Article

Hammill Institute on Disabilities

Assessment for Effective Intervention
2015, Vol. 40(2) 81–95

DOI: 10.1177/1534508414551404
aei.sagepub.com
SAGE

# Development of Early Measures of Comprehension: Innovation in Individual Growth and Development Indicators

**Alisha K. Wackerle-Hollman, PhD[1], Megan I. Rodriguez, MA[1], Tracy A. Bradfield, PhD[1], Michael C. Rodriguez, PhD[1], and Scott R. McConnell, PhD[1]**

**Abstract**

Early comprehension is an important, but not well-understood, contribution to early literacy and language development. Specifically, research regarding the nature of skills representative of early comprehension, including how they contribute to later reading success, is needed to support best practices to adequately prepare students. This article describes the process involved in the creation and refinement of the newly developed comprehension Individual Growth and Development Indicators (IGDIs 2.0). Two theoretical models of early comprehension are discussed to highlight the inherent complexity of this domain. Results of three studies are presented: Study 1 outlines the initial piloting process, Study 2 represents a larger-scale investigation, and Study 3 describes further field testing and reveals the final IGDI 2.0 comprehension candidate: Which One Doesn't Belong (WODB). Results indicated WODB out-performed the other candidate measures across psychometric and pragmatic criteria. The utility of the WODB task within a Response to Intervention (RTI) framework is also discussed.

**Keywords**

early literacy, test construction

Discussions reflecting the transition from "learning to read" during the preschool years to a long-term goal of "reading to learn" are becoming increasingly prevalent in early childhood (Dooley, 2011; Kail, 2011; Robinson, Einav, & Fox, 2013; Storch & Whitehurst, 2002). It is during this transition that children move from a focus on decoding text to a focus on deeper understanding, or comprehension, of the text and images printed on the page. Literacy experts agree that the end result of acquiring skills to read is comprehension; however, there is much contention in the field about how and when the skills required to add meaning to words begin to develop and how the continuum of such comprehension skills is manifested in the early years (van den Broek et al., 2005).

At the same time, a renewed focus on conceptually and empirically validating models of comprehension from early childhood through high school completion has come to the forefront of research efforts with grant competitions such as the *Reading for Understanding* initiative hosted by the Institute of Education Sciences, and numerous projects recently funded to explore comprehension development and application such as *Making the Right Connections: Improving the Comprehension of Struggling Readers* (McMaster & van den Broek, 2011) and *The Language Bases of Reading Comprehension* (Justice, 2010). Yet few studies have demonstrated models that empirically and conceptually describe the nature of early comprehension. These initiatives illustrate the juvenile state of the field in discerning a robust continuum of pre-requisite and precursor skills representative of comprehension.

Simultaneous to the K–12 efforts to explore comprehension (e.g., Justice, 2010), a growing emphasis on the need to discern what comprehension is for preschool-age children has also been fostered as initiatives in the early childhood realm consider and explore the mechanisms contributing to achievement gaps in later elementary years (e.g., Tompkins, Guo, & Justice, 2013). Such efforts have led to re-examination of current conceptualizations of how comprehension skills are acquired during early childhood.

Investigations of early comprehension have been grounded in extrapolations from what we currently know and

[1]University of Minnesota, Minneapolis, USA

**Corresponding Author:**
Alisha K. Wackerle-Hollman, University of Minnesota Urban Research, Outreach and Engagement Center (UROC), 2001 Plymouth Avenue North, Minneapolis, MN 55411, USA.
Email: wacke020@umn.edu

understand about comprehension in the early elementary school years. Comprehension skills characteristic of later elementary reading abilities are supported by research demonstrating that there are necessary precursors and in some cases pre-requisite skills representative of learning to read (e.g., early literacy skills: alphabet knowledge, vocabulary, and phonological awareness; National Early Literacy Panel [NELP], 2008; Snow, Burns, & Griffin, 1998; Storch & Whitehurst, 2002; Whitehurst & Lonigan, 1998). Specifically, evidence suggests early vocabulary development predicts later comprehension (Dickinson & Smith, 1994; Sénéchal, Ouellette, & Rodney, 2006), and early phonological awareness and alphabet knowledge skills predict later decoding skills in the first and second grades (e.g., Bradley & Bryant, 1983), with further studies demonstrating the connection between decoding and comprehension (Aarnoutse, Van Leeuwe, Voeten, & Oud, 2001).

Early literacy skills represent at least four key domains (NELP, 2008; Sénéchal, LeFevre, Smith-Chant, & Colton, 2001; Whitehurst & Lonigan, 1998) including (a) alphabet knowledge and concepts about print, or the ability to recognize and produce letter names and sounds and understand conventions of written text (McBride-Chang, 1999); (b) phonological awareness, or the ability to detect and manipulate words at the level of phonemes, the smallest units of spoken language (Anthony, Williams, McDonald, & Francis, 2007); (c) oral language, or a child's expressive and receptive vocabulary and communication skills (Dunst, Trivette, Masiello, Roper, & Robyak, 2006); and (d) comprehension, or the ability to gain information and draw inferences from written and/or spoken language (Snow et al., 1998). However, the continuum of how early comprehension skills connect and contribute to later comprehension is still unclear. A limited number of conceptual models are present in the research to demonstrate the development of early comprehension, with even fewer supported by empirical evidence.

## Theoretical Models of Early Comprehension

Current research indicates there are factors that might contribute to the development of early comprehension, such as aspects of socio-cultural theory (Rojas-Drummond, Torreblanca, Pedraza, Velez, & Guzman, 2013) and social-cognitive learning theory (Glenberg & Gallese, 2012). Here, we review two current perspectives to aid in understanding the progression of young children's early comprehension.

Dooley and Matthews (2009) posit that early comprehension is the cumulative sum of a tri-part processing system used to decode graphic symbols represented in words and images. The three parts of this system include the child (the early reader or decoder), the symbol present for the new concept (the actual text or picture), and the relationship between the child and the symbol (the activity in which the child encounters the new word or picture). Together, these components work in dynamic interaction to provide meaning for the child. As the child has additional experiences with the specific symbol within his or her socio-cultural context, meaning evolves, and the child develops a more complete understanding of the concept. For example, if a child's first encounter with the word *glasses* is when a child sees a person in a book wearing glasses, his or her understanding may be limited to the context of their experience within this interaction and will not yet include the meaning of glasses as the plural of drinking glass. However, a similar child might come from an environment where both parents and the child wear glasses, so encountering the word *glasses* would present a different level of understanding but may still be limited to only the concept of glasses as visual aids. Further, a third child who encounters the word *glasses* and already has an understanding of glasses in the context of drinking vessels would add the new meaning of glasses (as visual aids) to his or her repertoire of information—supporting an evolved meaning to the word *glasses.*

To support this tri-part system, Dooley and Matthews (2009) introduced three principles that set the stage for how comprehension may develop. First, they posited that young children develop comprehension in ways that are complementary, but different from adults. This notion is also supported by van den Broek and others in the field, indicating that preschool-age children are able to engage in the same inferential process and make meaningful connections, just as adults do (van den Broek, Kendeou, Lousberg, & Visser, 2011). However, those inferences are limited relative to the child's age. For example, preschool-age children are able to make relations between concrete, external, and individual events, but have a more difficult time comprehending relations between abstract, internal, and clusters of events (van den Broek et al., 2011).

Second, as events occur in children's lives, they create a system of symbolic reference. Therefore, a child's concepts of the words they encounter develop across time and experience. As age and experience interact, children build a schema around each word they learn. As children age, these schemas become more homogeneous, conforming to the societal definition of each new word. In this way, symbols or referents for words evolve over time until a complete understanding is achieved (Dooley & Matthews, 2009).

