

Redefining Individual Growth and Development Indicators: Phonological Awareness

Journal of Learning Disabilities
2015, Vol. 48(5) 495–510
© Hammill Institute on Disabilities 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022219413510181
journaloflearningdisabilities.sagepub.com



Alisha K. Wackerle-Hollman, PhD¹, Braden A. Schmitt, MA¹, Tracy A. Bradfield, PhD¹, Michael C. Rodriguez, PhD¹, and Scott R. McConnell, PhD¹

Abstract

Learning to read is one of the most important indicators of academic achievement. The development of early literacy skills during the preschool years is associated with improved reading outcomes in later grades. One of these skill areas, phonological awareness, shows particular importance because of its strong link to later reading success. Presented here are two studies that describe the development and revision of four measures of phonological awareness skills: Individual Growth and Development Indicators Sound Blending, Syllable Sameness, Rhyming, and Alliteration 2.0. The authors discuss the measure development process, revision, and utility within an early childhood Response to Intervention framework.

Keywords

preschool age, assessment, early literacy reading, early literacy, phonological awareness, assessment, Response to Intervention, Individual Growth and Development Indicators

Learning to read is one of the most important indicators of academic success (Snow, Burns, & Griffin, 1999). A growing body of research highlights the association between the development of early literacy skills during the preschool and very early elementary grades and improved reading outcomes in later grades (Snow et al., 1999). During the preschool years (ages 3 to 5), research indicates a defined set of skills as precursors, and in some cases prerequisites, to establish the groundwork for learning to read (e.g., National Early Literacy Panel, 2008). These skills, termed *early or emergent literacy*, capture the foundational elements essential for reading in the elementary grades (Whitehurst & Lonigan, 1998).

Both empirical and theoretical research suggests early literacy comprises at least four key domains (McConnell, Wackerle-Hollman, & Bradfield, in press; Senechal, LeFevre, SmithChant, & Colton, 2001; Whitehurst & Lonigan, 1998). These domains include (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) comprehension, or the ability to gain information and draw inference from written and/or spoken language (Snow et al., 1999); (c) oral language, or a child's expressive and receptive vocabulary (Dunst, Trivette, Masiello, Roper, & Robyak, 2006); and (d) 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).

Phonological Awareness

Of the four domains of early literacy identified here, phonological awareness holds particular importance to educators because of its strong link and contribution to later reading success (Anthony & Lonigan, 2004; Muter, Hulme, Snowling, & Stevenson, 2004). Study findings have illustrated that students with strong phonological awareness skills at the preschool and kindergarten level are likely to be more proficient readers at third grade (Muter et al., 2004; Wagner et al., 1997). Furthermore, research recognizes specific skills characteristic of phonological awareness, including rhyming, alliteration, blending, and elision, contribute to robust reading performance. For example, student ability to isolate and identify phonemes at 4 and 5 years of age has been shown to predict student performance on word reading and comprehension tasks in second grade (Muter et al.,

¹University of Minnesota, Minneapolis, USA

Corresponding Author:

Alisha K. Wackerle-Hollman, University of Minnesota Urban Research Outreach and Engagement Center, 2001 Plymouth Avenue, Minneapolis, Minnesota, 55411, USA.
Email: wacke020@umn.edu

2004). Similarly, Lonigan, Wagner, Torgesen, and Rashotte (2007) have found that preschool-age children's performance on blending tasks is highly correlated with beginning reading assessments at the end of first grade.

Based on the critical link between phonological awareness and later reading success, a considerable effort has been placed on identifying the component skills and the conceptual and pragmatic trajectory of the development of those skills within phonological awareness. At least three models exist in the research literature that align phonological awareness with a continuum of early literacy skill development. Goswami (1990) proposed that phonological awareness skills develop in three consecutive phases moving from largest to smallest units of text, where development during the preschool years (Phase 1) is represented by rhyme and alliteration awareness, followed by the development of phoneme-level knowledge and phonemic awareness (Phase 2). Goswami and Bryant close the continuum with the development of beginning reading skills, and as a result, spelling skills, finally culminating in a fluent reading experience (Phase 3; Carroll, Snowling, Hulme, & Stevenson, 2003; Goswami & East, 2000).

