A Read-Aloud Storybook Selection System for Pre-Readers at the Preschool Language Level: A Pilot Study

Journal: Journal of Speech, Language, and Hearing Research

Manuscript ID: JSLHR-L-15-0056.R1

Manuscript Type: Research Article

Date Submitted by the Author: 27-May-2015

Complete List of Authors: Schwarz, Amy Louise; Texas State University, Communication Disorders van Kleeck, Anne; The University of Texas at Dallas, School of Behavioral and Brain Sciences Beaton, Derek; The University of Texas at Dallas, School of Behavioral and Brain Sciences Horne, Erin; The University of Texas at Dallas, School of Behavioral and Brain Sciences MacKenzie, Heather; Private practice, NA Abdi, Hervé; The University of Texas at Dallas, School of Behavioral and Brain Sciences

Keywords: Children, Early intervention, Literacy, Language, Education
A Read-Aloud Storybook Selection System for Pre-Readers at the Preschool Language Level: A Pilot Study

Amy Louise Schwarz, Ph.D.¹
Anne van Kleeck, Ph.D.²
Derek Beaton, M.S.²
Erin Horne, M.S.²
Heather MacKenzie, Ph.D.
Hervé Abdi, Ph.D.²

¹Texas State University
²The University of Texas at Dallas
Abstract

**Purpose:** Many well-accepted systems for determining difficulty level exist for books children read independently, but few are available for the wide range of difficulty of storybooks read aloud to preschoolers. Also, the available tools list book characteristics based only on parents’ or authors’ opinions. We created an empirically derived book difficulty level system based on 22 speech-language pathologists’ (SLPs) judgments of specific storybooks used in preschooler read-alouds.

**Method:** SLPs sorted 11 storybooks into ranked stacks based on how difficult they thought the storybooks would be for preschoolers to understand when read aloud to them. SLPs described each stack globally, as well as why they assigned each storybook into particular a stack. From transcriptions of the explanations, we derived a glossary of book characteristics using content analysis. We created a difficulty level scale using a multivariate analysis technique that simultaneously analyzed book sorts and glossary terms.

**Results:** The book selection system includes a glossary of book characteristics, a four-level difficulty scale, and exemplar books for each level.

**Conclusion:** This empirically derived book difficulty level system created for storybooks read aloud to preschoolers represents a step toward filling a gap in the read-aloud literature.

**Key words:** read-aloud, storybooks, difficulty level, glossary
A Read-Aloud Storybook Selection System for Pre-Readers at the Preschool Language Level: A Pilot Study

Reading aloud to children is a critical activity for fostering the knowledge they need for later reading success (Anderson, 1985). The literature is replete with research indicating how adults should tailor their talk and interactions when reading books aloud to preschoolers to help children develop the foundational oral language skills needed for later text comprehension (e.g., see van Kleeck, 2014, 2015 for discussion). By contrast, little attention has been paid to book complexity (van Kleeck, 2003) and to how books should be selected for interactive read-alouds. To foster preschoolers’ oral language skills—foundational for later text comprehension—professionals should select books that will increase vocabulary (e.g., McGee & Schickedanz, 2007; Whitehurst, Crone, Zevenbergen, Schultz, & Velting, 1999), and that will improve children’s abilities to predict and infer story events and character motivation (e.g., Dickinson & Smith, 1994; van Kleeck, 2008).

Professionals who work with preschoolers often struggle to select appropriate books, particularly when the professionals’ purpose is to increase oral language skills (e.g., see Beck & McKeown, 2001; McGee & Schickedanz, 2007). For example, in studying 243 read-alouds at 39 preschools in Sweden, Damber (2014) found that educators selected books randomly. Other researchers have found that books selected for read-alouds are often too simple to elicit predictions and inferences about the story (e.g., McGee & Schickedanz, 2007). Even when books are selected that lend themselves to predictions and inferences, Beck and McKeown (2001) observed that the illustrations often communicated too much information and thus hampered children’s efforts to grapple more independently with meaning.
For these reasons, professionals working with pre-readers at the preschool language level need a system for selecting storybooks for read-alouds when the purpose of the read-aloud is to increase oral language skills important for later text comprehension. Although well-accepted systems for determining difficulty level exist for books children read independently, few are available for the wide range of difficulty of storybooks read aloud to preschoolers. Several scholars have suggested factors to consider when selecting books for this purpose (e.g., Beck & McKeown, 2001; Elster, 1998; Griffin, 1970; Martinez & Roser, 1985; McGee & Schickedanz, 2007), but none of these suggestions have been operationalized. The systems that are available (Anderson, Anderson, Shapiro, & Lynch, 2001; Griffin, 1970; MetraMetrics’ online database) have several shortcomings impacting their utility.

To provide the rationale for our read-aloud book selection system, we explain how researchers and publishers determine the difficulty level of books, why a qualitative approach is most suited for a read-aloud book selection system, and the attributes of two systems that have been developed for read-alouds. We then describe how Chall et al. (1996) created their book selection system for matching literary texts to the reading levels of children who are reading books independently because we used it as a model when creating our system for pre-readers.

**Quantitative and Qualitative Text Complexity Systems**

When determining the difficulty level of written material, researchers and textbook publishers take an approach that is quantitative, qualitative, or a combination of both. The quantitative approaches traditionally reduce the level of text difficulty to a single metric based on one or more measures of sentence complexity and vocabulary difficulty (e.g., Fry, 2002; Gunning, 2003; Peterson, 1991). More recent readability systems (e.g., Source Rater, Reading Maturity, Co-Metrix) include additional measures, such as text coherence (see Nelson, Perfetti,
Liben, & Liben, 2012 for a discussion). These systems cannot capture qualitative information depicted in illustrations, which are a defining attribute of books used in read-aloud activities with children at a preschool language level (Corrigan & Surber, 2010).

Qualitative approaches include both objective and subjective measures of text complexity (e.g., Fry, 2002; Gunning, 2003; Peterson, 1991; Weaver, 2000). The objective measures include quantifiable metrics of vocabulary and sentence complexity. The subjective measures include adult opinions about the physical features of books, the illustrations, and the cognitive requirements of texts (e.g., Fry, 2002; Gunning, 2003; Peterson, 1991; Weaver, 2000). For example, MetaMetrics’ lexile scores rate the difficulty level of storybooks using both quantitative and qualitative measures. Each book is assigned a lexile score (scaled from below 200 to above 1700), which represents both a given text’s difficulty level and a person’s reading level (MetaMetrics, 2014b). The variables used to calculate the lexile scores are word frequency and sentence length (Nelson et al., 2012). MetaMetrics assigns an Adult-Directed (AD) code to indicate books that are suitable for read-alouds (MetaMetrics, 2014a). The company’s text measurement specialists make this determination based on the book’s layout and formatting along with its intended age range and audience, both of which are supplied by the publisher (personal communication, Bianco & Redman, October & November 2014). Because educators read books aloud to children from pre-kindergarten into upper elementary school years (e.g., Santoro et al., 2008; Tomlinson, 2009), it is unclear for which developmental level books with this designation are suitable.

Qualitative book selection systems are used to match books to the reading level of children who are learning to read and who will be reading the selected book independently. Key features of these systems include: (a) a glossary of terms that relate to a book’s difficulty level,
(b) a scale that educators can use to rank books based on difficulty level, and (c) exemplar passages or books that illustrate each level of the scale (e.g., Chall et al., 1996). Glossaries used in book selection systems generally have a two-tiered structure. The first tier contains the major book characteristics of the glossary, which typically include language complexity, physical features of the books, illustrations, and cognitive requirements of the text. The second tier contains item descriptions that further specify the major book characteristics. For example, Giffin’s (1970) book characteristic of illustrations includes the following items: (a) number of illustrations, (b) placement of illustrations (near the text they represent or clustered), (c) visual appeal, and (d) level of text support the illustrations provide.