Third, comprehension is supported by socio-cultural factors, such that children are motivated to learn new words based on the influence of primary caregivers and the early social environments in which they interact (Dooley & Matthews, 2009; Rojas-Drummond et al., 2013). Primary caregivers and peers in a young child's life directly influence a child's conceptions about new experiences. These influences also guide the development of comprehension skills over time to provide a context for understanding. It is because

of these socio-cultural factors that children develop unique and specific vocabularies related to their environment. For example, a child who is raised in a home of musicians may have an extensive vocabulary including musical instruments, tempo and tone, just as a child who is raised in a home with a baseball aficionado might have an unusual depth and breadth of vocabulary around baseball players or pitches.

Much like Dooley and Matthews (2009), van den Broek and colleagues contributed to conceptual arguments by further supporting that comprehension develops in preschool-age children in relationship to other knowledge but is experienced on a separate but complementary trajectory (2005). That is, van den Broek et al. reported comprehension development occurs alongside early reading skills, not in succession to them, suggesting assessment and instructional practices should attend to comprehension during the early years as a unique construct, rather than embedded in existing research regarding reading and knowledge acquisition.

Foundations from Dooley and Mathews's (2009) and van den Broek et al.'s (2005) model of comprehension suggest that early comprehension is a multi-dimensional phenomenon, created by building a family or collection of ideas, skill, and activities. For example, when students are asked, "What is a good teacher?" each could produce a different family of characteristics or skillset that may be necessary to their own definition. Students could identify a teacher as a "good teacher" for different reasons—because they are academically rigorous, compassionate, patient, or yield a high percentage of high school graduates. In this way, children build different concepts of how they understand a word, question, or experience.

Complementary to understanding comprehension as a collection of ideas, skills, or activities, is the more specific notion of making meaningful relations in text or narrative. Evidence suggests children use the context of the story, or the other sentences close in proximity to the target sentence to create relations. For example, if a teacher read two sentences: "The milk spilled on the table" and "The girl was all wet," van den Broek et al. (2011) indicated that adults and children alike quickly create the relation that the girl's wet clothing was attributed to the milk. This type of interaction happens in one of two ways, either an easy and nearly effortless process that happens automatically, as suggested with the process in the previous sentence, or in a slow, intentional, and strategic process that requires the attention and cognitive focus of the child that occurs when sentences have distal and indirect connections (van den Broek et al., 2011).

Taken together, Dooley and Mathews's (2009) conceptual presentation of early comprehension and van den Broek's further contributions to early comprehension (van den Broek et al., 2005; van den Broek et al., 2011) suggest a foundation comprised of multi-dimensional concepts and meaningful relations, supplemented with three core process variables: type, efficiency, and content. First, van den Broek notes that adults and children engage in the same processes, but types of inferences are different as children age. For example, when completing story recall tasks, children and adults alike can recall similar static information, but the relations they establish between the referent information may be different. That is, inferences occur through early childhood in a continuum of complexity. Initially, children's inferences are limited to concrete physical relations that occur close together in time. As children age, they move to concrete physical relations between distant events and then to casual relations involving character goals, emotions, and desires. Finally, children experience comprehension as hierarchical and thematic relations between clusters of events and to the most developed of inferences such as translation of a story theme into a moral or lesson (van den Broek et al., 2011).

Second, parallel to Dooley and Matthews's (2009) model, van den Broek suggests that the core process variable of comprehension efficiency improves over time, such that experiences of drawing on a particular concept or word are reinforced or reduced over time. That is, experiences that align with correct interpretations of a word are reinforced by others, while experiences that include misinterpretations are extinguished. For example, if a child uses the word "long" to describe something in the context of its length, he will be reinforced of this use by others understanding and agreeing with his statement ("Yes, that is a very long rope!"). However if he then generalizes the word long to be used to illustrate height by saying, "David is very long!" others will extinguish the use of long by noting that a person is not long, instead the appropriate term is *tall*. The more experience a child has, the more complete the concept or schema. In this way, children build efficiency in developing accurate representations of a word presented in narrative or written text.

Finally, comprehension is also affected by the content delivered. Similar to Dooley and Matthews's (2009) presentation of socio-cultural influences, van de Broek presents each opportunity for comprehension as a function of the information with which the child has had experience with. Evidence suggests children can and do gather large amounts of information with targeted interest groups (e.g., dinosaurs, chess, etc.; Chi & Koeske, 1983). For example, children interested in dinosaurs may independently form a rich understanding of dozens of different species and characteristics of dinosaurs, allowing for a deep capacity in dinosaur content.

In sum, early comprehension is a valuable and meaningful component of early literacy that develops in conjunction with other skills (e.g., alphabet knowledge, phonological awareness etc.). However, the preliminary understanding of

the nature of comprehension described here is not enough to support effective instruction and intervention during the preschool years. We also need robust assessments to aid in understanding the progression of early comprehension skills with pre-kindergarten (Pre-K) students. In the following sections, we describe how such an assessment model may be pursued within the context of Response to Intervention.

## Response to Intervention

To appropriately address the needs of all children's comprehension skills, assessment and intervention practices must be tailored to provide a match between a student's skill level and instructional content (D. Fuchs, Fuchs, & Compton, 2012; C. R. Greenwood et al., 2011). The Response to Intervention (RTI) model is uniquely suited to address the varied needs of young learners by implementing a tiered system of assessment and intervention.

RTI is a framework to identify, monitor, and intervene with students based on individualized student academic need (D. Fuchs & Fuchs, 2006; C. R. Greenwood, Kratochwill, & Clements, 2008). Students are assessed to determine level of current performance, and intervention services are provided to match this performance level in one of three tiers. Tier 1 features high-quality evidence-based instruction, with complementary periodic screening. Tier 2 provides increased support for those students not making adequate progress in the general universal Tier 1 curriculum and is often presented as small group instruction along with more frequent progress monitoring to evaluate student performance. Tier 3 provides intensive, targeted, and individualized intervention and complementary progress monitoring for those students who continue to make limited progress with additional intervention.

In an RTI model, measures used to assess early literacy skills must function in two ways (D. Fuchs & Fuchs, 2006; C. Greenwood, Carta, McConnell, Goldstein, & Kaminski, 2008). First, data from measures must be able to be utilized to identify individual students who might require a more intensive level of intervention (Margolis, 2012). Second, for those students who are candidates for more intensive instruction and intervention, measures must accurately monitor progress over brief periods of time to continually evaluate if students are improving relevant skills during intervention (Margolis, 2012). Therefore, these measures must be psychometrically robust and logistically feasible, allowing educational professionals to gather meaningful data to inform instructional and intervention decisions. At the same time, assessments that demonstrate utility in an RTI model must also achieve additional empirical and pragmatic criteria. To demonstrate psychometric utility in assessing performance over brief periods of time, measures should obtain less than 20% of children with a score of zero, and produce skew and kurtosis values less than an absolute value of 1 (Wackerle-Hollman, Schmitt, Bradfield, Rodriguez, & McConnell, 2014). In addition, measures must also adhere to General Outcome Measurement (GOMs) tenets, including development of measures that are brief, (between 1 and 2 min per task), easy-to-administer, easy-to-interpret, related to long-term goals, and are inexpensive (or easily attainable). GOM instruments also demonstrate longevity (can be used for at least one academic year), reliability, validity, and sensitivity to growth over time, with utility as a progress monitoring measure (L. D. Fuchs & Deno, 1991).

Finally, pragmatic standards specific to RTI must also be achieved. First, measures must provide teachers and administrators opportunities to examine data and make relevant and efficient instructional changes. Second, measures used in an RTI model may benefit from utilizing a criterion-referenced standard of performance, rather than comparison with a normative peer group, which limits the assessor's ability to evaluate a student's absolute skill level. Third, measures must include a psychometrically robust pool of items to represent the construct of interest, rather than a small number of items which may not provide enough information about areas of skill deficit to inform instruction and intervention.

