, following a core curriculum (McGraw-Hill 2014).

The school followed the district procedure for identifying those in need of ESL services: at the beginning of the year, all parents in the school completed a home language survey indicating the primary language(s) spoken in the home. If a language other than English was indicated, the child was assessed using the English IDEA language proficiency test (IPT) 1 Oral Language Proficiency Test (Ballard and Tighe 1999). Children qualified for ESL services if they were designated non-English speaking or limited English speaking according to the IPT scoring criteria. Children who were designated fluent English speaking using IPT criteria did not qualify for ESL services.

Most participating children qualified for and received English language learner (ELL) services (N = 49), and received daily push-in or pull-out English language development instruction with an ELL teacher or a bilingual instructional assistant. A smaller number either did not qualify or had already exited services by the time the study was conducted (N = 13). ESL qualification data was missing for three children. All kindergarten, first, and second graders whose parents identified them as ‘speaking Spanish as a first language’ on a home language survey were given consent forms. Sixty-five children returned signed forms and were consented in to the study. Nothing was known about the amount of each language spoken in the home, but all participating children had been enrolled in the DLI program at least since the beginning of the school year in which assessments were conducted (Table 1).

Because of the small number of Spanish-speaking children in this community, data was collected in two waves (spring 2013, 2015) in order to reach the desired number of participants. All participating children were assessed at only one time point and were able to complete the ONR assessment in either Spanish or English. However, 10 children were unable to complete the assessment in both languages due to low proficiency in one or the other, so complete data is available for 55 students. Therefore, quantitative analyses were conducted on the 55 students who completed ONRs in both languages (using listwise deletion), whereas qualitative findings are discussed in terms of the overall sample (N = 65).

Table 1. Participants (N = 65).

<table>
<thead>
<tr>
<th>Grade</th>
<th>Wave 1 (n = 26)</th>
<th>Wave 2 (n = 39)</th>
<th>Total (N = 65)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>0</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>First</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>Second</td>
<td>15</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>12</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>24</td>
<td>38</td>
</tr>
<tr>
<td>ELL status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>17</td>
<td>32</td>
<td>49</td>
</tr>
<tr>
<td>Not qualified</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Missing data</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
Oral narrative retell assessment

Assessments for this study were conducted by the principal investigator (Spanish and English assessments) and a trained graduate research assistant (English assessments only). Two frog wordless picture books (Mayer 1969, 1974) and the Strong Narrative Retell Assessment Procedure (Strong 1998) were used. Because all children were heritage Spanish speakers, they were assessed in Spanish first; approximately a week later the procedure was repeated in English. Books were counterbalanced so that half the children (N = 33) heard Frog goes to dinner (Mayer 1974) in Spanish and Frog, where are you? (Mayer 1969) in English. The opposite was true for the other 32 students. Scripts for these two books are designed to be comparable in length, sentence complexity, and level of cohesion (Greenhalgh and Strong 2001) and they have been used with Spanish–English bilinguals in other studies (Simon-Cereijido and Gutiérrez-Clellen 2009; Bedore et al. 2010). Each assessment session lasted approximately 20 minutes, and only the target language was spoken throughout the session.

To begin the session, the assessor read the title of the book and told the child they would be asked to retell the story in their own words after listening and that they would not be able to look at the pictures while doing so. Children then listened to the story on headphones while looking at the pictures in the book. The assessor stayed in the room but did something unrelated to the task (like work on a computer) and did not listen to the story. This naïve listener condition is intended to lead to more detailed retellings because it implies a lack of shared knowledge (Berman and Slobin 1994; Strong 1998; Boyd and Nauclé 2001).

After listening to the story, each child was given the option to review the pictures in the book before giving it to the assessor to put away. About half the children took the option to do so. The audiorecorder was turned on, and children were prompted to begin the retell. The assessor remained silent throughout, intervening only after pauses of more than three seconds, at which point she gave a general prompt such as ‘tell me more’ (‘dime más’) or ‘anything else?’ (‘¿algo más?’) (Miller, Andriacchi, and Nockerts 2011). Once the child retold most of the story or paused for longer than five seconds, the researcher asked ‘is that all you remember?’ (‘¿es todo lo que recuerdas?’) to end the assessment (Justice et al. 2006; Justice, Bowles, and Gosse 2010).