Similarly, Gombert (1992) suggested phonological awareness can be separated into two types: epilinguistic awareness and metalinguistic awareness. Epilinguistic awareness is the global awareness of similarities between speech sounds gathered from previous knowledge or environmental stimuli. Metalinguistic awareness is the conscious awareness of phonological segments within words (e.g., phonemes). Tasks that require compartmentalizing words by deleting, combining, or replacing sounds in words are metalinguistic (Carroll et al., 2003).

Another pragmatic model, put forth by Anthony, Lonigan, Driscoll, Phillips, and Burgess (2003), more directly specifies a continuum of skill development from preschool to fluent reading that includes word awareness, syllable awareness, onset-rime awareness, and phoneme awareness. Each skill set includes specific tasks related to phonological and phonemic awareness. For example, onset-rime awareness includes tasks such as alliteration (Phillips, ClancyMenchetti, & Lonigan, 2008). The continuum of skills moves from largest to smallest units as well as from least to most complex. Together these theories and related skills demonstrate that phonological awareness is a dynamic domain, with relevant assessments capturing a variety of contributing skills.

Response to Intervention

To appropriately target the needs of all children's phonological awareness skills, assessment and intervention practices must be tailored to provide a match between a student's skill level and instructional content (Fuchs, Fuchs, & Compton, 2012). The Response to Intervention (RTI) model is uniquely

suited to address these varied needs by implementing a three-tiered system of assessment and intervention.

RTI is a framework to identify, monitor, and intervene with students based on individualized student academic need (Fuchs & Fuchs, 2006; 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 (Fuchs & Fuchs, 2006; Greenwood, Carta, McConnell, Goldstein, & Kaminski, 2009). First, measures must be able to *identify individual students* who might require a more intensive level of intervention. 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. Both identification and progress monitoring 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 achieve standard deviations below 50% of the mean to ensure the scale produces scores in the highest and lowest regions of the scale, representing lower and higher levels of performance. Additionally, 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. From a qualitative perspective, measures must also adhere to General Outcome Measurement (GOM) tenets (McConnell & Wackerle-Hollman, 2013). GOM provides a unique approach for developing measures in that it captures both empirical and functional standards by providing hallmarks of measurement creation, including being brief (between 1 and 2 minutes per task), being easy to administer and easy to interpret, being related to long term goals, having longevity (can be used for at least one academic year), having reliability, having validity, being inexpensive (or easily attainable), and being sensitive to growth over time, as

well as demonstrating utility as a progress monitoring measure (Fuchs & Deno, 1991; McConnell & Wackerle-Hollman, 2013).

Finally, measurement pragmatic standards specific to RTI must also be considered. First, teachers and administrators may find the most utility in measures that are brief and provide meaningful data to inform relevant and efficient instructional changes. Second, although RTI models can be successful with a normative or a criterion reference standard approach, we suggest a criterion reference standard of performance may be more robust (i.e., benchmarks, or attainment or lack of attainment of specific skills, and/or odds ratios that indicate likely mastery of future standards) rather than comparison to a normative peer group, which may limit the assessor's ability to evaluate a student's absolute skill level. Third, measures must include a psychometrically robust scale of items to represent the construct of interest, rather than a small number of items that may not provide enough information about areas of skill deficit to inform instruction and intervention.

Currently, a number of standardized assessments of phonological awareness for preschool-age children exist, including the Pre-Reading Inventory of Phonological Awareness (Dodd, Crosbie, McIntosh, Teitzel, & Ozanne, 2003) and the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007). Although these types of measures might adequately support screening decisions in an RTI model, as a stand-alone measure they do not demonstrate utility in meeting the specified needs of RTI.