Currently, few standardized assessments of comprehension for preschool-age children exist, including, *Renfrew Bus Story* (Cowley & Glasgow, 1994), *Clinical Evaluation of Language Fundamentals*–2nd Edition (*CELF Preschool–2*; Wiig, Secord, & Semel, 2004), and the *Test of Auditory Comprehension of Language*–3rd Edition (Carrow-Woolfolk, 1999). However, based on the measurement criteria previously described, these types of standardized assessments do not meet the characteristics of RTI. In contrast, a wealth of GOMs exist for use in K–12 settings including the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and Curriculum-Based Measures (Deno, 1992) with complementary measures designed for Pre-K use, such as the Individual Growth and Development Indicators (IGDIs, McConnell, Priest, Davis, & McEvoy, 2002). IGDIs include two versions, the 1.0 version, designed for use in instructional decision making through fluency-based assessments of alliteration, rhyming, and picture naming, and the 2.0 version specifically designed for use within an RTI framework (CEED@UROC, 2011). IGDIs 2.0 include five tasks: Picture Naming, Rhyming, Which One Doesn't Belong (WODB), First Sounds, and Sound Identification. All IGDI 2.0 tasks, and more specifically the comprehension measure, WODB, uses item response theory (Gorin, Embretson, & McKay, 2008; Wackerle-Hollman, Bradfield, McConnell, Albano, & Rodriguez, 2011) within Wilson's (2005) measurement construction framework. Wilson's framework was used to conceptually support item design, define item characteristics such as target responses, scoring rules, and to statistically model item performance (for a detailed description of

Wilson's model, see Wilson, 2005). Following Wilson's suggestions, work on IGDIs 2.0 used Rasch modeling (Albano, Rodriguez, McConnell, Bradfield, & Wackerle-Hollman, 2011). Rasch modeling provides an approach that considers both student ability (person parameter) and item-level statistics (item parameter). By locating items and student abilities on the same scale, we can better examine if items are available for students at their given ability levels. Therefore, creating items that surround and include the distribution of student ability offers the most parsimonious assessment of ability.

This article presents the process used to develop, pilot, and validate the new comprehension IGDI 2.0 measures. These measures, utilizing the strengths of GOM and psychometric advances related to Wilson's model and Rasch modeling, may provide a foundation for a robust measurement system for use in an early childhood RTI model.

Comprehension, as previously described, is an area of early literacy that has proven difficult to capture developmentally, theoretically, and empirically (Dooley & Matthews, 2009; van den Broek et al., 2011); therefore, the measurement of comprehension skills was conducted over the course of three studies. Study 1 includes the primary development and initial measure design and selection, pilot psychometric and practical examination of several potential measures of comprehension. Study 2 includes further refinement and a selection of measures to undergo further testing. Study 3 includes the collection of evidence for validity and reliability with the refined set of measures.

## Study 1: Developing and Piloting New Measures

The purpose of Study 1 was to develop, pilot, and evaluate the comprehension IGDIs. Specifically, this study sought to answer the following research questions:

*Research Question 1:* To what extent do the measures demonstrate characteristics of General Outcome Measurement (GOM)?

*Research Question 2:* What are the demonstrated item-level characteristics?

*Research Question 3:* What is the relation between the comprehension IGDI measures and also to other standardized measures of early comprehension?

To begin, we investigated the essential knowledge, skills, and abilities related to comprehension among preschool children.

### Defining Comprehension

A literature review was completed to gain more insight into the comprehension construct. A keyword search for the term *reading comprehension* within the Psych Info database yielded 8,810 results. These were limited to 185 articles in peer-reviewed journals that were published in English and were focused on preschool-aged children. The titles and abstracts were reviewed to identify 33 articles that included an operational definition of one of the following terms: discourse, listening, or narrative comprehension, text, passage, or story comprehension. These words and phrases were chosen to provide an inclusive lens of all research articles to date that offer a perspective on early comprehension. Target words and phrases were documented and supplied from seminal research articles including Dooley and Matthews's (2009) and van den Broek's work from 2005 to 2011.

A review of the literature identified two skill areas representing the construct during the preschool years: text comprehension and listening comprehension. Text comprehension is defined as the ability to understand and interpret text, including pictures and symbols (Dunst et al., 2006; Storch & Whitehurst, 2002). Listening comprehension is the ability to understand and interpret what is spoken aloud and infer meaning from what is heard (Skarakis-Doyle, Dempsey, & Lee, 2008).

### Method

*Participants and setting.* A total of 44 preschoolers enrolled in child care centers around a midwestern metropolitan area participated in this study. The child care centers selected represented a diverse sample of students based on socioeconomic status and ethnicity with 64% representing private-pay preschool environments, 23% representing urban public Pre-K programming, and 13% representing Head Start. Students recruited for this study were 68% Caucasian, 22% African American, 4% Asian American, and 6% other ethnicities. Data collection occurred during the winter of 2010 with children ranging in age from 39 to 63 months. Twelve of the children were 3-year-olds (36–47 months), 27 children were 4-year-olds (48–59 months), and 5 children were 5-year-olds (60–71 months). In all, 55% were male.

*Measures.* Four IGDI comprehension tasks and one standardized criterion measure are described here. The IGDI comprehension measures were designed using an iterative process that initially included seven methods for assessing comprehension. Initial measure conception was driven by examining existing K–12 measures and a literature review summarization that yielded the extrapolation of current K–12 methods down to early childhood in adapted approaches and by developing novel approaches to the construct of comprehension through new task development. Initial measures included sentence comprehension, story comprehension, a story titles task, sequencing, picture comprehension, a thematic inferences task and a categories task. Initial piloting reduced the pool to the three measures described here by eliminating those tasks that children did

not understand, that children did not respond to, or that were cumbersome to deliver. Once the three tasks were selected, an item-level analysis was completed to develop measures that included strategic foils (maximally or minimally different), reduced construct irrelevant variance, and were developmentally appropriate for Pre-K students.

*Picture Comprehension*. The first IGDI, Picture Comprehension, is an individually administered task in which the child is presented with an illustrated scene on a stimulus card. Following directions and administration of sample items, the assessor asked the child standardized "wh" questions relevant to the item (e.g., What is happening in this picture? What is he doing? Where are they going? etc.). For example, when shown an item featuring an illustration of a girl wearing a bathing suit and carrying a beach towel, the examiner asked the child, "Where is she going?" The number of cards answered correctly in 2 min was recorded as the child's score.

*Sequencing*. Sequencing, the second IGDI tested, is an individually administered task in which the child is shown three related illustrations and is asked to place them in chronological order, given the first illustration in the sequence. Following instruction on how to complete the task and successful completion of two sample trials, the administrator presented the first image in the sequence (the target image) followed by two choices with the prompt: "This picture comes first" (points to the target image), "Which picture comes next?" (points to choice images). The assessor was provided with all correct responses on the back of each card as indicated by a star on the back of the correct image. The number of cards answered correctly in 2 min was recorded as the child's score.

*Sentence Comprehension*. Sentence Comprehension is an individually administered task in which the child is shown a card depicting three illustrations and is asked to choose (by pointing) the image that best matches the sentence that is presented. Following instruction on how to complete the task and successful completion of two sample trials, the administrator presented the card and described one of the three illustrations. For example, "The bird has flown away." An item with three images is depicted: first a tree with a nest full of eggs and no bird; second, a bird flying toward the tree with the nest; and third, a bird sitting inside the nest. The assessor was provided with all correct responses on the back of each card as indicated by a star on the back of the correct image. The number of cards answered correctly in 2 min was recorded as the child's score.