The difficulty level scales that have been developed thus far are subjectively determined and—depending upon the particular book selection system used (e.g., Chall et al., 1996; Fountas & Pinnell, 2006)—include from 2 (easy or difficult) to 24 levels (containing many fine gradations of difficulty level). An exemplar passage or book is one that best represents a holistic application of the glossary (i.e., tier one: book characteristics and tier two: item descriptions) to a particular ranking on the difficulty level scale. These key features (i.e., a glossary, difficulty-level scale, and exemplar passages) were identified in previous research for book selection systems with the purpose of matching books to independent readers. We consider them essential components of a book selection system intended to help adults select books to read-aloud to pre-readers at the preschool language level when the purpose is increasing the oral language skills important to later text comprehension.

**Glossaries for Selecting Books to Read Aloud to Children**

Scholars have created glossaries based on parent interviews and expert judgment to select books for read-aloud activities with preschoolers (Anderson et al., 2001; Griffin, 1970). Because
Running Header: READ-ALOUD STORYBOOK SELECTION SYSTEM

we focus on children at the preschool language level, we excluded from our review another
glossary that was developed for read-aloud activities with first graders (Santoro et al., 2008).

The two systems of Anderson et al. (2001) and Griffin (1970) contain only a glossary.
These two systems lack scales of book difficulty and benchmark passages that Chall et al. (1996)
consider essential features for consistent application of a book selection system. Also, the
systems are meant for books across genres, including storybooks written as narratives (with and
without sound play), poetry, alphabet and counting books, and information books written as
expository text. Combining multiple genres into a single glossary can be problematic, because
different genres often include quite different book characteristics and hence might be expected to
differ in their glossaries (see Chall et al., 1996 for a discussion).

Another important consideration is the source of the glossary. Griffin’s (1970) glossary
appears to be based solely on the author’s opinion. Although one author’s opinion has some
value, a glossary based on the judgment of several persons offers a greater breadth of
perspective. This is the approach that Anderson et al. (2001) took when developing their
glossary. They conducted a study to determine how parents selected books to share with their 4-
year-old children, and whether those preferences varied based on the gender of the children.
Twelve mothers and twelve fathers were given 14 books across the various genres mentioned
earlier, with female and male characters equally represented. The parents were asked to choose
five books from the array that they would read to their 4-year-old child, as well as any books
they would not read to their child. They were then asked why they selected the books.

To discover trends in the data, Anderson et al. (2001) analyzed the data two ways. They	
tabulated which books the parents would and would not select. From these interviews, Anderson et al. (2001) extracted a set of key book characteristics and item descriptions the parents used.
Because the authors did not describe the manner in which they extracted the glossary, it is not possible to evaluate the validity of their glossary. Also, instead of randomizing the books for each participant, the researchers gave the parents the books in alphabetical order. This leaves open the possibility that the order of books may have influenced what book, language, illustration, and cognitive attributes the parents discussed.

The existing glossaries used to select books for read-aloud activities have the additional weaknesses of lacking difficulty-level scales and benchmark passages or books that holistically represent the book characteristics and item descriptions for each level of the scale. These omissions potentially make them difficult to apply (see Chall et al., 1996; Hiebert, 2002 for discussions). The only book selection system that contains all of the optimal design features and that achieved acceptable validity was developed by Chall and her colleagues (1996)—but it was designed for children who would be reading the books independently. We used their system, which we describe next, as the basis for our book selection system for read-aloud activities with prereaders at the preschool language level, however we also made some revisions to adapt and improve their design.

**The Qualitative Assessment of Text Difficulty (QATD).** Chall et al. (1996) created several book selection systems known collectively as the Qualitative Assessment of Text Difficulty (QATD) to help beginning teachers, parents and others match books to children’s independent reading levels. The QATD provides book selection systems in the three content areas of literature, science, and social studies. The QATD addresses text complexity within each content area by genre. For example, the literature content area contains the two genres of literary texts and popular fiction texts. For each genre within each content area, the QATD provides a book selection system that includes: (a) a glossary of book characteristics and item descriptions,
(b) a reading level difficulty scale, and (c) exemplar passages representing the holistic qualities of each reading level. We focus on how Chall et al. (1996) created the book selection system for matching literary texts to children’s independent reading levels because it is the most similar system in content to the one we created for adults to use when selecting storybooks to read aloud to pre-readers at a preschool language level.

To create the literary book selection system of the QATD, Chall et al. (1996) generated the following paraphrased questions to use when evaluating passages from books they considered to be on a literary continuum: (a) How common is the vocabulary and how many levels of meaning do the words include? (b) How complex are the sentence structures? (c) How much life experience is required for the reader to understand the passage? (d) How much cultural and literary knowledge is required to understand the passage? And (e) How much skill in literary analysis is required for the reader to understand the passage? The concepts in these questions (e.g., vocabulary, sentence complexity, relation to life experience) constitute the glossary of book characteristics for the QATD’s literary book selection system.

Chall et al. (1996) created the reading level difficulty scale for literary texts (which ranges from beginner level through college level) in two stages. First, they reviewed a multitude of passages from books across the literary continuum they identified, and then asked themselves the five questions above. They identified exemplar passages representing extreme ends of the scale and then identified exemplar passages representing the intervening reading levels. Recall that the book characteristics in their glossary are the concepts from the five questions they asked themselves when selecting the exemplar passages. Recall, too, that glossaries in book selection systems have a two-tier structure, including book characteristics (e.g., vocabulary knowledge) and item descriptions (e.g., non-literal word meanings). For each book characteristic, Chall et al.
(1996) generated item descriptions that explain how a given book characteristic changes at different reading levels of the scale. Although Chall et al. (1996) did not explain the process they followed when creating the item descriptions, we assume that the item descriptions explain how the book characteristics are instantiated within each exemplar passage.

The QATD’s literary book selection system is the best example of existing book selection systems because it contains a brief glossary of: (a) book characteristics and item descriptions according to book genre, (b) a scale that provides gradations of difficulty level, and (c) exemplar passages for each gradation of the difficulty level scale. Despite these strengths, the QATD’s book selection system for literary texts has two weaknesses. First, the selection of exemplar passages and the gradations of reading levels within the literary text scale reflect only Chall et al.’s (1996) opinion. We addressed this issue by basing our book selection system on the judgment of practicing speech-language pathologists (SLPs) who read storybooks aloud to preschoolers. Second, Chall et al. (1996) did not provide information about how they determined which books were appropriate to use as exemplar passages. We addressed this issue: (a) by using books from four preschool curricula and (b) by determining the exemplar storybooks after our statistical analysis determined the gradations of the difficulty level scale.

Research Questions

Our purpose was to create a system for selecting books for read-alouds based on how difficult they are to understand by pre-readers at the preschool language level. Following Chall et al. (1996), we narrowly focused our book selection system on one genre – storybooks without sound play and without prominent print salient features. Print salient features include labels within illustrations, sounds represented in illustrations, and words in a speech balloon to indicate which character is speaking (Zucker et al., 2009). Our specific research questions were:
(1) After sorting books that could potentially be read aloud to preschool-aged pre-readers, what terms do SLPs use when describing their reasons for: (a) creating each stack of books and (b) placing each book in a particular stack?

(2) What are the patterns of similarity and dissimilarity in the stacks created by the SLPs when sorting the storybooks?

(3) How do the patterns of similarity and dissimilarity of the book characteristics (research question #1) and the stacks of storybooks (research question #2) compare?

(4) As a group, what are the overall levels of difficulty SLPs identified and which storybooks best represent each level of the difficulty scale?