One set of measures that demonstrate potential within an RTI framework are the early literacy Individual Growth and Development Indicators (IGDIs). IGDIs are a set of brief tasks that evaluate early literacy performance in preschool-age children. With a widespread distribution of users across the nation, supported by a long history of research support, the original set of these measures, IGDIs 1.0, have utility in assessment, screening/identification, evaluation, and intervention studies (Greenwood et al., 2008; McConnell & Missall, 2008). IGDIs 1.0 were created using the tenets of GOM as a guiding framework. The IGDIs 1.0 feature two phonological awareness measures, Rhyming and Alliteration. Existing data for Rhyming and Alliteration indicate they measure phonological awareness, with moderate levels of convergent and discriminant validity including the Picture Vocabulary Test-3 (Dunn & Dunn, 2007; $r = .40$ to $.62$), Concepts About Print (Clay, 1985; $r = .34$ to $.64$), and the Test of Phonological Awareness (Torgesen & Bryant, 2004; $r = .44$ to $.79$), demonstrated in empirical contributions from McConnell, McEvoy, and Priest (2002), Missall (2004), and Priest, Silbergliitt, Hall, and Estrem (2000) (Early Childhood Research Institute on Measuring Growth and Development [ECRI-MGD], 1998). Similarly, moderate to high test-retest coefficients were obtained for both tasks (.83 to .89 for Rhyming and .46 to .80 for Alliteration; ECRI-MGD, 1998).

These research findings suggest the measures demonstrate some degree of utility and psychometric adequacy; however, they have significant shortcomings (McConnell & Missall, 2008).

The development of the IGDIs 1.0 was not based on classical test theory. Given this, the data accompanying performance are based on sample-dependent observed scores and typically are used to track normative development, limiting assessors from evaluating student performance based on absolute skill level.

Additionally, because Rhyming and Alliteration are measures that require students to respond to onset rime and/or syllable-level units (fitting within Goswami's [1990] first phase of phonological awareness with a moderate level of word complexity), the tasks may be too difficult for many preschool-age students. Studies demonstrate that students who are 3 to 4 years of age often receive scores of zero, suggesting the Rhyming and Alliteration IGDIs 1.0 have limited utility with young preschool children (Roseth, Missall, & McConnell, in press; Wackerle-Hollman, 2009). Finally, the standardized administration instructions for IGDIs 1.0 specify that the entire set of items is randomly shuffled prior to administration, and information is not consistently gathered about item-level performance. Information is also not gathered about item difficulty or discrimination, preventing assessors, test developers, and users to make more fine-grained evaluations of child performance.

To address these challenges, the measures developed here followed an iterative research and development process using item response theory (Gorin & Embretson, 2008; Albano, Rodriguez, McConnell, Bradfield, & Wackerle-Hollman, 2011) generally and Wilson's (2005) measurement construction framework specifically to respond to the first function of assessment within an RTI model: identifying students in need of additional intervention

Wilson's framework was used to conceptually support and design items, define item characteristics such as responses of interest and parameters for scoring, and statistically model performance (for a detailed description of Wilson's model, see Wilson, 2005). Wilson's model employs four processes: construct mapping, defining the item response, defining the outcome space, and selecting a measurement model. Prior to Study 1 we developed a construct map to provide a strong guide for interpretation. The construct map included an operational definition of *phonological awareness*, as previous described, through comprehensive literature review and expert contributions. We then used the construct map as a foundation for item design. Items were constructed as manifestations of the construct to capture student performance in each domain and revised through an iterative process across the three studies presented here. These three studies then gathered student performance data that have provided important information about how items at different levels of child performance

function. The outcome space is where we applied rules for scoring responses and evaluated features of responses to items we constructed. The outcome space facilitated the identification of student responses corresponding to a particular level on the construct and to make meaning of student performance. Finally, following Wilson's suggestions, work on IGDIs 2.0 employed Rasch modeling and Item Response Theory (Albano et al., 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.

We based this work on Rasch modeling for two reasons. First, pragmatic constraints prevented the use of other IRT models because of the nature and limitation of the sample. In addition, Rasch models provide one-to-one correspondence between Rasch scaled scores and raw scores, to facilitate the ease of number-correct scoring, a hallmark of ease of use in early childhood assessment. In addition, research demonstrates that one-parameter and two-parameter item difficulties are generally correlated at very high levels, typically .98 (de Ayala, 2009, p. 152). Second, the Rasch model provides a strong framework for instrument design, in that the model facilitates evaluation of item functioning by allowing for the creation and identification of items that fit the model rather than identifying a model to fit the data. By employing the Rasch model to help construct the IGDI 2.0 measures, a conceptually robust foundation for producing appropriate and meaningful scores was utilized, rather than identifying a measurement model to explain variation and perhaps inadequacies in data due to less objectively constructed measures.