*Story Comprehension*. Story Comprehension is an individually administered task in which the child is asked to listen to a brief story available to general preschool populations through public media resources (e.g., bookstores, online etc.) such as Leslie Patricelli's *The Birthday Box* or Ruth Krauss's *The Carrot Seed*. The administrator read the book aloud to the child and asked the child a series of 15 brief questions about the story, including prediction questions, direct recall questions, and inference questions. The child's score was the number of correct responses.

*CELF Preschool–2*. Three subtests of the *CELF Preschool–2* (Wiig et al., 2004) were administered to all children as a criterion measure of comprehension. The Core Language score is derived from the three subtests administered. This standardized assessment is an individually administered test of language used to evaluate a child's general language ability. Although the *CELF Preschool–2* is marketed as a language measure, it assesses facets of the comprehension domain, such as listening comprehension, making it an ideal criterion measure. For example, the Sentence Structure subtest evaluates the child's ability to interpret spoken sentences of increasing length and complexity. The remaining two subtests (Word Structure and Expressive Vocabulary) evaluate the child's ability to apply word structure rules and to label illustrations of people, objects, and actions. Test–retest reliability coefficients for 3- to 5-year-olds range from .75 to .79 for Sentence Structure, .79 to .85 for Word Structure, and .78 to .95 for Expressive Vocabulary. The test–retest reliability coefficients for the Core Language score are high, ranging from .87 to .95 for 3- to 5-year-olds (Wiig et al., 2004).

*Procedure*. Measures were individually administered by undergraduate and graduate students who received comprehensive training on all measures to ensure standardization and fidelity of implementation across the study. All assessors were monitored using fidelity checklists during training, received feedback regarding administration errors, and were required to remedy errors before using the assessments with participating children. Initial fidelity statistics indicated assessors reached 81% fidelity on initial attempts and all assessors reached 100% fidelity by the second attempt during training. All assessment sessions were conducted on-site in a quiet hallway near the classroom. The assessments were conducted over the course of three sessions, each lasting from 15 to 20 min. After each session, the participants were allowed to select a small toy from a prize box.

All measures were tested with a pool of 44 students with the exception of Story Comprehension, which included a reduced sample of 26 students. While the Story Comprehension was attempted with all 44 students, 18 students simply did not respond to the prompts indicating that they either were not comfortable with the administration setting or that they did not understand the prompt. These students did not respond to the assessor in any way. As a

**Table 1.** Descriptive Statistics of Comprehension IGDIs Across Three Studies.

| Measure | Wave | *N* | *M* | *SD* | Skewness | Kurtosis | Percentage zero scores |
|---|---|---|---|---|---|---|---|
| Study 1 | | | | | | | |
| Picture Comprehension | N/A | 43 | 13.5 | 4.2 | −0.9 | 1.7 | 1 |
| Sequencing | N/A | 43 | 10.4 | 3.2 | 0.4 | 0.5 | 0 |
| Sentence Comprehension | N/A | 42 | 9.8 | 5.0 | 0.1 | 0.0 | 2 |
| Story Comprehension | N/A | 26 | 5.0 | 2.1 | −0.9 | 1.5 | 2 |
| Study 2 | | | | | | | |
| Sentence Comprehension A | N/A | 88 | 9.1 | 2.2 | −0.1 | −0.5 | 12 |
| Sentence Comprehension B | N/A | 106 | 8.5 | 2.4 | −0.2 | −0.5 | 7 |
| Study 3 | | | | | | | |
| Sentence Comprehension | Winter | 224 | 19.3 | 3.9 | −0.6 | −0.1 | 12 |
| WODB | Winter | 191 | 51.9 | 13.0 | −1.0 | −0.2 | 7 |
| Sentence Comprehension | Spring | 80 | 20.3 | 3.8 | −0.6 | −0.3 | 10 |
| WODB | Spring | 126 | 53.1 | 13.4 | −1.1 | −0.1 | 5 |

*Note.* IGDIs = individual growth and development indicators; WODB = Which One Doesn't Belong.

result, we did not include these occurrences as valid data or zero scores.

## Results

*Evaluation of measure criteria.* Descriptive statistics for the comprehension IGDI measures are presented in Table 1. The descriptive statistics listed align with the suggested measurement criteria of GOMs (e.g., percentage zero scores, skew, kurtosis, etc.). Results indicate the IGDI measures met or exceeded the majority of psychometric criteria for utilization in an RTI model previously described.

*Item-level performance.* Item-level means and item-total correlations for each measure are displayed in Table 2. Item-level means provide information about individual item difficulty. Item-total correlations indicate the degree to which an item contributes to the overall measure and discriminates between those that do or do not have a trait (e.g., comprehension ability). For Sentence Comprehension and Sequencing, 80% of the item means fell between .20 and .80. Conversely, only 36% of the item means for Picture Comprehension fell within the desired range. Sentence Comprehension had the highest percentage of items with item-total correlations greater than .20, while Picture Comprehension again had the fewest.

*Relation between measures.* Intercorrelation coefficients for the comprehension IGDI measures are provided in Table 3. Correlations ranged from low ($r = .28$) to moderate ($r = .53$). The correlations between the IGDI comprehension measures and the *CELF Preschool–2* were moderate ($r = .41$–$.74$), with the exception of Sequencing, yielding a coefficient of $r = .16$.

**Table 2.** Range of Item Means, Means, and Item-Total Correlation Ranges by Measure Across Three Studies.

| | Item means | | Item-total correlations | |
|---|---|---|---|---|
| Measure | Range | % from .20 to .80 | Range | % .20 or above |
| Study 1 | | | | |
| Picture Comprehension | .17–1.00 | 36 | −.35–.67 | 30 |
| Sequencing | .28–.91 | 80 | −.47–.44 | 40 |
| Sentence Comprehension | .19–1.00 | 80 | −.13–.54 | 63 |
| Study 2 | | | | |
| Sentence Comprehension A | .46–.96 | 77 | −.02–.63 | 52 |
| Sentence Comprehension B | .30–.89 | 71 | .04–.64 | 67 |
| Study 3 | | | | |
| Sentence Composition | .30–.98 | 63 | .07–.42 | 70 |
| WODB | .26–.90 | 56 | −.13–.70 | 97 |

*Note.* WODB = Which One Doesn't Belong.

## Discussion

The purpose of this pilot study was to examine the basic psychometric properties of the IGDI comprehension measures and to determine which measures best demonstrated the characteristics of GOMs. Evaluation of each measure included a comparison of descriptive statistics to predefined GOM criteria, examination of correlations both within IGDI comprehension measures and with standardized criterion measures (e.g., *CELF Preschool–2*) to evaluate

**Table 3.** Intercorrelation Coefficients for Study 1 and Study 3.

| Measure | Picture Comprehension | Sequencing | Sentence Comprehension | Story Comprehension | Which One Doesn't Belong |
|---|---|---|---|---|---|
| Picture Comprehension | 1 | | | | |
| Sequencing | .28 | 1 | | | |
| Sentence Comprehension | .49** | .37* | 1 | | |
| Story Comprehension | .53** | .47* | .40* | 1 | |
| Which One Doesn't Belong | N/A | N/A | .54* | N/A | 1 |
| CELF Sentence Structure | .43** | .44* | .63** | .46* | .74** |
| CELF Word Structure | .54** | .16 | .55** | .60** | .57** |
| CELF Expressive Vocabulary | .66** | .41* | .58** | .66** | .69** |

*Note*. Which One Doesn't Belong was the only measure added in Study 3 and therefore results are presented collapsed across Study 1 and Study 3. Which One Doesn't Belong correlations represent data reported in Study 3, whereas all other correlation values represent data reported in Study 1. CELF = Clinical Evaluation of Language Fundamentals.
*$p < .05$. **$p < .01$.

validity. Finally, item-level performance data were also examined to provide additional support for selecting measures for further development and testing in Study 2.