Quantitative analysis

All narratives were audiorecorded and transcribed by either the principal investigator or trained and experienced university students who were highly proficient in the target language. The Systematic Analysis of Language Transcripts (SALT) computer program and its accompanying conventions were used for coding (Miller, Andriacchi, and Nockerts 2011; Miller 2012). In accordance with SALT conventions, all measures were calculated using only complete and intelligible utterances. Maze behaviors such as repetitions, reformulations, and code-switches were omitted and did not count in any of the measures (Lucero 2015). Code switches were considered maze behavior in order to facilitate the analysis of a child’s proficiency in the two languages separately.

Following SALT conventions, transcripts were segmented into C-units using Loban’s rules (1976); a C-unit includes a main clause and any subordinate clauses. Coding of C-units only differed in the case of coordinated clauses with omitted subjects in the second main clause, as recommended in the literature on ONRs with Spanish-speaking children (Gutiérrez-Clellen and Hofstetter 1994; Miller et al. 2006). Utterances that contained a succession of verbs without repeating the subject were segmented into separate C-units, even if they would be considered a single utterance using standard coding conventions. For example, the frog jumped and landed in the water would be considered a single utterance using standard C-unit coding. However, in the present study it was coded as two utterances: the frog jumped/and landed in the water. This modified coding accounts for the pronoun-drop nature of Spanish – the subject pronoun is not necessary because it is encoded in the verb that follows. Modified C-unit coding results in a greater overall number of utterances than standard coding, but protects against the over inflation of grammatical complexity. In order to maintain consistency and
comparability across languages, modified C-unit coding was used to segment all transcripts in both languages (Miller et al. 2006), as recommended in the literature (Gutiérrez-Clellen and Hofstetter 1994; Gutiérrez-Clellen et al. 2000; Heilmann et al. 2008).

Vocabulary breadth was measured by number of different words (NDW; Miller et al. 2006), a count of unique, uninflected root words. NDW is useful as a cross-linguistic measure because it allows for direct comparison across languages (Simon-Cereijido and Gutiérrez-Clellen 2009). NDW is considered a developmentally sensitive and robust indicator of a child’s vocabulary and has been used in research with bilingual children (Heilmann et al. 2010). It has also been found to be positively related to reading achievement in Spanish-speaking children (Miller et al. 2006; Rojas and Iglesias 2013). NDW is calculated automatically by the SALT program.

Grammatical complexity was measured by mean length of utterance at the word level (MLUw), widely considered a reliable measure of grammar for this population of learners (Castilla, Restrepo, and Perez-Leroux 2009; Rojas and Iglesias 2013). Grammatical analysis is important because syntactic knowledge plays a role in reading comprehension, especially with regard to children’s ability to synthesize information (Gutiérrez-Clellen 1998; Verhoeven 2011). MLUw is widely considered a general measure of syntactic complexity and has been used in many ONR studies with bilingual children (Fiestas and Peña 2004; Miller et al. 2006; Simon-Cereijido and Gutiérrez-Clellen 2009; Bedore et al. 2010; Iluz-Cohen and Walters 2012). Further, Gutiérrez-Clellen et al. (2000) suggested that MLUw is an appropriate and reliable measure of grammar for cross-linguistic analyses. MLUw is analyzed automatically by the SALT program.

**Qualitative analysis**

Macrostructure ability can be measured in various ways (Reese et al. 2010; Schachter and Craig 2013; Terry et al. 2013), many of which correlate with or predict reading proficiency. In this study, macrostructure was assessed using the Narrative Scoring Scheme (NSS; Heilmann et al. 2010), a tool to determine a child’s ability to produce a coherent, sequential, and detailed narrative. The NSS incorporates features of story grammar with other narrative elements. It consists of 7 criteria that are qualitatively scored on a scale of 1 (minimal/immature) to 5 (proficient): introduction, character development, mental states, referencing, conflict resolution, cohesion, and conclusion. There are a total of 35 possible points. Transcripts were coded for NSS by the principal investigator, a trained graduate student, or a SALT professional.