This article presents the development and related processes used throughout three phases of iterative development to define, pilot, and validate newly revised IGDI 2.0 measures of phonological awareness. This iterative process was designed to maximize efficiency in the development of useful measures for the identification function of RTI assessment; as a result, early phases included smaller samples and broad selection of measures and procedures, with each successive phase deepening methodological rigor and narrowing analytic procedures. These measures, utilizing the strengths of GOM and psychometric advances related to Wilson's (2005) model and Rasch modeling, may provide a foundation for a robust and seamless measurement model for use in an early childhood RTI model, but represent the first three steps in a continuous process of revision and validation.

Three studies were conducted under the auspices of the Center for Response to Intervention in Early Childhood (CRTIEC) as part of a larger effort to expand IGDIs to capture all domains of early literacy and to revise existing measures as appropriate. Study 1 describes the primary

development and initial measure design, selection, and piloting, including examining psychometric and practical properties of several potential measures of phonological awareness; this first study represents an empirical "test of concept" for possible identification measures. Study 2 describes the revision and validation of the most promising measures and testing with a larger, more diverse sample, providing a more robust test of psychometric characteristics of full prototypes of RTI identification item pools. Finally, Study 3 describes the final iteration of revised items designed to uniquely examine performance to discern Tier 1 and Tier 2 or Tier 3 intervention candidates.

Study 1: Developing and Piloting New Measures

The purpose of Study 1 was to review procedures for initial measure design and selection, to examine the psychometric and practical properties of newly developed IGDIs 2.0, and to select individual measurement formats for further research and development. Specifically, this study sought to answer the following questions: (a) To what extent do the selected measures relate to one another? (b) To what extent do individual measures relate to standardized measures of phonological awareness? And (c) how do the IGDI 2.0 measures perform at the item level?

Literature Review

A comprehensive literature review was completed in order to determine a consistent definition of "phonological awareness" (see McConnell et al., in press, for more information on this review). A keyword search using *phonological awareness* and *phonemic awareness* within Education Full Text and Psych Info yielded nine peer-reviewed articles published after 2006 that featured the conceptual development of either phonological awareness or phonemic awareness and targeted preschool-age children. The variety of definitions within the articles yielded common elements, including an understanding that words are made up of individual sounds and the ability to recognize and manipulate sounds. Synthesis of the articles revealed phonological awareness may be best defined as "the ability to detect and manipulate the sound's structure of words independent from their meanings" (Phillips et al., 2008, p. 3).

Method

Measure Design: Phonological Awareness

To capture phonological awareness skills, four tasks were created or revised (see Table 1 for a description of each measure and a summary of quantitative and qualitative pre-pilot responses to the tasks). In addition to two previous

Table 1. Phonological Measure Development and Qualitative Results.

Measure Description	Type	M (Range)	Qualitative Results Summary (n = 10)	Selected for Study 1
Syllable Segmenting required the child to correctly clap the syllable pattern of two-, three-, or four-syllable words, presented verbally (e.g., “elephant” should elicit three consecutive claps).	1, 3	8.25 (0 to 18)	Syllable Segmentation received a dichotomous response pattern where responses were either consistently two claps regardless of the prompt, or nearing 100% accuracy. Syllable Segmentation took 5 to 6 minutes to administer and score.	No
Rhyming required the child to identify a word that rhymes with a target word, given three choices illustrated pictorially.	2, 4	N/A	No results were obtained in the prepilot because the format of Rhyming did not substantially change from IGDIs 1.0.	Yes
Alliteration required the child to identify a word that starts with the same sound as the target word, given three choices illustrated pictorially. Each item included an example alliterative, emphasizing a name or adjective that starts with the same initial sound as the target (e.g., Dan the Dog). Whenever possible, the target pictures name or adjective were monosyllabic.	2, 4	5.82 (4 to 9)	Alliteration response patterns suggested students enjoyed and could complete the task. Assessors reported Alliteration was highly engaging and easy to administer and score. Alliteration took 6 to 7 minutes to administer and score.	Yes
Sound Blending required the child to produce a word given a prompt including a word, syllable, or phoneme blend (e.g., the syllables /wa/ /ter/ should elicit the word water).	2, 3	4.75 (0 to 14)	Sound Blending received a dichotomous response pattern; students seemed to understand the task (i.e., nearing 100% accuracy) or clearly did not grasp the concept at all and could not blend (i.e., repeating the prompt verbatim). Assessors reported students struggled more with phoneme-level items compared to syllable- and word-level items. Sound Blending took 5 to 6 minutes to administer and score.	Yes

Note. IGDI = Individual Growth and Development Indicator. Each task was timed for 2 minutes and a score was given as the number correct. 1 = Manipulation; 2 = detection; 3 = production; 4 = multiple-choice.

versions of Rhyming 1.0 and Alliteration 1.0, four new tasks were tested: Rhyming 2.0, Alliteration 2.0, Syllable Sameness, and Sound Blending.