Method

Participants

Through email solicitations, we recruited SLPs who: (a) worked at the Callier Center for Communication Disorders, (b) received their clinical training at the University of Texas at Dallas, or (c) worked for major employers of SLPs in the Dallas Metroplex, including several school districts, clinics, and private practices. Participants were 22 experienced SLPs who used storybooks in therapy. Their experience ranged from two to 30 years ($M = 12.39$ years, $SD = 8.73$ years). Nineteen participants had completed a master’s degree (86.36%) and three participants had doctoral degrees (13.64%). Twenty-one participants were female (95.45%) and one participant was male (4.55%). Twenty participants worked full-time as SLPs (90.91%), and two participants worked part-time as SLPs (9.09%). Prior written consent to participate was obtained from all participants in accordance with the standards of the Institutional Review Board of The University of Texas at Dallas.

Storybook selection
When selecting the storybooks for this study, we searched the Internet for “storybook-based curriculum.” Curricula were selected that included a list of storybooks in the online description. We found 58 readily available books marketed as appropriate for preschoolers from four curricula: (a) *the Storybook-Based Curriculum* (Fetherston & England, 1998), (b) *Read, Play, and Learn!* (Linder, 1999), (c) *Read It Once Again* (Schaper, 2000-2009), and (d) *Happily Ever After* (Bannister, Preston, & Primozich, 2006). To control for the genre of books, we excluded 35 books because they were alphabet books, counting books, and/or storybooks including sound play (e.g., rhyming, alliteration). This control was necessary because adults read books differently to children based on book genre (e.g., Bradley & Jones, 2007; Zucker et al, 2009). The remaining 23 books were storybooks without sound play.

As a measure of whether these storybooks had wide appeal, we verified that all 23 storybooks were available through Amazon.com and the MetaMetrics’ database. MetaMetrics’ database was chosen because it is a metric that the National Governors Association Center for Best Practices and Council of Chief State School Officers (2010) recommends for leveling texts for children learning to read independently (e.g., Nelson et al., 2012) and it includes over 155,000 titles from over 200 publishers (MetaMetrics, 2014a).

For a sorting task to have the possibility of finding a stable pattern, the task needs to include between 9 and 20 items, with an optimal number of 12 items (Chollet, Lelièvre, Abdi, & Valentin, 2011). For this reason, we separated the 23 books into two groups, one containing 11 books and one containing 12 books, having plans to replicate our current study in a future study. To separate the books into two relatively equal groups, we used the number of different words (NDW) because it is a variable in quantitative readability systems, such as CohMetrix (Graesser, McNamara, Louwerse, & Cai, 2004), and because it is a measure of vocabulary complexity with
which SLPs are familiar. We used the Systematic Analysis of Language Transcription Software (Miller & Iglesias, 2012) to calculate the NDW for each book. We paired each set of two books with the closest NDW and then randomly assigned them to one of the two groups. We verified that the NDW distribution of each group of books was normally distributed using the Lillifors Test and had equal variance using Levene’s Test. A two-sample $t$-test confirmed that the two sets of books were not significantly different on NDWs $t(21) = -0.10, p = .922, r = .02, 95\% CI[-.39, .43]$. By flipping a coin, we decided that the set of 11 storybooks would be used in the current study, and the set of 12 books would be used in a follow-up replication study. See Appendix A for a list of the 11 storybooks used in this study and the curricula from which they were selected.

**Procedure**

A member of the research team met participants individually either at a university clinic or at the participant’s place of business in a conference room and audio-recorded the session. Storybooks were pre-randomized for each participant. The procedure took between 30 and 60 minutes ($M = 37$ minutes, $SD = 12$ minutes, $range = 19 - 60$ minutes) to complete, and included four steps. See Appendix B for the examiner’s script for each experimental step.

In the first step, each participant was asked to familiarize him or herself with each book. To ensure that the participants familiarized themselves with all the storybooks before attempting to sort them, we gave each participant a survey that asked the following about each book: (a) whether the participant was familiar with the book, and if so, (b) whether the participant liked the book. As the participants were completing these two survey questions, we gave them one book at a time to evaluate. After each participant completed the survey on all the books, the examiner reordered the books so that the books were in that participant’s unique randomized order.
In the second step, we administered the sorting task, which we adapted from the Santosa et al. (2010). We gave all 11 storybooks to each participant at one time and asked him or her to create ranked groups by sorting the books into stacks according to how difficult the books would be for a pre-reading preschooler to understand when hearing the books read aloud. The participants were not provided with any guidelines for sorting the storybooks.

In the third step, we asked each participant to describe the characteristics of each stack (i.e., stack-level descriptions). Participants were allowed to re-sort the books when they were describing the characteristics of each stack. However, when participants did re-sort one or more books, they were asked to describe again the characteristics of each stack of books in case their initial description had changed. We analyzed the participants’ descriptions of each stack to determine through content analysis the categories associated with each stack, or level of difficulty, they had created.

In the fourth step, we asked participants to describe why they placed each book in each stack (i.e., book-level descriptions). As before, participants were allowed to re-sort books when they were describing the characteristics of each book. However, when participants did re-sort one or more books, they were asked to describe again the characteristics of each stack (the third step in the procedure) as well as the characteristics of each book in all the stacks. This re-description step was necessary in order to capture possible changes in the glossary underlying the participants’ sorting decisions. We elicited descriptions at both the stack-level and individual book-level to provide the participants the opportunity to increase the precision of their descriptions.

**Fidelity.** Two graduate students administered the experimental script to 22 participants. To measure fidelity in script administration across examiners, three transcripts from each
graduate student’s administration of the script were randomly selected, equaling 27% of the data. A graduate student blind to the purpose of the experiment calculated fidelity by comparing the transcription of each examiner’s administration of the experimental script to the master script on a word for word basis. The percent correct ranged from 91% to 100%.

**Analysis plan.** We identified the glossary from transcriptions of the SLPs descriptions using content analysis (Weber, 1990). We used DiSTATIS—a factor-analytic technique particularly suited to sorting tasks (Abdi, Valentin, Chollet, & Chrea, 2007; Abdi, Williams, Valentin, Bennani-Dosse, 2012)—for three purposes: (a) to analyze patterns of similarities and dissimilarities in the stacks created by the SLPs when sorting the storybooks, (b) to compare these patterns with the glossary of terms we extracted from the SLP descriptions, and (c) to identify the exemplar books for each level of the difficulty scale. See Appendix C for a detailed description of how we conducted our content analysis and used DiSTATIS to create the book selection system.

**Results**

**Research Question #1:** After sorting books that could potentially be read aloud to preschool-aged pre-readers, what terms do SLPs use when describing their reasons for: (a) creating each stack of books and (b) placing each book in a particular stack?

**Results for Phase I**

**First round of data reduction.** Using 11 randomly selected transcripts (50% of the data), we generated four decision rules and 27 categories. The decision rules are bulleted below:

- When participants repeated ideas captured by a category when describing a particular stack or books within a stack, only the first mention of the category was coded. When
participants repeated ideas captured by a category across stacks and across books in
different stacks, those repetitions were coded.

- Participant utterances that did not pertain to the stack-level and book-level descriptions
  were not coded.
- If the participant contradicted him or herself, the original idea was not coded, but the
  final idea they expressed was coded.
- Utterances were coded as vague when they were uninterpretable (e.g., “it’s more visual
  than it is auditory”), ambiguous (e.g., “variation of text”), or referred to the overall
difficulty level rather than to a particular aspect of a stack and or storybook (e.g., “it’s
  simple”).

Two graduate students separately coded the remaining 11 participant transcripts (50% of
the total transcripts) for the 27 categories. After all remaining transcripts had been coded
separately, the coders met to compare their work and resolve any differences. Questions the two
coders could not easily resolve were discussed in a general laboratory meeting to reach
consensus.

Second round of data reduction. We reduced the 27 possible categories by selecting
only those categories mentioned by at least 50% of the SLPs during their stack-level and book-
level descriptions. This process identified the nine categories (i.e., book characteristics) shown in
Table 1.