To achieve inter-rater reliability, approximately 20% of transcripts (n = 12 in Spanish, n = 11 in English) were randomly selected to be transcribed and coded by both the principal investigator and the student or SALT professional. In Spanish, word-to-word match transcription was 92%, point-to-point identification of modified C-units was 89%, and agreement on NSS was 87%. In English, word-to-word match transcription was 93%, point-to-point agreement on identification of modified C-units was 95%, and agreement on NSS was 90%. In all cases, inter-rater reliability was considered adequate.

**Findings**

Quantitative and qualitative data were analyzed separately but are combined here in order to present a robust, integrated description of the narrative retelling abilities of heritage Spanish-speaking emergent bilingual children in the early years of dual language education. In the sections that follow, findings for the two research questions are presented.

**Research question 1: relations across languages**

The first research question was: how were microstructural and macrostructural components of children’s narratives related across languages? Table 2 shows descriptives for the overall sample, as well
results of paired-sample t-tests for each component. Children performed significantly better in macrostructure in Spanish than English (SpNSS $M=19.69$, $SD=6.70$; EngNSS $M=18.80$, $SD=6.89$), $t(54)=1.033$, $p<.001$ (two-tailed), eta squared = .019. In contrast, children performed significantly better in English on one microstructure element: vocabulary (EngNDW $M=62.00$, $SD=22.70$; EngNDW $M=65.49$, $SD=29.58$), $t(54)=-.1057$, $p<.001$, eta squared = .02. There were no significant differences across languages for grammar (SpMLUw/EngMLUw).

Table 3 shows within and cross-language Pearson product-moment correlations. Vocabulary scores (NDW) were significantly correlated across languages ($r=.588$, $p<.01$), suggesting that children who used a wider range of words in the Spanish story also tended to use a large variety of words in the English story. Unlike vocabulary, grammar (MLUw) was not significantly correlated across languages ($r=.116$, $p=.399$). This was true despite the fact that means were similar (7.31 in Spanish, 7.92 in English). Finally, like vocabulary, macrostructure (NSS) performance was significantly correlated across languages ($r=.557$, $p<.01$).

**Research question 2: different patterns of bilingualism**

The second research question was: what patterns of bilingual performance did children exhibit on oral narrative retells? The key finding was that despite significant correlations in NDW and NSS performance across languages for the overall sample, individual children exhibited a range of proficiencies in each language. This would be expected in any dual language program and is not, in itself, concerning. Indeed, research has shown heterogeneity in narrative performance even within the ELL-designated population (Hipfner-Boucher et al. 2015). Given this diversity, individual children’s bilingual performance can provide insight into their strengths and challenges in oral language, which may forecast future reading proficiency (Cain 2003; Boudreau 2008). This information can be especially helpful when investigating the performance of children who struggle with the ONR task in one or both languages.

The strengths and challenges of individuals in this sample are visually evident in Figures 1–3, scatterplots which illustrate the relations between Spanish and English performance on each of the ONR components (macrostructure, vocabulary, grammar) for each child. These scatterplots illustrate the

<table>
<thead>
<tr>
<th>Table 2. Means and standard deviations on ONR components for children with complete data ($N = 55$).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>NSS</td>
</tr>
<tr>
<td>NDW</td>
</tr>
<tr>
<td>MLUw</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Note: NSS = Narrative Scoring Scheme; NDW = number of different words; MLUw = mean length of utterance-word.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Correlations among ONR components for children with complete data ($N = 55$).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>SpNSS</td>
</tr>
<tr>
<td>SpNDW</td>
</tr>
<tr>
<td>SpMLUw</td>
</tr>
<tr>
<td>EngNSS</td>
</tr>
<tr>
<td>EngNDW</td>
</tr>
<tr>
<td>EngMLUw</td>
</tr>
<tr>
<td>Note: Sp = Spanish; Eng = English; NSS = Narrative Scoring Scheme; NDW = number of different words; MLUw = mean length of utterance-word.</td>
</tr>
<tr>
<td>$^* p &lt; .05$ (two-tailed).</td>
</tr>
<tr>
<td>$^{**} p &lt; .01$ (two-tailed).</td>
</tr>
</tbody>
</table>
similarities between a child’s two retellings; dots at the top right corner represent children with high scores on the given component in both languages, while those in the lower left corner represent those with lower scores in both languages. The linear regression lines for NSS and NDW show the expected positive relations, while the linear regression line for MLUw indicates the non-significant cross-linguistic relation in the domain of grammar. In addition, the scatterplots show the diversity in how individual children’s retells were related across languages.