*Essential GOM criteria.* There are a number of essential criteria to consider when evaluating the extent to which a measure demonstrates the characteristics of a GOM. As described previously, a GOM is quick and easy to use and interpret. The four comprehension measures evaluated in this study are brief in their administration (2 min each) and are easy to use. Another defining characteristic of a GOM is that it is psychometrically sound. Specifically, a psychometrically sound measure has a skew and kurtosis value that is less than the absolute value of one, and has a low number of zero scores. Results suggested that all the measures had reasonable skew and kurtosis and a low number of zero scores. The Story Comprehension measure yielded limited variability in total scores as a result of a constrained sample ($n = 26$), therefore inferences about utility are restricted.

*Validity evidence.* To examine the relation between the measures, intercorrelation coefficients were calculated. The correlations ranged from low (Sequencing and Picture Comprehension) to moderate (Story Comprehension and Picture Comprehension). The variability in the Intercorrelation coefficients suggests that the measures are not capturing the comprehension domain in the same way.

Criterion-related validity evidence was examined by calculating correlations between each comprehension IGDI and the core subtests of the *CELF Preschool–2*. With the exception of Sequencing, all of the measures demonstrated moderate correlations with the *CELF Preschool–2*.

*Item-level functioning.* Item-level information was not calculated for Story Comprehension due to the restricted amount of information obtained from that measure. Item means, or $p$ values, and item-total correlations were examined to understand how each measure functioned at the item level. The item mean indicates the specific difficulty of the item. The acceptable range of item means was within .20 and .80. Any item means outside this range did not meaningfully contribute to the assessment because they were either too difficult or too easy. Picture Comprehension had the lowest percentage of items that fell within the desired range (36%), whereas many were too easy.

Item-total correlations are another criterion that helps determine how an item contributes to an assessment. Specifically, item-total correlations indicate the extent to which the items discriminate between children with high ability in the construct and children with low ability in the construct. Items with an item-total correlation of .20 or higher were determined to be contributing to the assessment. Sentence Comprehension had the highest percentage of items (63%) with item-total correlations greater than .20, suggesting that this measure had more items that were able to successfully discriminate between children's ability levels. Picture Comprehension again had the lowest percentage of items that met the desired criteria (30%).

Considering all the relevant data, it was determined that Sentence Comprehension would move on to Study 2 for further testing and refinement. This measure demonstrated the most consistent combination of GOM characteristics, validity evidence and item-level functioning. It was also decided to combine Picture Comprehension and Sequencing into one task. Item-level data were used to make the decision to combine the

two tasks because neither demonstrated the characteristics of an effective measure on its own. The best items from each task were therefore combined to create one measure.

## Study 2

Building on the results of the first study, Study 2 sought to answer similar research questions with a larger sample of children. Study 2 involved the following research questions:

*Research Question 1:* What are the basic psychometric properties of the IGDI comprehension measures?
*Research Question 2:* What are the demonstrated item-level characteristics?
*Research Question 3:* What is the relation between the comprehension IGDI measures and standardized measures of early comprehension?

With Study 2 came a shift in the measurement framework from Classical Test Theory to the Rasch Model. The Rasch Model places items on a scale based on item difficulty defined as the ability required to have a 50% probability of correctly responding to the item, locating the average item at 0 (typically resulting in an ability scale from −4 to +4). Based on their performance on the IGDI items, children are assigned Rasch scores, which reflect their ability in the given domain, relative to the location of the items. Thus, items and children are placed on a common scale, defined by the items as representation of the construct.

### Method

*Participants and setting.* Participants for Study 2 were recruited from early child care centers around a midwestern metropolitan area. Participants were 46% Caucasian, 37% African American, 7% Hispanic, 5% Native American, 4% Asian American, and 1% other ethnicities. Study 2 occurred during the summer of 2010 and included 196 participants ranging in age from 36 to 69 months. Seventy of the children were 3-year-olds (36–47 months), 93 children were 4-year-olds (48–59 months), and 32 children were 5-year-olds (60–69 months). There were limited data available on the gender distribution of the sample; for this reason, gender information is described for 91 of the 196 participants. Of the 91 participants, 61% were female and 39% were male.

*Measures.* The comprehension IGDIs administered in Study 2 included Sentence Comprehension and the combined Picture Comprehension and Sequencing (PCS) measures. Before being tested in Study 2, item-level revisions were made to each measure. Specifically, poorly functioning items were discarded or edited.

For Sentence Comprehension, 27 items were tested. Three items were discarded from the previous study because of poor item-level functioning, and 19 were edited. Edits to the items included both changes to the prompt and changes to the images. Eight items received prompt changes. As an example, a card from Study 1 originally read, "The dog is barking at the cat." For Study 2, the prompt was changed to read, "The dog is chasing the cat." The change was made because it was too difficult to capture the original sentence with an image. A dog that is chasing a cat is better depicted through an image than a dog barking at a cat.

Items also received image changes. Figure 1 shows an example of the kind of image changes made to a given item. The prompt for this particular item read, "Dad works." The first set of images (top row) represents how the item looked in Study 1, whereas the second set of images (bottom row) represents the item in Study 2. The target image depicting dad working, the middle image, was changed to make the room look more like an office. The other two images, the distractor images, were altered so that they were maximally different from the target image.

For PCS, a total of 38 items were tested. The best functioning items from each measure were combined (17 from Picture Comprehension, 21 from Sequencing). The administration instructions were also altered to reflect both measures.

*Procedure.* Study 2 participants were part of a larger study where each student was administered nine IGDI 2.0 measures across four domains (Phonological Awareness, Oral Language, Alphabet Knowledge, and Comprehension) during the summer of 2010. Other tasks administered included three oral language tasks: Picture Naming, Definitional Vocabulary, and WODB; two phonological awareness tasks: First Sounds and Rhyming; and two alphabet knowledge tasks: Sound Identification and Letter Identification. Only the IGDI early comprehension measures are reported here, which included two tasks: Sentence Comprehension and PCS. Measures were individually administered by undergraduate and/or graduate students who received comprehensive training on all measures to ensure standardization and fidelity of implementation across the study. All administration procedures and procedures for assessor fidelity of implementation were engaged in the same process described in Study 1. Initial fidelity statistics indicated assessors reached 84% fidelity on initial attempts and 100% fidelity by the second attempt during training. To decrease the burden on children (given each task had between 20 and 30 items), a bundling procedure was created. The bundles were constructed such that each item was administered to at least 100 children. This was critical to collect sufficient item-level data to meet Rasch model requirements. Each bundle had four sample cards and five common cards that remained constant across bundles.