Participants and setting. A total of 47 children enrolled in three childcare centers in the Upper Midwest participated in this investigation. An economically diverse sample was recruited from two centers located in suburban areas serving a predominately Caucasian population and one located in an urban area serving a predominately Asian American population. Children ranged in age from 36 to 71 months. Fourteen of the children were 3 years old (36 to 47 months), 19 children were 4 years old (48 to 59 months), and 14 children were 5 years old (60 to 71 months). Twenty-one (44.6%) were female and 26 (55.4%) were male.

Measures

Alliteration 1.0. Alliteration 1.0 was drawn from existing IGDIs (ECRI-MGD, 1998) and is an individually administered assessment in which the child identifies from three

alternatives a word that starts with the same sound as a provided target. Participants were presented with an 8.5 × 5.5 inch card with four pictures arranged in one row and three alternatives below the centered target. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read standardized directions to the child: “Point to the picture that starts with the same sound as ____.” The number of cards answered correctly in 2 minutes was recorded as the child’s score.

Rhyming 1.0. Rhyming 1.0 is an existing individually administered IGDI measure (ECRI-MGD, 1998) in which the child identifies from three alternatives a word that rhymes with a provided target word. Participants were presented with an 8.5 × 5.5 inch card with four pictures arranged in one row and three alternatives below the centered target. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read standardized directions to the child: “Point to the picture that rhymes with or sounds the same as ____.”

The number of cards answered correctly in 2 minutes was recorded as the child's score.

Alliteration 2.0. Alliteration 2.0 is a revised version of the original IGDI and is an individually administered assessment in which the child identifies from three alternatives a word that starts with a provided target sound. Participants were presented with an 8.5 × 5.5 inch card with four pictures arranged in one row and three alternatives below the centered target. The target picture also included an adjective or name that started with the same sound (e.g., Dan the Dog) that provided the child with two opportunities to hear the target sound. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read standardized directions to the child: "Point to the picture that starts like ____." The number of cards answered correctly in 2 minutes was recorded as the child's score.

Rhyming 2.0. Rhyming 2.0 is also a revised version of the original IGDI and is an individually administered assessment in which the child identifies from three alternatives a word that rhymes with a provided target word. Participants were presented with an 8.5 × 5.5 inch card with four pictures arranged in one row and three alternatives below the centered target. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read the standardized directions to the child: "Point to the picture that rhymes with ____." The child was shown two example tasks, modeled by the administrator. The number of cards answered correctly in 2 minutes was recorded as the child's score.

Sound Blending. Sound Blending is an individually administered assessment in which the student is prompted to blend word segments at the word, syllable, and phoneme level. Sound Blending was presented verbally using two blocks as manipulatives to assist with demonstrating the tasks. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read the standardized directions to the child: "I'm going to say some words in a funny way. See if you can say them the real way." Children were then presented with words that had been segmented into two parts (e.g., cow-boy), with the administrator tapping one block with her finger for each sound presented. The number of words blended correctly in 2 minutes was recorded as the student's score.

Syllable Segmentation. Syllable Segmentation is a verbal, individually administered assessment in which the child is prompted to clap once for each syllable of simple words. Following an introduction, instruction on how to complete the task, and four practice trials, the administrator read the standardized directions to the child: "When we say words,

we can say their parts using claps. We can say elephant like this: /el/-/e/-/phant/." Administrators clapped one time for each syllable in the word. Children were verbally presented with two-, three-, and four-syllable words. The number of words segmented correctly in 2 minutes was recorded as the child's score.