[Table 1]

These nine categories totaled 79% of all the coded categories for the stack-level
descriptions and 60% of the total coded categories for the book-level descriptions. Note that the
vague category was one of these nine categories. Given that the vague category does not help us
create a glossary skilled SLPs use when classifying storybooks based on difficulty level, we excluded this category from further analysis. The remaining eight categories shown in Table 1 totaled 71% of all coded categories for the stack-level descriptions and 55% of the total coded categories for the book-level descriptions.

The same two graduate students who arrived at consensus coding for the 27 categories coded reliability in the second round. They coded 11 participant transcripts (50% of the total transcripts) for the eight key category descriptors. Reliability for each of the eight categories was calculated on four randomly selected participant transcripts using Cohen’s Kappa (Cohen, 1960), with Kappa values ranging from 0.65 to 0.96 (M = 0.81, SD = 0.12). Three categories did not reach Cohen’s preferred level of 0.80: infer (0.65), structure (0.70), and ill.sup (0.70); however the overall average Kappa was within the preferred range (0.81). Because of the three categories with somewhat low inter-judge reliability, the two coders resolved all differences, with more difficult questions being discussed in a general laboratory meeting in order to reach consensus.

Third pass. The glossary in Table 1 only includes key cognitive and perceptual characteristics the SLPs identified. To determine how many qualifiers the SLPs used when describing the categories for the individual storybooks, we again analyzed the book-level descriptions of all the participants. Note that some participants did not have additional comments on the individual books after describing each of their stacks. Four levels of difficulty—low, medium, high, and uninformative—were determined.

Two graduate students independently coded the participants’ book-level descriptions into the four qualifier categories. Cohen’s Kappa for the overall scale was 0.61; the majority of disagreement involved the middle of the scale, reflecting the difficulty of interpreting participants’ statements regarding medium levels of the eight descriptor categories. We used
information from the statistical analysis discussed in the next three sections to refine the intervening levels of the difficulty scale.

**Research Question #2: What are the patterns of similarity and dissimilarity in the stacks created by the SLPs when sorting the storybooks?**

The DiSTATIS analysis revealed two major patterns that explained 29.96% and 14.25% of the total variance, respectively (see Figure 1).

**Figure 1**

*Combined two-dimensional map with storybooks (dots) and 95% confidence intervals.*

[FIGURE 1]

In Figure 1, storybooks that are plotted closely together were frequently sorted together by the SLPs. Component 1—plotted as the horizontal axis—explains the largest amount of variance (29.96%) in both the stack and book data. Component 2—plotted as the vertical axis, explains the second largest amount of variance (14.25%) in both the stack and book data. The center of the map (where the two axes cross) corresponds to component scores of 0. Items near (or at) the origin contribute little to no variance in the data. Dots on the component maps represent storybooks.

**Confidence Interval Results.** When confidence intervals do not overlap, storybooks are considered significantly different from one another at \( p < .05 \). Many of the pairs of the books’ confidence intervals do not overlap (e.g., Friends and Abiyoyo; Kissing and Corduroy; Pizza and Rainbow) and this indicates that most books significantly differ from each other. In addition, the configuration of confidence intervals around each storybook reveals three clusters of storybooks: (1) lower left quadrant (Flower, Caterpillar, and Somebody), (2) upper left quadrant (Pizza and Snowy), and (3) the right side of the two-dimensional map (the remaining storybooks). Alone,
the two-dimensional map and confidence intervals—derived strictly from the participants’ sorting data, which was based on difficulty—suggests three levels of difficulty. However, to fully understand this configuration, we also analyzed the qualified book characteristics.

**Research Question #3: How do the patterns of similarity and dissimilarity between the SLPs’ reasons for sorting the storybooks (research question #1) and the actual stacks they created when sorting the storybooks (research question #2) compare?**

When the SLPs described why they sorted the books as they did, they often included “qualifiers” using vocabulary denoting easy, moderately easy, or difficult, and they sometimes gave uninformative responses such as “it’s simple” without specifying which aspects of the book were simple. From these descriptions, we extracted four qualifiers (low, middle, high, uninformative). We excluded the “uninformative” qualifier from the analysis because it did not help create the book selection system. In order to know why SLPs sorted these storybooks as they did, we overlaid (projected) onto the two-dimensional map the eight key book characteristics and three of the qualifier levels (low, middle, high) that the SLPs used in their book-level descriptions. Figure 2 displays the combined two-dimensional map of the storybooks with the qualified book characteristics. Note that in Figure 2 the book characteristic precedes the qualifier levels (e.g., infer.low = requires the child to make very few inferences to understand the story when it is read-aloud).

**Figure 2**

Combined two-dimensional map of storybooks (gray dots) and qualified book characteristics (colored in red, yellow, and green). Easy characteristics appear in green, difficult characteristics appear in red, and middle characteristics appear in yellow.
Analysis of the book characteristics showed that: (a) the lower left quadrant of the two-dimensional map is more associated with easy qualifiers (e.g., low inference, short in length, supportive illustrations, and simple vocabulary), (b) the lower right is more associated with hard/difficult qualifiers (e.g., high inference, storybooks with complex structures, and books that are—according to our participants—long), and (c) the upper right quadrant of the two-dimensional map is more associated with “medium” or “middle” qualities (e.g., density, length, and vocabulary).

The qualified book characteristics (Figure 2) help clarify some of the confidence interval results (Figure 1). The lower left quadrant in Figure 2 contains the easy books: Flower, Caterpillar, and Somebody. Book characteristics qualified as easier for preschoolers to process (lower left) are: simple sentences (lang.comp.simple), little information per page (density.low), simple vocabulary (vocab.simple), and short book length (length.short). The lower right quadrant contains the difficult books: Rainbow, Engine, and Abiyoyo. Book characteristics qualified as difficult (lower right) for preschoolers to process are: long books (length.long), lots of information per page (density.high), complex vocabulary (vocab.complex), and complex sentences (lang.comp.complex).

Given that the lower left and lower right quadrants reflect easy and difficult, respectively, the remaining books are qualified as “middle” or “medium” difficulty. The results from the confidence interval analysis (Figure 1) indicate significantly different sorting behavior between the storybooks in the upper right (Friends, Corduroy, and Kissing) and upper left quadrants (Snowy and Pizza) of the map. Figure 2 shows that all of the “middle” qualifiers (e.g., infer.med, lang.comp.med) mentioned by the SLPs are plotted exclusively in the upper right quadrant of the map, not the upper left quadrant. Therefore, the data indicate that the SLPs identified two
“middle” categories, one based on sorting (upper left quadrant) and another based on verbal descriptions (upper right quadrant). This means that there is a difference between how the SLPs sorted the books—which represents their judgment—and how the SLPs described their judgment about the books with respect to this “middle” set of books. Together, the results from the confidence interval analysis (Figure 1) and qualified book characteristics (Figure 2) suggest that there are two “middle” categories. We assigned the “moderately easy” label to the books in the upper left quadrant (Snowy and Pizza) because they are plotted closest to the books the SLPs identified as easy (Flower, Caterpillar, and Somebody). We assigned the “moderately difficult” label to books in the upper right quadrant (Friends, Corduroy, and Kissing) because they are plotted closest to the books the SLPs identified as difficult (Engine, Rainbow, and Abiyoyo).

**Component 1: Extreme gradations of the difficulty level scale.** Our results showed that Component 1 reflects the extreme gradation of the difficulty level scale, where the left is “easy” (e.g., Flower) and the right is “difficult” (e.g., Engine; as seen in Figures 1 and 2).

**Component 2: Middle gradations of the difficulty level scale.** Our results showed that Component 2 reflects something more complex than just easy versus difficult (Component 1; see Figure 2). Component 2 separates the easy and difficult books—in the lower left and right quadrants respectively—from the “middle” books in the upper left (Snowy, Pizza) and right quadrants (Friends, Corduroy, and Kissing).