Of interest in the present analysis was whether children performed similarly in each language relative to their peers. Therefore, additional analyses were conducted on the retells of children who were considered to be struggling with narrative retelling in either Spanish, English, or with both languages. To identify these children, a benchmark was set at $-1.5$ SD on two or more ONR components within a language (Gutiérrez-Clellen 2002). Children who were unable to complete the assessment in either language were also considered to be struggling in that language. For example, a child who had SpNDW and SpNSS scores that were 1.5 SD below the sample mean was identified as struggling in Spanish. A child who performed 1.5 SD below the sample mean on EngNDW and EngMLUw would be identified as struggling in English. A child who scored $-1.5$ SDs on two or more components in both Spanish and English was considered to be struggling with the task in both languages. Using these criteria, three groups of students with different patterns of struggle were identified: (1) those whose challenges were limited to Spanish ($n = 4$); (2) those whose challenges were limited to English ($n = 5$); and (3) those who struggled in both languages ($n = 5$). Table 4 shows the Spanish and English scores on all components for children – identified by their initials and grade level – in each of the three groups. A dash indicates that the child was not able to complete the task in that language.

Figure 1. NSS scatterplot.
Figure 2. NDW scatterplot.

Figure 3. MLUw scatterplot.
As can be seen in Table 4, four students – three kindergartners and one first grader – met the criteria to be considered struggling in Spanish ONR performance. All four children could not complete the task in Spanish at all, but performed within the expected range (±1.5 SD) in English. This was somewhat unexpected given that all participants came from homes where Spanish was spoken and were enrolled in a DLI program. However, three of these children were in kindergarten, so perhaps their experience telling stories in Spanish was still somewhat limited. Only one first grader was unable to complete the ONR in Spanish, and no second graders were considered Spanish strugglers.

### Children struggling in Spanish

As can be seen in Table 4, four students – three kindergartners and one first grader – met the criteria to be considered struggling in Spanish ONR performance. All four children could not complete the task in Spanish at all, but performed within the expected range (±1.5 SD) in English. This was somewhat unexpected given that all participants came from homes where Spanish was spoken and were enrolled in a DLI program. However, three of these children were in kindergarten, so perhaps their experience telling stories in Spanish was still somewhat limited. Only one first grader was unable to complete the ONR in Spanish, and no second graders were considered Spanish strugglers.

### Children struggling in English

Five students – three kindergartners and two first graders – met the criteria to be considered struggling in English ONR performance (Table 4). Two of these children could not complete the task in English at all, but performed within the expected range in Spanish. Three others performed below −1.5 SD in two or more English ONR components. An illustrative example of a retell told by one of these children serves to highlight the discrepancy in one child’s two languages. E.HR. was a first grade girl who struggled to retell *Frog, where are you?* (Mayer 1969) in English.

C (They um) they were X the frog.
C And the frog have babies.
C (Um and um h*um he um) the frog was gone.
C And then (they) they found him.
C That’s it.

At the macrostructure level, E.HR. received an EngNSS score of 9, compared to a sample mean of 18.8 (35 possible). She launched into the story without introducing the setting or characters, and did not include any of the events critical to advancing the plot. She gave a brief, undetailed conclusion. At the microstructure level, she used only 13 different words (EngNDW sample $M = 65.49$) and had an EngMLUw of 4.0 (sample $M = 7.92$). The single area of strength in E.HR.’s English retell was her consistent and correct use of referencing, including the appropriate use of definite articles. However, if someone did not know the story already, they would not be able to reconstruct it strictly from the information given in this retell.

Table 4. Performance of struggling students by group.