TOPEL. Participants were given the Phonological Awareness subtest of the TOPEL (Lonigan et al., 2007) as a criterion measure of phonological awareness. The subtest includes elision tasks, which required the participant to remove part of a word (e.g., "say sandbox without sand"), and sound blending tasks, in which the participant must blend two parts of a word together (e.g., "What do these sounds make: /Ba/-/t/?"). Raw scores from the Phonological Awareness subtest were used for the purposes of these analyses, to allow for variance due to age. The Phonological Awareness subtest of the TOPEL has a test-retest reliability coefficient of .83. The correlation coefficients for the TOPEL and the elision and blending subtests of the Comprehensive Test of Phonological Processing, were .59 and .65, respectively (Lonigan et al., 2007). Child performance was represented by scale score.

Procedures. All measures were administered one-on-one with each child by trained undergraduate and/or graduate students. Prior to administration of the measures, the undergraduate and graduate students were trained in standardized procedures for each measure in order to ensure consistent administration 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.

All assessment sessions were conducted on-site at each participating childcare center, either in an empty classroom, conference room, or quiet hallway area. All children were administered four IGDI measures and one criterion measure. To compare the functionality of Alliteration 2.0 and Rhyming 2.0 with existing measures, about half of the children ($n = 21$) also received two additional measures: Rhyming 1.0 and Alliteration 1.0. In order to decrease the burden on children's attention, assessments were conducted in two separate sessions, each lasting from 15 to 20 minutes. After each session, the children selected a small toy from a prize box.

Results

Evaluation of Measure Criteria

Descriptive statistics for each Phonological Awareness IGDI and the TOPEL are presented in Table 2. The descriptive statistics listed in Table 2 were consistent with the

Table 2. Study 1 Early Childhood RTI Criteria: Descriptive Statistics.

Measure	N	M	SD	Skew	Kurtosis	% of Zero Scores
Alliteration 1.0	21	3.38	4.48	1.71	3.16	42
Rhyming 1.0	21	6.86	5.94	0.08	-1.58	33
Alliteration 2.0	47	3.38	4.34	1.22	0.47	47
Rhyming 2.0	47	4.68	5.09	0.55	-1.15	45
Sound Blending	47	7.13	9.17	0.78	-0.94	57
Syllable Segmenting	47	10.91	10.92	0.43	-1.23	38
TOPEL PA	47	14.38	6.21	-0.08	-0.50	2

Note. RTI = Response to Intervention; TOPEL PA = Test of Preschool Early Literacy, Phonological Awareness.

majority of the suggested qualitative criteria for GOMs (e.g., easy to use, brevity, longevity). We also evaluated these measures against empirical criteria for measurement within an RTI model (i.e., $SD < 50\%$ of the mean, less than 20% of children with a score of zero, and skew and kurtosis values less than an absolute value of 1). All IGDI measures had SD s that were relatively large compared to the means, all but one of the measures exceeded skew and kurtosis criteria, and all measures obtained zero scores in excess of the 20% standard, with Sound Blending resulting in the highest percentage of children receiving this score (57%). Alliteration 1.0 had the largest skew and kurtosis. Finally, when considering the measurement criteria for use within an RTI model, the IGDI measures were brief enough to provide data-based decision making, evaluate performance based on a criterion-referenced performance standard, and include test construction featuring solely phonological awareness tasks.

Relations Among Measures

Correlations between measures were calculated and are included in Table 3. Intercorrelations of all new measures with the older IGDI 1.0 were small. Intercorrelations with the revised IGDI 2.0 (Alliteration 2.0 and Rhyming 2.0) and new IGDI (Sound Blending and Syllable Segmenting) were all moderate or moderate to high. One exception was the correlation between Rhyming 1.0 and Rhyming 2.0 (.71), which was the highest correlation of the group. The lowest correlation was between Rhyming 1.0 and Alliteration 1.0 (.16). Criterion-related validity correlation coefficients for the TOPEL Phonological Awareness were at or above .27, with Sound Blending at .70.

Item-Level Performance

In addition to descriptive information and correlations, item-level means and item-total correlations for each measure were also examined (Table 4). 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., phonological awareness ability). For Alliteration 2.0 and Syllable Segmenting, 80% or more of the item means fell between .20 and .80. This was not the case for Rhyming 2.0 and Sound Blending, where less than 60% of the item means fell between .20 and .80. This indicates that overall the items in Rhyming 2.0 and Sound Blending were too difficult for this sample of children as compared to the items in the Alliteration 2.0 and Syllable Segmenting. Rhyming 2.0 and Sound Blending had fewer items that positively discriminate between children who did and did not have the skills assessed by these measures, as compared to Alliteration 2.0 and Syllable Segmenting.