When Components 1 and 2 are interpreted together, they show a difference between an “intuitive middle” (via sorting) set of books (Pizza and Snowy) and a “descriptive middle” (via characteristics) set of books (Friends, Corduroy, and Kissing). We interpret this “intuitive middle,” which we labeled “moderately easy” and “descriptive middle,” which we labeled “moderately difficult” as two middle gradations of the scale.
Research Question #4: As a group, what are the overall levels of difficulty SLPs identified and which storybooks best represent each level of the difficulty scale?

Recall that the Component 1 scores (horizontal axis) define the extreme gradations of the difficulty level scale and the Component 2 scores (vertical axis) differentiate two middle gradations of the scale. Taken together, these four gradations are: easy (lower left), moderately easy (upper left), moderately difficult (upper right), and difficult (lower right) as shown in Figure 3.

Figure 3

Combined two-dimensional map of the SLPs sorting data that reflects the four difficulty levels: easy, moderately easy, moderately difficult, and difficult.

Therefore, we adopted a four-point Likert scale for our difficulty level scale and applied the four-point Likert scale to each of the eight category descriptors shown in the table in Appendix E. In this way, we used the judgment of SLPs and how they select storybooks to read aloud to pre-readers at the preschool language level to create the difficulty level scale.

To select the exemplar storybooks for each gradation of the difficulty level scale, we applied the following rule. When there were two books in a gradation, as in the moderately easy gradation (Snowy, Pizza), either book could serve as the exemplar. For the remaining gradations, which included three books, we decided that the storybook with the intermediate component score best represented that gradation. So, the exemplar books are as follows: (a) easy gradation: Carle’s (1987) *Very Hungry Caterpillar*, (b) moderately easy gradation: either Keat’s (1976) *The Snowy Day* or Steig’s (1999) *Pete’s a Pizza*, (c) moderately difficult gradation: Freeman’s (2008) *Corduroy*, and (d) difficult gradation: Piper’s (2005) *The Little Engine that Could*.
Summary. We answered our first research question by making explicit the key categories (i.e., book characteristics) skilled SLPs intuitively use when selecting storybooks to read aloud to preschool-aged pre-readers based on the level of difficulty of those storybooks for preschoolers to understand. These key categories (shown in the top portion of Table 1) are based on both the stack-level and book-level descriptions provided by the SLPs. From the SLPs’ book-level descriptions, we also extracted the four qualifier categories of low, medium, high, and uninformative that can be applied to each of the eight key categories (i.e., book characteristics). We used DiSTATIS to identify patterns of similarity and dissimilarity in the stacks of books the SLPs created. We also used DiSTATIS to simultaneously analyze the SLPs’ ranked sorting data and the eight qualified book characteristics they used to explain their sorting decisions, which answered our third research question. This analysis confirmed that the SLPs sorted the books based on level of difficulty because it explains the greatest source of variance in the analysis. From this last analysis, we identified a four gradation difficulty scale and exemplar books for each gradation, which answered our fourth research question. We used the cluster of books in each quadrant of the average two-dimensional map of all the SLP sorting data as the basis for the difficulty level scale. We determined the exemplar books for each gradation either by selecting the book with the intermediate component score for gradations that contained three books or by selecting both books when the gradation contained only two books.

Discussion

Although scholars suggest book characteristics that professionals should consider when selecting books for read alouds to young children (e.g., Beck & McKeown, 2001; Griffin, 1970; McGee & Schickedanz, 2007), many professionals struggle to select appropriate books. Some select picture books for read alouds at random (Damber, 2014), whereas others select books that
lack challenging vocabulary or story structure (Beck & McKeown, 2001; McGee & Shickedanz, 2007). This lack of standard practice suggests that professionals need an evidence-based tool that operationalizes the book selection process for read alouds. In this study, we contributed to this need by introducing into a book selection system the criteria used by experienced SLPs when determining how difficult a storybook will be for a pre-readers at the preschool language level. This system includes a glossary, a difficulty level scale, and exemplar books for each gradation of the scale. In this section, we discuss how our system aligns with scholarly advice on book selection and how the levels of difficulty in our scale compare to the lexile scores assigned to these books by MetaMetrics. After discussing the study’s limitations, we discuss possible clinical implications and future directions.

**How Our System Aligns with Advice from Scholars on Book Selection**

The easy end of our difficulty level scale is defined by the book characteristics of supportive illustrations, simple story structure, vocabulary, and language structures as well as plots that draw on experiences familiar to preschoolers. These characteristics align with those mentioned by Richards (2010) as ones that increase children’s interest and enjoyment in books. The moderately difficult and difficult ends of our difficulty level scale are defined by the book characteristics of moderately high and high levels of inferencing; complex story structure, vocabulary, and language structure; as well as plots that do not draw on experiences familiar to preschoolers, and plots that do not reveal inferences and predictions required by the text. Many scholars have suggested that these characteristics are likely to help children who are acquiring the kind of oral language skills they will later need to understand more complex texts (Beck & McKeown, 2001; Dickinson & Smith, 1994; McGee & Schickedanz, 2007; van Kleeck, 2014, 2015; Whitehurst et al., 1999). The judgment of SLPs captured by our book selection system,
then, is congruent with the book characteristics scholars have suggested are important for educators to attend to when reading books aloud to children.

**MetaMetrics’ Lexile Scores and Adult-Directed Designation**

Recall that MetaMetrics—a privately-owned company that markets lexile scores as a way to match text difficulty to the reading level of independent readers—identifies picture books appropriate for read-alouds to children by giving them the *Adult Directed* (AD) designation. Books receive the AD designation based on the judgment of the company’s text measurement specialists (personal communication Bianco & Redman, October & November 2014). We cross-referenced the storybooks used in our study, which were selected from four readily available preschool curricula, with the MetaMetrics’ database. We wanted to know whether the books in our study received the AD designation and how their relative level of difficulty (as measured in lexile scores) compared to the level of difficulty assigned collectively by the SLPs in our study.

Only 55% of the books used in this study (6 out of the 11 books) received the AD designation from MetaMetrics. Given that the books used in this study come from existing pre-kindergarten curriculum, the MetaMetrics’ AD designation does not align well with the judgment of professionals designing preschool curriculum.

The lexile scores of the books used in our study ranged from 150L, which corresponds to a second-grade independent reading level, and 740L, which corresponds to a seventh-grade independent reading level (MetaMetrics, 2014b). These lexile scores are sensitive to the distinction between the easy and moderately easy levels of our scale and the easy and difficult extremes of our scale, with the exception of one book, Pfister’s (2010) *Rainbow Fish*. Based on word frequency and sentence length, MetaMetrics assigns this book a lexile score of 410L, which falls midway between the 150L and 740L range for the books in our study. The SLPs in
our study considered *Rainbow Fish* as one of the harder books of the 11 they sorted. They associated this book with complex vocabulary, complex language, complex story structure, and high amounts of inferencing, all of which are book characteristics lexile scores do not measure. Lexile scores also fail to capture the distinction between the moderately difficult and difficult gradations of our scale. This is not surprising considering that lexile scores are only based on word frequency and book length. As shown in Figure 3, the book characteristics distinguishing the moderately-difficult and difficult gradations of our scale are the characteristics crucial for interactive read alouds. Those characteristics are medium to high amounts of inferencing, complex vocabulary, complex language, and complex story structure. For these reasons, we suggest that practicing clinicians, educators, and researchers should not rely on MetaMetrics’ AD designation when selecting books to read aloud to pre-readers at the preschool language level.