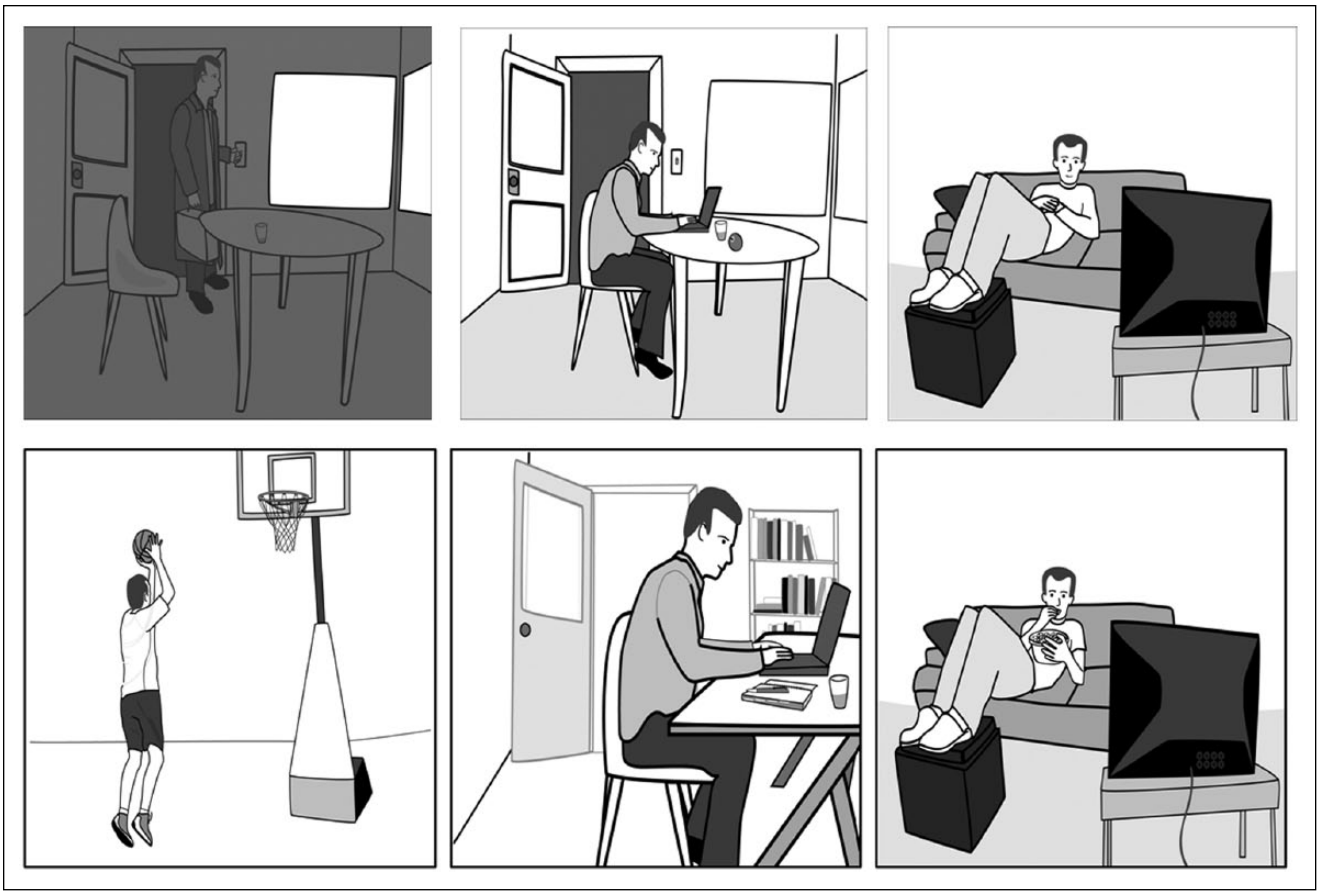

**Figure 1.** Schematic of how images from an item in Study 1 (top row) changed in Study 2 (bottom row).
*Note.* The prompt for the item read, "Dad works."

## Results

*Characteristics of measures.* Results of the PCS measure are not presented because administration of the measure discontinued midway during the data collection period. The measure did not perform as expected and did not align with the characteristics of GOMs in terms of brevity and ease of use.

Table 1 provides the descriptive statistics for Sentence Comprehension. Results are presented by bundle. Results suggested Sentence Comprehension yielded adequate mean, standard deviation (*SD*), skew, kurtosis, and percentage of zero scores.

*Item-level performance.* Item-level statistics for Sentence Comprehension are provided in Table 2. To reiterate, the item means provide information about individual item difficulty. Item means ranged from $r$ = .30 to .96.

*Relations between measures.* To address the second research question, correlations with the *CELF Preschool–2* were calculated. Table 4 demonstrates moderate correlations for each bundle of Sentence Comprehension with each subtest of the *CELF Preschool–2*. Correlations yielded values of $r$ = .36 to .58.

**Table 4.** Correlations With *CELF Preschool–2*.

| Measure | CELF Sentence Structure | CELF Word Structure | CELF Expressive Vocabulary |
|---|---|---|---|
| Sentence Comprehension A | .56** | .36** | .37** |
| Sentence Comprehension B | .58** | .49** | .56** |

*Note.* CELF = Clinical Evaluation of Language Fundamentals.
**$p$ < .01.

## Discussion

Study 2 represented a larger-scale investigation into the utility and validity of the IGDI comprehension measures. The purpose of the study was to examine the basic psychometric properties of the IGDI comprehension measures and to ensure that they align with the characteristics of GOMs. To accomplish this, descriptive statistics and criterion-related

validity coefficients between the measures and the *CELF Preschool–2* were examined. Item-level performance data were also examined to provide additional support for further development.

*Essential GOM criteria.* Adopting the same evaluation criteria used in Study 1, each set or bundle of Sentence Comprehension demonstrated reasonable ease of use, brevity, percentage of zero scores, skew and kurtosis values. This piece of evidence increases our confidence in the measure and indicates that Sentence Comprehension meets the criteria of a GOM.

*Validity evidence.* Criterion correlations with the *CELF Preschool–2* were conducted as a source of criterion-related validity evidence. Each bundle of Sentence Comprehension demonstrated moderate correlations with the *CELF Preschool–2* suggesting that Sentence Comprehension continues to be an adequate indicator for measuring early comprehension skills.

*Item-level functioning.* Item means were examined to get a better sense of how each measure functioned at the item level. The item mean indicates the specific difficulty of the item. The acceptable range of item means was within.20 and .80. Any item falling outside this range is not contributing in any meaningful way to the overall measure. More than 70% of items in each Sentence Comprehension bundle fell within the desired range.

Using the information collected from the previous studies, measures were selected for Study 3 based on a number of important factors. First, the extent to which the measures exhibited strong validity evidence with an established criterion measure within the comprehension domain was considered. A second factor that was considered was the degree to which the measures aligned with the GOM characteristics. Third, each measure's item-level information was considered before deciding which measures moved forward in the study. As stated previously, the PCS measure failed midway during data collection and therefore was not selected for further testing. It was determined that Sentence Comprehension remained the best candidate for the comprehension domain and required further testing and refinement.

## Study 3

The purpose of Study 3 was to further examine the feasibility and utility of Sentence Comprehension and to begin evaluating the WODB task as a measure of early comprehension. The WODB task, previously developed as an oral language measure, transitioned to the comprehension domain after criterion correlations demonstrated higher relations between this measure and the *CELF Preschool–2*. In Study 2, WODB was used within the Oral Language domain (as reported in Bradfield, Besner, Wackerle-Hollman, Rodriguez, & McConnell, 2014). Given the lack of transparent transition from oral language to comprehension, the WODB task was originally conceptualized as an oral language task because of its requirement to know and use language to make categories. However, after review of Study 2 performance, it was reasoned that WODB performance more accurately represented a child's level of categories or schemas for grouping and developing relations between words and concepts. Therefore, the WODB task was moved to the comprehension domain. Study 3 sought to answer the following research questions:

*Research Question 1:* What are the demonstrated item-level characteristics?

*Research Question 2:* What is the relation between the comprehension IGDI measures and also to other standardized measures of early comprehension?

### Participants and Setting

A total of 599 children participated in Study 3. Data collection occurred over two time periods, winter (March) and spring (May) of 2011. Participants were selected from local and distant preschool sites across the United States in eight states: Alaska, Michigan, Minnesota, Kansas, Missouri, Ohio, Indiana, and Oregon. Demographic data regarding ethnicity and SES were not available at the student level for this study because participating site teachers administered the IGDI tasks as part of their educational assessment package. However, all sites were included in school districts that reported aggregated data as part of state reporting requirements (information on each district's ethnicity and SES may be requested from the author). Data to calculate age were not available for all children in the study. From the data that were available, the children's ages ranged from 36 to 69 months. Thirty-six of the children were 3-year-olds (36–47 months), 134 children were 4-year-olds (48–59 months), and 154 children were 5-year-olds (60–69 months). In all, 54% were male.

*Measures.* Based on the conclusions from Study 2, Sentence Comprehension and WODB were selected for use to assess early comprehension in Study 3. A complete description of the Sentence Comprehension measure can be found in Study 1.