Discussion

This study involved a small-scale examination of the IGDI 2.0 phonological awareness measures, conducted to capture preliminary information such as student response rate, zero responses, and basic descriptive statistics, and to determine which IGDI 2.0 measures were the best candidates for further development and large-scale field testing in Study 2. During item development for pilot testing, GOM features were maintained as much as possible, and as a result the item sets remained timed tasks (1 or 2 minutes). Evaluation of each measure included a comparison of descriptive statistics to predefined GOM and measure criteria, examination of correlations both between measures within the domain and with standardized criterion measures (e.g., TOPEL) to evaluate validity.

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

Essential GOM Criteria

GOM criteria remain an important tenet of IGDI 1.0 and 2.0 measures because they align the measurement tools with real-world academic goals and provide the end user with tools that are both socially and psychometrically valid, but also are engaging and brief to administer. Of the GOM characteristics described previously, all six evaluated phonological awareness assessments are quick (2 minutes each) and easy to administer and interpret. Assessors reported Rhyming 2.0 and Alliteration 2.0 were generally easy to administer and interpret. Assessors reported challenges with Syllable Segmenting because the nature of the assessment elicited a dichotomous response set from children. Either children were nearly always accurate and clearly understood the task or children demonstrated a lack of understanding entirely, as illustrated by continually clapping, instead of clapping along

Table 3. Study I Correlation Between Measures.

Measure	Alliteration 1.0	Rhyming 1.0	Alliteration 2.0	Rhyming 2.0	Sound Blending	Syllable Segment
Rhyming 1.0	.16*	—				
Alliteration 2.0	.43	.61**	—			
Rhyming 2.0	.37	.72**	.51**	—		
Sound Blending	.39	.22	.42**	.55*	—	
Syllable Segment	.45*	.29	.52**	.55**	.59**	—
TOPEL PA	.56**	.27	.63**	.42*	.70**	.53**

Note. TOPEL PA = Test of Preschool Early Literacy, Phonological Awareness.

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

Table 4. Study I Mean Number of Responses per Item, Range of Item Means, Means, and Item-Total Correlation Ranges by Measure.

Measure	Mean Number of Responses per Item	Item Means		Item-Total Correlations	
		Range	% Between .20 and .80	Range	%.20 or Above
Alliteration 2.0	5.79	.25 to 1.00	80	-.39 to .94	79
Rhyming 2.0	6.03	.33 to 1.00	56	-.53 to .97	61
Sound Blending	8.43	.40 to 1.00	37	-.28 to .70	43
Syllable Segmenting	20.75	.28 to .83	100	.17 to .84	98

with the syllables of the word and very low or zero scores. Sound Blending demonstrated additional challenges because of the nature of the task (manipulation of cubes), pronunciation of separated words, and related pacing of stimuli.

In addition, results suggested nearly all of the measures met few of the empirical Early Childhood (EC) RTI criteria. Alliteration 1.0 met none of the criteria; Rhyming 1.0, Rhyming 2.0, and Syllable Segmenting only met the skew criterion; Alliteration 2.0 met only the criterion for kurtosis; and Sound Blending met the criteria for both skewness and kurtosis. It should be noted that none of the measures met the criterion for a standard deviation less than 50% of the sample mean. Because mean performance on the IGDI phonological awareness measures was low relative to the standard deviations, performance at the lower end of the distributions could not be appropriately captured, as illustrated by a significant proportion of zero scores and visual analysis of sample distributions suggesting items were too difficult for this sample. Together, these three criteria (skew, kurtosis, and SD/M ratio) describe the shape of the distribution. Taken together, these findings suggest that the IGDI 2.0 measures are superior to the IGDI 1.0 measures; however, in general, the phonological awareness measures performed poorly among statistical EC RTI criteria, indicating the need for improvement in the measures to accurately capture child performance. In particular, post hoc analyses suggested that items located higher on the ability scale than did children and that future instrument development would require more items at the lower or earlier level of ability.