<table>
<thead>
<tr>
<th></th>
<th>Spanish ONR</th>
<th></th>
<th>English ONR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SpNSS</td>
<td>SpNDW</td>
<td>SpMLUw</td>
<td>EngNSS</td>
</tr>
<tr>
<td>1. Struggling with Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.S. (K)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>9</td>
</tr>
<tr>
<td>E.Cr. (K)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>13</td>
</tr>
<tr>
<td>O.D. (K)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>20</td>
</tr>
<tr>
<td>L.KP. (1st)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>27</td>
</tr>
<tr>
<td>2. Struggling with English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.LC. (K)</td>
<td>21</td>
<td>69</td>
<td>7.1</td>
<td>–</td>
</tr>
<tr>
<td>Er.SC. (K)</td>
<td>15</td>
<td>41</td>
<td>7.4</td>
<td>9</td>
</tr>
<tr>
<td>L.VV. (K)</td>
<td>9</td>
<td>21</td>
<td>6.0</td>
<td>7</td>
</tr>
<tr>
<td>E.HR. (1st)</td>
<td>17</td>
<td>45</td>
<td>8.5</td>
<td>9</td>
</tr>
<tr>
<td>L.V. (1st)</td>
<td>13</td>
<td>27</td>
<td>7.8</td>
<td>–</td>
</tr>
<tr>
<td>3. Struggling in both languages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.LC. (K)</td>
<td>8</td>
<td>23</td>
<td>5.1</td>
<td>7</td>
</tr>
<tr>
<td>D.HP. (K)</td>
<td>9</td>
<td>10</td>
<td>4.3</td>
<td>–</td>
</tr>
<tr>
<td>El.M. (K)</td>
<td>7</td>
<td>13</td>
<td>4.5</td>
<td>–</td>
</tr>
<tr>
<td>R.SR. (K)</td>
<td>7</td>
<td>7</td>
<td>4.0</td>
<td>–</td>
</tr>
<tr>
<td>O.G. (1st)</td>
<td>12</td>
<td>12</td>
<td>3.7</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Sp = Spanish; Eng = English; NSS = Narrative Scoring Scheme; NDW = number of different words; MLUw = mean length of utterance-word.
In contrast, E.HR.’s Spanish retell of *Frog goes to dinner* (Mayer 1974) was more detailed and informative. Her NSS score was 17, which was much closer to the sample mean of 19.69. She used 45 different words (SpNDW sample $M = 62.00$) and her SpMLUw was an impressive 8.5 (sample $M = 7.31$). An English translation follows the Spanish transcript.

```
C (Cuando) su mama le dijo (que um) que se calle.
C Y después cuando fueron a la cena (y) la rana se subió en la ensalada.
C Eso es todo.
C Y algo (ah)>
C Y después cuando (le dijo al se* al que) le dio la ensalada a la (m*)>
C (Le ponía el la la la que) le dio la ensalada *y vio la rana.
C Y se brincó la rana en el vaso.
C Y después se ahogaron *y el señor lo agarró.
C Y el niño dijo que es su rana.
C Y su mamá que le dijo que se calle.
C Y (um) después (uh) les dijo el señor (que se) que tenga su rana.
C Y se ve con toda la familia estaba enojada.
C Y el papá le dijo que vaya a su cuarto.
C Y después al último (um) los animales se rieron.
C And later when they went to dinner (and) the frog jumped into the salad.
C That’s all.
C And something (um)>
C And then when (he said to the se* to that who) he gave the salad to the m*>.
C (He put the the the the that) he gave the salad *and saw the frog.
C And the frog jumped in the glass.
C And after that they drowned *and the man grabbed it.
C And the boy said it was his frog.
C And his mom told him (to um) to be quiet.
C And then at the end (um) the animals were laughing]
```

Similar to her English retell, E.HR. struggled to provide an introduction, and there were some points of confusion throughout the story. Nonetheless, her Spanish ONR was more detailed and informative overall. E.HR. included key events that led logically to the conclusion, and indicated a deeper understanding of the story by noting the mental states of characters – the family was angry, but the boy and his animals thought the situation was funny.

Taken together, E.HR.’s two retells show that she had an understanding of story structure and a Spanish vocabulary that was diverse enough to retell a coherent narrative. They also highlight the fact that she struggled with introductions but was proficient at referencing in both languages. Analyzing her bilingual retelling skills using quantitative and qualitative methods therefore provides deep insight into her overall oral language proficiency and potential future literacy development.

**Children struggling in both languages**

Five students – four kindergartners and one first grader – met the criteria to be considered struggling overall with the ONR task. Four children could not complete the task in English at all, but unlike children in the first group, they also performed below the expected range in Spanish. Only one child in this group, a kindergartner, was able to even complete the task in both languages, and in both he performed $-1.5$ SD below the sample mean on two or more components.