**Study Limitations**

The two major limitations of this study are that the storybook selection system is unvalidated and that the system is based on the judgment of only 22 SLPs. We plan to validate the book selection system in two ways. First, we plan to have student clinicians use the difficulty level scale in Appendix E to rank the storybooks used in this study to see whether the scale helps them approximate the judgment of the master clinicians who sorted the storybooks in our study. If the scale helps student clinicians approximate the judgment of master clinicians, we will know that the clinical judgment captured in our analysis of experts is transferable to students in training as SLPs. Second, we plan to have the same master clinicians sort a different set of storybooks. Recall that, in the methods section, we mentioned having identified two sets of storybooks matched in their NDW distributions. We plan to have the same master clinicians who
participated in this study sort this new set of storybooks to determine whether the statistical model of their thinking is stable. It also will be important to repeat the experiment with a larger number of master clinicians from different geographical regions to determine whether the judgments of the 22 SLPs included in this study are truly representative.

Clinical Implications

Once we have replicated our storybook selection system, we will need to test whether it has clinical utility for SLPs. We imagine some SLPs may want to use this storybook selection system before these studies are conducted and published. Keeping in mind the preliminary nature of these findings, we have provided the following suggestions on how to apply the book selection system to storybooks not included in this study as, well as different ways the system might be used in clinical practice.

To use our storybook selection system, SLPs would need to familiarize themselves with the benchmark storybooks for each gradation of the difficulty scale shown in Appendix E, paying particular attention to how the books differ from each other in terms of difficulty level for each of the book characteristics. When selecting a storybook to read aloud that was not included in this study, SLPs will need to compare the new storybook to the exemplar storybooks for each of the book characteristics and to identify in which gradation of difficulty the new storybook belongs. If the clinician’s purpose in conducting a read aloud is for enjoyment and for encouraging children to draw on their background knowledge (e.g., Richards, 2010), that clinician should select a storybook that is most similar to the exemplar books associated with the easy (*Very Hungry Caterpillar*) and moderately easy (*Snowy Day, Pete’s a Pizza*) gradation of the scale. Based on the language level and background knowledge of the child or small group of
children matched on language level (Beck & McKeown, 2001), the SLP would decide which
gradation on the easy end of the scale to use.

However, if an SLP’s purpose is to read aloud a storybook to increase a child’s command
of complex vocabulary (e.g., McGee & Schickedanz) and/or to improve the child’s ability to
make predictions about the story structure and inferences about character motivation (e.g., Beck
& McKeown, 2001), the SLP should select a storybook that is similar to the benchmark books
associated with the moderately difficult (Corduroy) and difficult (Engine) gradations of the
scale. The SLP would decide which gradation to use on the hard end of the difficulty scale based
on the ability level of the child or ability-matched small group of children she or he serves.

When using this tool with small and large groups that include children with varying
language levels, we suggest selecting books based on the ability of children with the most
developed language skills in the group. This will insure that the practitioner consistently provides
the high-level of interactive talk children need during read-alouds to acquire vocabulary and to
make inferences and predictions essential for later text comprehension. The practitioner can still
include children with less developed language skills during a group read-aloud by directing a few
questions to them that draw on information shown in the illustrations and or directly stated in the
text. By taking this approach, the practitioner can support continued growth of children with
more developed language skills and uplift those with less developed language skills.

We also believe that this scale could be a training tool that SLPs can use when consulting
with early childhood educators and when supervising student clinicians who are evaluating
various aspects of storybooks to read aloud in treatment. For example, supervising clinicians and
student clinicians could use the scale in Appendix E to rate independently a stack of storybooks.
Then, they could have a discussion about similarities and dissimilarities in their thinking to better
align their judgments. This use of the scale could make the clinical judgment of the supervising clinician obvious to the student clinician, thus deepening their understanding and depth of training.

Future Directions

This study included only the genre of storybooks read aloud to pre-readers at the preschool language level. There also is a need for separate book selection systems to help clinicians select read-alouds for different book genres. Recall that Chall et al. (1996) created the Qualitative Assessment of Text Difficulty (QATD), which includes a collection of book selection systems by genre for professionals and parents to use when selecting books for children to read independently. Besides including book selection systems for fiction and popular fiction, the QATD also includes book selection systems for the genres of life science, physical science, narrative social studies, and expository social science. For preschoolers, the genres for creating additional book selection systems might include rhyming books, alphabet books, and a variety of types of expository books.

Given the increasing cultural and linguistic diversity among our clients at the preschool language level, there also is a great need for separate read-aloud book selection systems across genres for dual language (e.g., English-Spanish books) and minority language books (e.g., Spanish only books). Although dual language and minority language read-alouds are often recommended, clinicians who serve English Language Learners (ELLs) lack an evidence-based tool for selecting books based on difficulty level.

A final area for future research is to determine whether and how choosing books at different levels of difficulty affects the quality of talk surrounding the book and the overall success of the read-aloud activity. It will be important to measure not only the amount and
quality of talk generated by the practitioner and the children but also the level of support the practitioner provides and the children’s level of engagement.

**Conclusion**

Selecting the appropriate storybook to read aloud to pre-readers depends upon the SLP’s purpose in reading aloud the storybook and the SLP’s ability to both identify and rate the difficulty level of key characteristics that are manifested in the storybook. The results of this study isolated and scaled these key characteristics, which are the amount of inferencing required, the density of ideas per page, the child’s level of experience with concepts included in the storybook, the degree of conceptual support provided by the illustrations, the complexity of vocabulary, language, and story structure as well as book length. The Likert scale in Appendix E provides SLPs with a way to systematically approach book selection. This evidence-based tool has the potential of being easy to incorporate into classroom and therapy contexts. Future studies will determine its validity and clinical utility.
References


Appendix A

| TABLE A1 |
Appendix B

[TABLE B1]
Appendix C

Research Question #1: After sorting books that could potentially be read aloud to preschool-aged pre-readers, what terms do SLPs use when describing their reasons for: (a) creating each stack of books and (b) placing each book in a particular stack?

To identify the glossary from this corpus, we conducted a content analysis. Content analysis is a qualitative research method that reduces preserved forms of human communication, such as written texts and transcribed audio-recordings of interactions, into key content categories that allow the researcher to make valid inferences about the essential message in the text (Weber, 1990). In our analysis, we included all of the transcribed data from each participant (i.e., an all-inclusive content analysis) because sampling only part of each transcript might have threatened the semantic cohesiveness of the corpus (Weber, 1990), and might have caused us to miss one or more of the key book characteristics important to the glossary we hoped to identify. When conducting a content analysis, data can be reduced by having the research team hand-code the transcripts or by having a computer program code the transcripts. Following Weber (1990), we chose to hand-code the transcripts because hand-coding allowed us to incorporate both syntactic and semantic aspects of the participants’ stack-level and book-level descriptions into our reduced data set.

Our unit of analysis was the category. As discussed in Weber (1990), researchers who use content analysis define category differently. In this study, we define category as a mutually exclusive collection of words, phrases, sentences, and series of sentences with different meanings, that when taken together, describe categories SLPs consider important when assessing the difficulty level of storybooks. Recall that glossaries contain the two tiers of book characteristics (tier one) and item descriptions (tier two). The categories we extracted from the
data were the book characteristics of our glossary. The item descriptions were words, phrases, sentences and series of sentences that provided different dimensions of meaning for the book characteristics (e.g., easy to difficult vocabulary, simple to complex sentence structure). To identify the glossary and item descriptions, we made three passes of the transcripts.

**First pass.** We reduced each participant’s stack-level and book-level descriptions twice in order to arrive at the key categories (i.e., book characteristics) SLPs use to assess the difficulty level of storybooks they imagined reading aloud to preschool-aged pre-readers. In the first round of data reduction, we generated a hand-coded list of categories and our decision rules for placing information from the transcripts into those categories using 11 randomly selected transcripts. Then two graduate students coded whether each participant mentioned (present = 1, absent = 0) each category during his or her stack-level and book-level descriptions.