*WODB.* WODB is an individually administered task in which the child is shown cards with three colored pictures. The child is asked to identify which of the three pictures does not belong with the others. Before the task begins, the assessor demonstrates how to correctly respond using four practice cards. The first two sample cards are modeled by the assessor; the child is asked to respond to the last two

sample cards with constructive feedback given for incorrect responses. The child's final score is the total number of questions answered correctly.

*Procedure.* Study 3 participants were included as part of a larger validation study that included the administration of six IGDIs, representing the domains of early literacy (Phonological Awareness, Oral Language, Alphabet Knowledge, and Comprehension) during the winter and spring of 2011. The six tasks administered included the two comprehension measures, WODB and Sentence Comprehension; two phonological awareness tasks, Rhyming and First Sounds; one oral language task, Picture Naming; and one alphabet knowledge task, Sound Identification. Data collection for Study 3 occurred over the course of two waves with the criterion measure administered during the winter data collection period. Assessment protocols mirrored those described in Studies 1 and 2; however, for all states besides Minnesota, classroom staff provided assessment interactions rather than trained graduate research associates (GRAs). In addition, all training for other sites was provided digitally using Adobe Connect software. Initial fidelity statistics indicated assessors reached 79% fidelity on initial attempts and 100% fidelity by the second attempt during training.

For each comprehension measure, item-level revisions were made to improve image quality and to remove any construct irrelevant features that could inappropriately influence how children respond to the tasks. For example, Sentence Comprehension items changed from low resolution, black and white images in Study 1 to high resolution, color images in Study 3. Select WODB items were improved by removing unnecessary background features and to the extent possible, all illustrated images were replaced with photographic images. Administration instructions were not modified.

During administration, each IGDI, excluding Sentence Comprehension, had roughly 60 items. There were a total of 68 items for WODB and 27 items for Sentence Comprehension. To reduce the burden placed on children, each measure was administered across 3 days with a maximum of 20 items presented in each interaction at each wave. In addition, for select sites, a matrix sampling technique system was instituted to further reduce child fatigue. As a result, any one student could have received measures in two domains, with the criterion measure sampled across the subsample ensuring that close to 60 participants received the *CELF Preschool–2*.

## Results

*Characteristics of measures.* Descriptive statistics for Sentence Comprehension and WODB are provided in Table 1. Results are presented by each wave of data collection. For both Sentence Comprehension and WODB, distributions suggest the measures do not indicate significant concern as illustrated by skew, kurtosis. However, the percentage of zero scores indicates a small percentage of students (5%–12%) were unable to respond to any of the WODB and Sentence Comprehension items.

*Item-level performance.* Item-level means and item-total correlations for each measure are displayed in Table 2. For Sentence Comprehension, 63% of the item means fell between .20 and .80. For WODB, 56% of the item means fell within the desired range. WODB had the highest percentage of items with item-total correlations greater than .20.

*Relation between measures.* The correlation between the two comprehension IGDIs was moderate ($r$ = .54). In addition, criterion correlations are presented in Table 3. Correlations with the IGDI measures and the *CELF Preschool–2* were moderate, ranging from $r$ = .55 to .74.

## Discussion

This study presented a large-scale field test of the comprehension IGDIs. The purpose of the study was to further examine the feasibility and utility of Sentence Comprehension and WODB as a measure of early comprehension. Descriptive statistics, criterion correlations, and item-level statistics are offered as evidence of the measures' utility. This study represented a diverse sample of students across five states in the United States.

*Essential GOM criteria.* As described earlier in this manuscript, the hallmarks of a GOM are that the measure is brief, easy to score and administer, and psychometrically robust. Examining the descriptive statistics is one way of evaluating a measure's alignment with GOM characteristics. The score distributions of each measure suggest that neither indicate significant concern as illustrated by skew, kurtosis. In general, Sentence Comprehension showed an improved distribution of scores over WODB. However, it is evident by the percentage of zero scores that some student abilities are not captured on the Sentence Comprehension or WODB tasks, indicating a floor effect for the lowest performers. In addition, data collectors reported both tasks were easy to collect and interpret, brief to administer, and engaging for students.

*Item-level functioning.* Item means and item-total correlations were examined to understand task functioning at the item level and how each item contributes to the overall measure. Whereas Sentence Comprehension had more item means that fell between the .20 and .80 range, WODB had the highest percentage of items with item-total correlations greater than .20. Although WODB had more items that were outside the item mean range, it had more items that were able to successfully discriminate between children's ability levels (i.e., item-total correlations).

*Validity evidence*. To establish criterion-related validity for the comprehension measures, a subsample of students received the *CELF Preschool–2*. WODB yielded higher correlations with the comprehension criterion measure than did Sentence Comprehension, indicating that WODB may be a better measure of early comprehension skills.

Based on the information presented here, as well as additional considerations, including cost of development (Sentence Comprehension is significantly more expensive to develop because it requires an illustrator), potential for writing new items, and adequacy of the items in the current scale, WODB was suggested as the comprehension measure candidate for inclusion in the next cadre of IGDIs.

## Implications for Practice

RTI in early childhood is gaining steam as a new and promising approach for addressing early academic challenges, shifting the emphasis away from a "wait to fail" approach and toward a responsive and preventive approach to child intervention and assessment (C. R. Greenwood et al., 2011). The IGDIs 2.0 have been designed to be uniquely suited for use within an RTI model by including items that can reliably detect the ability levels of preschool-age children and, thus, produce scores that can inform decisions about the need for intervention. As a result, IGDIs 2.0 demonstrate promise in an RTI framework and will be further developed through an iterative process—beginning with the methodology described within this article.

More specifically, this manuscript delineates the iterative development process of the IGDI 2.0 comprehension tasks. The IGDI 2.0 comprehension tasks were developed to align with current theories of early comprehension. In particular, the final IGDI 2.0 task, WODB, aligns with van den Broek's (2005) position that early comprehension represents a child's ability to build a family or collection of knowledge. By asking children to draw on their existing knowledge of categories and collections, we can gather important information about their ability to comprehend information for use within the early literacy domain.

However, even with the support of emerging theories of early comprehension it is important to note that the IGDI 2.0 comprehension tasks still face significant challenges and require additional research and development. First, measures of early literacy that demonstrate utility in an early literacy RTI model must be able to capture the full continuum of skill development within a given construct. Specific to early comprehension, an ideal measure for an RTI model would have a limited, if any, floor effect. That is, the abilities of the lowest performers would be present on the task. IGDI 2.0 WODB, at present, has a non-negligible floor effect, suggesting room for improvement in task design.

At the same time, the limited understanding of early comprehension available to inform task design provides a sort of catch-22. IGDI 2.0 comprehension work will continue to explore new ideas about how to understand how children go through the process of making collections and families of knowledge that may represent lower levels of early comprehension abilities.

It is also important to note that in the design process of IGDIs 2.0 comprehension tasks, we were aware of and attended to the relation between vocabulary and oral language development and comprehension. Whereas the theories presented here suggest they are complementary constructs, they also are entangled through the reciprocal relationship present between knowledge and understanding. Further, we did not develop the IGDI 2.0 comprehension tasks to the exclusion of vocabulary and oral language. Instead, a complementary IGDI 2.0 task, Picture Naming, was developed to parallel the comprehension task (see Bradfield et al., 2014).

Nevertheless, given the paucity of early GOM screening measures for comprehension, the work described within this manuscript provides evidence for the IGDI 2.0 comprehension tasks suggesting potential for use within an early childhood RTI model. Results indicate the current task is able to accurately capture the ability levels of most preschool-age students as demonstrated in Study 3 and may provide a resource for assessing the degree to which early comprehension skills have developed during the preschool years.