This group’s literacy profile is different from the other two groups in that the children’s struggles could not be attributed to low proficiency in one language or the other. Rather, they had universal struggles with oral language; not only were they unable to complete the task in English (their L2), but they did poorly in their L1 as well.
Discussion

The purpose of this article is to highlight patterns of bilingual proficiency on a composite oral language task among heritage Spanish-speaking emergent bilingual children enrolled in DLI programs. Little is currently known about how oral academic language skills are related across languages (Goodrich, Lonigan, and Farver 2013), and given the important role of oral language in later literacy outcomes, this is a significant omission in the field. The present study is one of the first to present detailed quantitative and qualitative analyses of the ONR performance of Spanish–English emergent bilingual children (for a notable exception, see Gutiérrez-Clellen 2002), although similar work has been conducted with other language groups (Schwartz and Shaul 2013; Westerveld 2014). To this end, illustrative examples of narrative transcripts have been presented alongside quantitative data to highlight prevalent patterns of bilingual narrative proficiency among children in the first three years of schooling. The value of this two-pronged analysis is to forefront the importance of instructing and assessing bilingual academic communicative competence for children learning two languages simultaneously. When assessing bilingual children, it is imperative to attend to oral language and literacy skills in both of their languages (Manis, Lindsey, and Bailey 2004; Miller et al. 2006; Gottardo and Mueller 2009). Oral narratives are an appropriate tool for this purpose because stories are commonly told and read aloud in primary classrooms, and children are expected to be proficient in retelling at an early age (Boyd and Nauclér 2001). Importantly, findings from this study also underscore the fact that high-level oral language skills in can develop side by side for emergent bilinguals enrolled in DLI education. Children seemed to be gaining bilingualism, rather than losing Spanish in the process of learning English.

In the present study, ONR performance was found to be significantly related across languages for the overall sample, especially in vocabulary and macrostructure. The finding that vocabulary was significantly related across languages contradicts some existing research that has reported non-significant relations (Pearson 2002; Bedore et al. 2010; Westerveld 2014). This is not surprising, given the different methods used to assess vocabulary and the different age groups involved. Children in the Westerveld (2014) study were four years old, and children in the Bedore et al. (2010) study were in kindergarten, so the findings from the present study begin to extend what is known about cross-linguistic ONR vocabulary proficiency among slightly older children with more school experience. In addition, the present finding that vocabulary is related across languages confirms other studies that have found such relations (Lindsey, Manis, and Bailey 2003; Proctor et al. 2006; Lucero 2015), although it is one the first to be shown using the ONR task.

The finding that grammar was not significantly correlated across languages is in line with the limited existing research on grammar – only one other ONR study investigated this relation, and the results were found to be non-significant (Bedore et al. 2010). Studies using other grammatical measures, such as oral cloze, have also found non-significance (Gottardo 2002; Gottardo and Mueller 2009), so there seems to be mounting evidence that grammatical proficiency is not related across languages. Reasons for this are currently unknown, and further research is warranted. Finally, the finding that macrostructure ability was significantly related across languages suggests that the ability to construct coherent stories is not bound by language and corroborates other research that has found cross-linguistic correlations at the macrostructure level (Pearson 2002; Uccelli and Paéz 2007; Iluz-Cohen and Walters 2012; Schwartz and Shaul 2013).

These findings underscore the need to consider bilingual skills holistically in order to understand strengths and challenges of emergent bilingual children and to create supportive environments for their academic success. There are a number of implications for instruction and assessment. Instructionally, research has shown that explicit attention to narrative retelling can improve reading comprehension (Garner and Bochna 2004; Boudreau 2008) and is therefore a key early literacy skill worth teaching. In the case of emergent bilingual children in DLI programs, such instruction could facilitate bilingualism and, ultimately, biliteracy. However, the form that instruction should take – and the language in which it should be provided – may be influenced by the pattern of bilingual proficiency
a child shows. Children who struggle primarily with Spanish oral language may need more support to continue developing L1 proficiency alongside English. A vast amount of research shows that developing a strong L1 will have long-term benefits for emergent bilinguals (August and Shanahan 2006; Goldenberg 2008; García and Kleifgen 2010). Without a strong foundation in Spanish, these children may be disadvantaged in the long run. In addition, they may not be identified for extra support because their English skills are within the expected range. This may have the effect of doubly disadvantaging such children. In the context of DLI education especially, they need exposure to complex Spanish language instruction to continue developing oral language.