**Second pass.** In the second round of data reduction, we narrowed the list of categories to those most often mentioned by the participants. We identified only those categories mentioned by greater than or equal to 50% of the participants when giving their stack-level and book-level descriptions. The categories identified in the second round of data reduction created the glossary for our book selection system.

**Third pass.** The glossary identified cognitive and perceptual categories without specifying how the participants qualified their use of the categories when providing their descriptions for each storybook they placed in a particular stack. For example, the glossary term “inference” was applied in both of the following cases. A participant might say Eric Carle’s *Very Hungry Caterpillar* (1987) requires a pre-reader to make very few inferences when listening to an adult reading the storybook aloud, while Don Freedman’s *Corduroy* (2008) requires the pre-reader to make many inferences to understand the story. We analyzed the book-level descriptions
a final time to identify any qualifiers (e.g., low, easy) the SLPs used when describing why they put each book in each stack. So in the above example, we keyed in on the qualifying terms “very few” and “lots” for the dimension of “inference.” From these data, we extracted a preliminary scale of difficulty.

**Research Question #2: What are the patterns of similarity and dissimilarity in the stacks created by the SLPs when sorting the storybooks?**

To answer this question, we used DiSTATIS—a factor-analytic technique particularly suited to sorting tasks (Abdi et al. 2007, 2012)—to analyze the similarities and dissimilarities of the stacks created by the SLPs when sorting the storybooks. See Appendix D for a technical discussion of the software package we used to conduct the analysis, and for examples of how the data were coded in our analysis.

**DiSTATIS.** DiSTATIS is part of a family of methods that analyzes multiple data tables via principal components analysis (PCA; Abdi & Williams, 2010, Abdi et al., 2012). DiSTATIS analyzes a set of dissimilarity (or distance) matrices measured on the same items (Abdi, 2007).

To answer our second research question, we used the stacks each participant created during the storybook sorting task. Each participant’s sorting data generated a unique table to represent how each sorted the storybooks, giving us 22 data tables (see Appendix D for a discussion of how the data were coded, transformed, and normalized before the DiSTATIS analysis was conducted). DiSTATIS combined these 22 tables into a single matrix that described the overall relationship of each storybook to one another. DiSTATIS created new orthogonal variables (à la PCA), called components, that best represent the pattern of similarity between storybooks, ranked by their variance. The first component explained the maximum variance in the data, and each subsequent
component explained the next largest amount of variance (under the constraint that components are pairwise orthogonal, i.e., uncorrelated).

DiSTATIS assigned component scores to the storybooks to reflect how much each book contributed to the variance of each component. DiSTATIS plotted the component scores onto two-dimensional maps that showed the spatial relationship between the storybooks across two components. On these maps, books that appeared close to each other were often sorted together. In contrast, books that appeared far apart were rarely sorted in the same grouping by the SLPs. The two-dimensional maps describe the overall relationship among storybooks: Items at the extremes of a component contribute a high amount of variance to that component and items close to the origin (i.e., 0), contribute little variance.

Confidence Intervals. We wanted to determine how stable the sorting pattern was across participants. To do this, we calculated confidence intervals using a bootstrap procedure. The bootstrap procedure resampled, with reselection (Efron & Tibshirani, 1993; Hesterberg, 2011), each SLP’s sorting data. Each SLP’s data could have been selected 0, 1, or many times. Following this resampling, the data table was recomputed (Abdi, Williams, Valentin, et al., 2012, Abdi et al., 2009). This procedure was repeated 1000 times to create distributions around each storybook that represented how much a storybook’s component scores changed under resampling. These distributions were used to build 95% confidence intervals (i.e., peeled convex hulls, Greenacre, 2007), which DiSTATIS plotted onto a two-dimensional map. When the confidence intervals did not overlap, books were considered to be significantly different from one another at the $p = .05$ level (Abdi et al., 2009).
Research Question #3: How do the patterns of similarity and dissimilarity between the SLPs’ reasons for sorting the storybooks (research question #1) and the actual stacks they created when sorting the storybooks (research question #2) compare?

To answer our third research question, we used DiSTATIS to project the key book characteristics extracted from the SLP interviews onto the two-dimensional map of storybook stacks created by the SLPs (see Appendix D for a discussion of how this was done and for an example of how the data were coded). We used a method called *barycentric projection* to project the book characteristics onto the DiSTATIS map. This process created *supplemental component scores* for the characteristics. The supplemental component scores of the key book characteristics did not contribute to the variance or the component structure. Rather, they indicate which qualified book characteristics occurred very frequently with a particular storybook.

Research Question #4: As a group, what are the overall levels of difficulty SLPs identified and which storybooks best represent each level of the difficulty scale?

We derived the overall difficulty level scale and the exemplar storybooks from the combined two-dimensional map that plotted the storybook stack data and the book characteristics extracted from the SLP interview data. We planned to base the difficulty-level scale on major sources of variance identified by the analysis and particular groups of storybooks. We planned to identify exemplar books based on which component score best represented each group of storybooks plotted by DiSTATIS on the two-dimensional map.
Appendix D

To conduct our DiSTATIS analysis, we used the MExPosition package (e.g., Chin Fatt, Beaton, & Abdi, 2012; Beaton, Chin Fatt & Abdi, 2014), with some additional in-house scripts, for the [R] statistical software (R Core Team, 2012). DiSTATIS is covered in comprehensive detail elsewhere (Abdi, Dunlop, & Williams, 2009; Abdi & Valentin, 2007; Abdi, et al., 2012; Lelièvre et al., 2009).

DiSTATIS’s Relationship to Other Component-Analytic Techniques

DiSTATIS is a special case of STATIS—a family of techniques with numerous special cases. STATIS performs an optimal PCA on multiple data tables. The differences between the special cases are specific to optimal weight selection and normalization methods (Abdi, Williams, Valentin, et al., 2012). In the case of DiSTATIS (as opposed to STATIS), each table is a distance table. DiSTATIS analyzes sets of distance matrices, where each participant provides a distance matrix (as described in the next section). Instead of an optimal PCA for multiple data tables, DiSTATIS is an optimal MDS for multiple distance tables (and therefore DiSTATIS is part of the 3-way factorization techniques such as Tucker-3, Tucker, 1966).

Coding Data for DiSTATIS

Active variables. For our analyses, stacks were labeled as described on a scale of perceived difficulty (i.e., 1 for easiest and all other group labels ascending from there). Each participant’s sorting data generated a unique table (hence multiple tables) to represent how each sorted the storybooks. Table D1 provides an example of how the data were initially coded. The first column contains abbreviations of the storybook names. The remaining columns include the grouping of the stacks (labeled as the stack rank order) each participant created during the task.

[TABLE D1]
To perform DiSTATIS, each participant’s data needed to be represented as a dissimilarity (distance) matrix. A distance, or dissimilarity, matrix is a symmetric table (i.e., same number of rows and columns with the same labels) where the diagonal is 0. A value of 0 means “0 dissimilarity (distance)”, and therefore two items with a distance of 0 are the same. The larger the value, the more dissimilar (or further apart) two items are.

Data in Table D1 were transformed into a set of indicator matrices, as shown in Table D2.

**[TABLE D2]**

Table D2 was then transformed into a set of distance matrices, as shown in Table D3.

**[TABLE D3]**

This was done with matrix algebraic manipulation (as seen in Eq. in Abdi & Valentin, 2007) or with a binary distance metric. For further detail on the transformation and analyses, see Abdi, Williams, Valentin, et al. (2012) and Abdi and Valentin (2007). The aggregate set of distance matrices is what is analyzed by DiSTATIS.