## Acknowledgments

The authors thank colleagues who assisted with this project, including participating child care centers and programs in the Minneapolis/St. Paul area. In addition, the authors express sincere appreciation for the work contributed by partner CRTIEC sites including Howard Goldstein and supporting staff at the Ohio State University, Ruth Kaminski and supporting staff at Dynamic Measurement Group, and Judith Carta and Charles Greenwood and supporting staff at the University of Kansas. Finally, the authors are indebted to the research team at the University of Minnesota who participated in measure design and data collection: Tony Albano, Amanda Besner, Kate Clayton, Laura Potter, and Megan Rodriguez.

## Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs. McConnell, Bradfield, and Wackerle-Hollman have developed assessment tools and related resources known as *Individual Growth & Development Indicators* and *Get it, Got it, Go!* This intellectual property is subject of technology commercialization and possible licensing agreements through the University of Minnesota. The authors may be entitled to royalties for products related to the research described in this article. This relationship has been reviewed

and managed by the University in accordance with its conflict of interest policies. The opinions and recommendations presented in this article are those of the authors alone, and no official endorsement from the Institute of Education Sciences should be inferred.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by Grant R324C080011, the *Center for Response to Intervention in Early Childhood*, from the Institute of Education Sciences, U.S. Department of Education, to the University of Kansas (Charles Greenwood and Judith Carta, principal investigators).

## References

Aarnoutse, C., Van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing*, *14*, 61–89.

Albano, A. D., Rodriguez, M. C., McConnell, S., Bradfield, T., & Wackerle-Hollman, A. (2011, April). *Scaling with measures of early literacy*. Paper presented at the meeting of the National Council for Measurement in Education, New Orleans, LA.

Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. *Annals of Dyslexia*, *57*(2), 113–137. doi:10.1007/s11881-007-0008-8

Bradfield, T., Besner, A., Wackerle-Hollman, A., Rodriguez, M., & McConnell, S. (2014). Redefining individual growth and development indicators: Oral language. *Assessment for Effective Intervention*. Advance online publication. doi:10.1177/1534508413496837

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, *301*, 419–421. doi:10.1038/301419a0

Carrow-Woolfolk, E. (1999). *Test for Auditory Comprehension of Language–Third Edition*. Austin, TX: Pro Ed.

CEED@UROC. (2011). *Individual growth and development indicators 2.0 identification assessments: Rhyming, first sounds, which one doesn't belong, picture naming and sound identification* (Unpublished assessment). Minneapolis: University of Minnesota, Center for Response to Intervention in Early Childhood.

Chi, M. T., & Koeske, R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*, 29–39.

Cowley, J., & Glasgow, C. (1994). *The Renfrew bus story: Language screening by narrative recall—Examiner's manual*. Centreville, DE: The Centreville School.

Deno, S. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure: Alternative Education for Children and Youth*, *36*(2), 5–10.

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, *29*, 105–122.

Dooley, C.M. (2011). The emergency of comprehension: A decade of research 2000-2010. *International Electronic Journal of Elementary Education*, *4*, 169–184.

Dooley, C. M., & Matthews, M. W. (2009). Emergent comprehension: Understanding comprehension development among young literacy learners. *Journal of Early Childhood Literacy*, *9*, 269–294. doi:10.1177/1468798409345110

Dunst, C., Trivette, C. M., Masiello, T., Roper, N., & Robyak, A. (2006). Framework for developing evidence-based early literacy learning practices. *CELLpapers*, *1*(1). Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v1_n1.pdf

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93–99.

Fuchs, D., Fuchs, L. S., & Compton, D. (2012). Smart RT: A next-generation approach to prevention. *Exceptional Children*, *78*, 263–279.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*, 488–499.

Glenberg, A., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension and production. *Cortex*, *48*, 905–922.

Good, R. H., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills 6th edition (DIBELS)*. Eugene, OR: Institute for the Development of Educational Achievement. Available from http:///dibels.uoregon edu

Gorin, J. S., Embretson, S. E., & McKay, D. (2008). Item response theory and Rasch models. In McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 59–74). SAGE.

Greenwood, C., Carta, J., McConnell, S., Goldstein, H., & Kaminski, R. (2008). *Center for response to intervention in early childhood*. Available from http://www.crtiec.org

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The Response to Intervention (RTI) approach in early childhood [Feature ABI: Y FTI: Y; P]. *Focus on Exceptional Children*, *43*(9), 1–22.

Greenwood, C. R., Kratochwill, T. R., & Clements, M. (2008). *Schoolwide prevention models: lessons learned in elementary schools*. New York, NY: Guilford Press.

Justice, L. (2010). *The language basis of reading comprehension*. Washington, DC: Institute of education sciences grant award, National Center for Educational Research.

Kail, M. (2011). The study of early comprehension in language development: New methods, findings and issues. *Language Interaction and Acquisition*, *2*, 13–35. Retrieved from http://dx.doi.org/10.1075/lia.2.1.01kai

Margolis, H. (2012). Response to intervention: RTI's linchpins. *Reading Psychology*, *33*, 8–10. doi:10.1080/02702711.2011.630600

McBride-Chang, C. (1999). The ABCs of the ABCs: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, *45*(2), 285–308.

McConnell, S. R., Priest, J. S., Davis, S. D., & McEvoy, M. A. (2002). Best practices in measuring growth and development for preschool children. In A. Thomas & J. Grimes (Eds.), *Best*

*practices in school psychology* (4th ed., Vol. 2, pp. 1231–1246). Washington, DC: National Association of School Psychology.

McMaster, K., & van den Broek, P. (2011). *Making the right connections: Improving the comprehension of struggling readers*. Washington, DC: Institute of Education Sciences grant award, National Center for Special Education Research.

National Early Literacy Panel. (2008). *Developing early literacy report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

Robinson, E., Einav, S., & Fox, A. (2013). Reading to learn: Prereaders' and early readers' trust in text as a source of knowledge. *Developmental Psychology*, *49*, 505–513. doi:10.1037/a0029494

Rojas-Drummond, S., Torreblanca, O., Pedraza, H., Velez, M., & Guzman, K. (2013). Dialogic scaffolding: Enhancing learning and understanding in collaborative contexts. *Learning Culture and Social Interaction*, *2*, 11–21.

Sénéchal, M., LeFevre, J., Smith-Chant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy: The role of empirical evidence. *Journal of School Psychology*, *39*, 439–460. doi:10.1016/S0022-4405(01)00081-4

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. *Handbook of Early Literacy Research*, *2*, 173–182.

Skarakis-Doyle, E., Dempsey, L., & Lee, C. (2008). Identifying language comprehension impairments in preschool children. *Language, Speech and Hearing Services in Schools*, *39*, 54–65.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Storch, S., & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947. doi:10.1037/0012-1649.38.6.934

Tompkins, V., Guo, Y., & Justice, L. (2013). Inference generation, story comprehension and language skills in the preschool years. *Reading and Writing*, *26*, 403–429.

van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M., & Lorch, E. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum.

van den Broek, P., Kendeou, P., Lousberg, S., & Visser, G. (2011). Preparing for reading comprehension: Fostering text comprehension skills in preschool and early elementary school children. *International Electronic Journal of Elementary Education*, *4*, 259–268.

Wackerle-Hollman, A., Bradfield, T., McConnell, S., Albano, A., & Rodriguez, M. (2011, April). *Task development and item analysis in innovative measures of early literacy*. Paper presented at the National Council for Measurement in Education, New Orleans, LA.

Wackerle-Hollman, A., Schmitt, B., Bradfield, T., Rodriguez, M., & McConnell, S. (2014). Redefining individual growth and development indicators: Phonological awareness. *Journal of Learning Disabilities*. Advance online publication doi:10.1177/0022219413510181

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*, 848–872.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals–preschool* (2nd ed.). San Antonio, TX: Psychology Corporation.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.