In contrast, the group of children who struggle primarily in English is more likely to be identified as being at-risk for school struggle than those in the first group, given their low oral English proficiency. Common curriculum-based measurement (CBM) tools are typically administered in English, and decisions about how to provide supplemental instruction are frequently based on CBM results. However, the students in this group did well in retelling in Spanish, suggesting that they have a strong foundation in literacy and may simply need more time and support to continue improving their English proficiency. The fact that they could construct a detailed and coherent retell in their L1 suggests that they had good listening comprehension skills and understood the concept of story. For these children, teachers can draw upon L1 proficiency as a resource to build on while also facilitating English language development.

Finally, children in the third group – those who struggled with oral language in Spanish and English – may be the ones to be most concerned about in terms of language development overall. Unlike those in the first two groups in this sample, these children are the most likely to encounter future difficulties with literacy (Roth et al. 1996; Roth, Speece, and Cooper 2002) because of the difficulties they exhibited with both microstructure and macrostructure language skills in both languages. Emergent bilingual children who struggle with ONR in both languages should be the focus of intensive oral language and retelling instruction in both languages. Given the body of research suggesting that early narrative production skills can predict later literacy outcomes (Feagens and Appelbaum 1986; Griffin et al. 2004; Boudreau 2008), these children may benefit immensely from such instruction.

There are also assessment implications to be drawn from the present study. Findings suggest that it is important to monitor the oral language proficiency of emergent bilingual children in two languages side by side (Montanari 2004; Westerveld 2014). Without doing so, we might incorrectly assume that those who have challenges in one language are struggling with oral language overall, when in fact they may only be hindered by low language proficiency in that language. The nuanced analysis presented in this article highlights the fact that bilingualism is not a static set of skills, but rather a continuum of proficiencies in various domains (Gutiérrez-Clellen 2002). Measures like ONR can be used by teachers to learn more about the oral language proficiency of their emergent bilingual students.

This study had two notable limitations: first, the sample was small and all children attended the same school. Therefore, the study was exploratory and descriptive in nature. Further investigations are needed to determine whether this sample is representative of the larger population of US heritage Spanish-speaking emergent bilinguals. The diversity in the language proficiency profiles of participating children suggests that this may be the case, but at this point it remains to be seen. A second limitation was that all participating children were enrolled in a DLI program. While this is an important context for language and literacy development, many Spanish-speaking children in the USA are enrolled in English Only programs, and it is not possible to know from this study whether such children would exhibit similar patterns of bilingualism.

Even given these limitations, the exploratory quantitative and qualitative analyses presented in this article support the contention that administering composite academic oral language measures to emergent bilingual children has both practical and theoretical benefits (Manis, Lindsey, and Bailey 2004; Gottardo and Mueller 2009; Sandberg and Reschly 2011). Practically, it can help identify a child’s strengths and challenges in literacy-related language skills, which can guide instruction.
Theoretically, such assessment can illuminate patterns of bilingual proficiency for the increasing number of children who are receiving heritage language instruction in 50:50 DLI programs nationwide.

**Note**

1. In ONR transcripts presented here, children’s utterances have been segmented into C-units, but are otherwise presented exactly as the child told the story, including unconventional vocabulary and grammar. Maze behavior is indicated by parentheses, * indicates a missing word or part of word, and > indicates an incomplete utterance. Incomplete utterances were not considered in calculating MLUw. X indicates an unintelligible word.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**Notes on contributor**

**Audrey Lucero** is an assistant professor whose work broadly investigates the experiences of Spanish-speaking Latin@s with the U.S. public school system. Three specific strands of inquiry guide her current research agenda: 1) oral language and reading achievement among young Spanish-English emergent bilingual children; 2) dual language immersion education programs as venues for biliteracy development and community empowerment; and 3) the experiences of Latin@ undergraduate students at predominantly White state universities.

**References**


Barnes, Adrienne E., Young-Suk Kim, and Beth M. Phillips. 2014. “The Relations of Proper Character Introduction to Narrative Quality and Listening Comprehension for Young Children From High Poverty Schools.” *Reading and Writing* 27: 1189–1205.