**Supplemental variables (qualified book characteristics).** Recall that once each participant had sorted the storybooks, the examiner asked the participant to describe their reasons for sorting the books into particular stacks. For this data, we extracted categories of book characteristics and a preliminary low, medium, high, uninformative scale. These characteristics were then dummy-coded—just like the sorting data (see below)—and projected as supplemental variables onto the DiSTATIS map (see Abdi & Williams, 2010; and Abdi, Williams, Valentin et al., 2012 for descriptions of supplemental projections and DiSTATIS). Because the supplemental variables were defined *per book*, these data were projected with respect to the partial projections (i.e., the individual book scores per participant), as shown in Table D4.
[TABLE D4]

For this analysis, DISTATIS computed *partial component scores*, through partial projection techniques (e.g., Abdi, Williams, Valentin, et al., 2012).
Appendix E

[TABLE E1]
Table 1

*Categories that greater than or equal to 50% of the participants mentioned when giving their stack-level and book-level descriptions in the second round of data reduction*

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Categories</th>
<th>Item descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infer</td>
<td>Inference</td>
<td>Amount of inference, problem solving, and abstract concepts</td>
</tr>
<tr>
<td>Length</td>
<td>Length</td>
<td>Length of the book</td>
</tr>
<tr>
<td>Vocab</td>
<td>Vocabulary</td>
<td>Complexity of vocabulary and amount of different vocabulary</td>
</tr>
<tr>
<td>density</td>
<td>Density</td>
<td>Compactness of text and/or information per page, including comments about font size</td>
</tr>
<tr>
<td>lang.comp</td>
<td>Language complexity</td>
<td>Complexity of language and sentence length</td>
</tr>
<tr>
<td>Exp</td>
<td>Experience</td>
<td>Familiarity of activities and experiences for preschoolers</td>
</tr>
<tr>
<td>ill.sup</td>
<td>Illustration support</td>
<td>Level of support illustrations provide the story</td>
</tr>
<tr>
<td>structure</td>
<td>Story structure</td>
<td>Complexity of story structure including one or more of the following: (a) number of characters in the story, (b) clarity of story sequence, (c) predictability, and (d) memory load.</td>
</tr>
<tr>
<td>Vague</td>
<td></td>
<td>Un-interpretable, ambiguous, or referring to overall difficulty level rather than a particular aspect of a stack and/or storybook</td>
</tr>
</tbody>
</table>

Category excluded from our glossary and our statistical analysis.
### Table A1

**Abbreviations, references, and curricula for the 11 storybooks.**

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Storybook references</th>
<th>Curricula</th>
</tr>
</thead>
</table>
Table B1

Script used across four steps of the procedure

First step: Familiarity and preference survey

(Gesture toward the stack of books.) *Take your time with each of the books. As you look at each book, I want you to fill out this form* (Give them the Familiarity & Preference Checklist*. You’ll notice that each book on this form has two questions (indicate on the paper). *Here’s the first title with two questions and the second title with the same two questions.* (Collect the checklist when they are finished). *Thank you.*

Second step: Sorting task

*Please sort all these books into stacks according to how difficult they are for pre-reading children to understand when you read the books aloud to them. There are 11 books. Sort them into as many stacks according to difficulty level as you want. You may place each book in only one stack (gesture). After you finish sorting the books, I am going to ask you to describe the characteristics that define the difficulty level of each stack. Then I will ask you why you placed each book in each stack.* (Give them 20 pages of paper and a pencil.). *Here is a pencil and some paper. Feel free to make notes about the groups you are creating. Once again, you are sorting the books into stacks according to how difficult they are for pre-reading children to understand when you read the books aloud to them. Please let me know when you are finished.*

Third step: Descriptive task for each stack of books

*Now I’m going to ask you to describe each stack you just created.* (Point to the stack of books on your far left). *Using adjectives, tell me why you grouped these books together.* (Move to the next stack and ask the same two questions.). *Let’s move to the next stack.*

Fourth step: Descriptive task for each book within each stack

*Now, I’m going to ask you why you placed each book in each stack.* (Start with the stack on participant’s far left. Pick up the first book in the stack. Read the title of the book.). *Using adjectives, tell me why you placed this book in this stack.* (Place the book in front of the stack. Pick up the next book in the stack. Read the title of the book and ask the same question. Continue this process until you’ve gone through all the books in the first stack. Move to the next stack). *Let’s move to the next stack.* (Repeat the same process until you’ve elicited a description about all the books in all the stacks.).

*Note.* Italicized text indicates what the examiner said to the participants. The regular text in parentheses indicates instructions to the examiner. *a = This checklist asked whether participants were familiar with each book and whether they liked each book.*
**Table D1**

*Example of data table containing each SLP’s ranked stack assignments for each storybook.*

<table>
<thead>
<tr>
<th>Storybooks</th>
<th>SLP1</th>
<th>SLP2</th>
<th>SLP3</th>
<th>SLP20</th>
<th>SLP21</th>
<th>SLP22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Flower</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Somebody</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pizza</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Rainbow</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Kissing</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>...</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Friends</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>...</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Corduroy</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>...</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Engine</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Abiyoyo</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

*Note. SLP = speech-language pathologist. See Appendix A for the full titles of each storybook that we shortened to a single word for our analyses.*
Table D2

*Example of a dummy-coded matrix.*

<table>
<thead>
<tr>
<th>Book</th>
<th>S01.1</th>
<th>S01.2</th>
<th>S01.3</th>
<th>S01.4</th>
<th>S02.1</th>
<th>S02.2</th>
<th>S03.1</th>
<th>S03.2</th>
<th>S03.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowy</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Flower</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Somebody</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pizza</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rainbow</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Kissing</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Friends</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Corduroy</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Engine</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Abioyo</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note.* Dummy coding means coding using 1’s and 0’s to represent presence and absence. Participants are labeled S0X. The ‘.’ reflects which stack a book belongs to for that person.
Table D3

*Example of a dissimilarity matrix for one participant (S01 from above).*

<table>
<thead>
<tr>
<th></th>
<th>Snowy</th>
<th>Flower</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>Rainbow</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>Engine</th>
<th>Abiyoyo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowy</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Flower</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Somebody</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pizza</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rainbow</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Kissing</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Friends</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Corduroy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Engine</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Abiyoyo</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table D4

Example of data table containing each SLP’s dummy-coded book characteristics. A ‘1’ indicates an SLP believes that characteristic belongs with that book.

<table>
<thead>
<tr>
<th>Storybooks</th>
<th>SLP1</th>
<th>SLP2</th>
<th>SLP3</th>
<th>SLP20</th>
<th>SLP21</th>
<th>SLP22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowy.Infer.Low</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Snowy.Infer.Med</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Snowy.Infer.High</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Snowy.Infer.Uninformative</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Abiyoyo.Vocab.Simple</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Abiyoyo.Vocab.Med</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Abiyoyo.Vocab.Complex</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Abiyoyo.Vocab.Uninformative</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exemplar Book</td>
<td>Exemplar Book</td>
<td>Exemplar Book</td>
<td>Exemplar Book</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Amount of inference, problem solving, and abstract concepts
| Low | Moderately Low | Moderately High | High |
| Complexity of story structure, including the number of characters, clarity of story sequence, predictability, and memory load
| Simple | Moderately Simple | Moderately Complex | Complex |
| Complexity of language and/or sentence length
| Simple | Moderately Simple | Moderately Complex | Complex |
| Complexity and/or amount of different vocabulary
| Simple | Moderately Simple | Moderately Complex | Complex |
| Density (or amount) of text/information per page, including font size
| Low | Moderately Low | Moderately High | High |
| Familiarity of activities/experiences for preschoolers
| Familiar | Moderately Familiar | Moderately Unfamiliar | Unfamiliar |
| Length of the book
| Short | Moderately Short | Moderately Long | Long |
| Levels of support illustrations provide for the story
| Supportive | Moderately Supportive | Moderately Unsupportive | Unsupportive |