Validity Evidence

Validity was examined by evaluating the relation between the IGDI measures within the phonological awareness domain. Intermeasure correlations ranged from weak (Rhyming 1.0 and Alliteration 1.0) to strong (Rhyming 1.0 and Rhyming 2.0). The dramatic variability in internal criterion-related validity correlations suggest some measures may be poor representations of the phonological awareness domain (Alliteration 1.0), while others may be adequate to strong representations (Rhyming 2.0). These findings further support the notion that the IGDI 2.0 measures outperformed the IGDI 1.0 measures; however, all of the current measures of phonological awareness skills had significant room for improvement.

The relation between IGDI performance and performance on the TOPEL was also examined to evaluate external criterion-related validity evidence. With the exception of Alliteration 1.0, all measures demonstrated significant correlations with the TOPEL, suggesting the IGDI measures may appropriately access the phonological awareness domain.

Item-Level Functioning

The p values, the proportion of children passing an item, ranged between .25 and 1.00. A p value within the range of .20 to .80 was considered acceptable. The p values outside this range indicate the item did not contribute to the test in a meaningful way, as a result of either being too difficult or

too easy. Similarly, item-total correlations can also be used to aid in determining if items contribute to a test in meaningful ways. Item-total correlations with values that were greater than .20 were determined to be discriminating well and items at or above .20 were retained in the potential item pool. With the exception of Sound Blending, the measures had over 60% of items with values at or above .20.

The phonological awareness measures demonstrate both strengths and weaknesses across the three contributing pieces of empirical and practical evidence. Three candidate measures—Rhyming 2.0, Alliteration 2.0, and Sound Blending—were selected for further refinement and iterative revisions for field testing in Study 2. The final three measures were chosen based on their superior fit with the GOM criteria, supporting criterion validity evidence, and item-level functioning. Alliteration 1.0 and Rhyming 1.0 were eliminated because their counterpart revisions were statistically improved, and Syllable Segmenting was removed because of poor performance within item level and descriptive evaluations and anecdotal reports noting a dichotomous response pattern.

Study 2

Building on the results of Study 1, Study 2 intended to examine the psychometric properties of newly developed IGDI 2.0 with an expanded set of items and revised procedures to conceptually support a reduced floor effect. Study 2 also drew a larger, more diverse sample of children. Specifically, this study sought to answer the following questions: (a) To what extent do the measures relate to one another? (b) What is the validity of the measures?

For Study 2, the Wilson (2005) “constructing measures” framework was implemented fully, employing the Rasch measurement model for analyses of child and item performance. The Rasch model places items on a scale based on item difficulty, locating the average item at zero (typically resulting in an ability scale from -4 to 4). Based on their performance on the IGDI items, children are assigned Rasch scores that 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

A total of 756 children participated in assessments in the fall, winter, and spring of the 2009–2010 academic year. Of the 756 participants, 633 children received scores above zero on the Rhyming 2.0 and Alliteration 2.0 measures and were included in the analyses presented here. Children in the larger study were enrolled in 65 classrooms in childcare

centers from four states in the East, Midwest, and Pacific Northwest. Early care and educational setting classrooms were targeted for recruitment. Children between 4 and 5 years of age (48 and 71 months) were eligible for recruitment. Parental consent forms were sent home with all eligible children. The mean age of children was 54 months. Exactly half of the children were male ($n = 378$) and half were female ($n = 378$). The distribution of race/ethnicity was as follows: 36% White, 30% African American, 20% Hispanic, 10% multirace, 2% Asian, 1.5% Other, and 0.4% Native American. Eighty-four percent of parents reported speaking to their child at home in English and 21% in Spanish.

Measures

Using information collected from Study 1, measures were selected for use in Study 2 based on their overall fit with the GOM characteristics, criterion validity correlation coefficients, item-level information, and the professional judgment of the research team regarding the feasibility of each measure. For those measures that were considered for Study 2, poorly functioning items were discarded or edited to remove for construct-irrelevant features (Albano et al., 2011). Construct-irrelevant features are elements of an item that influence child response but do not relate to the domain. Construct-irrelevant features include aspects within items such as unnecessary backgrounds, unnecessary borders around items, or differences in image type (e.g., illustration vs. photograph). Additional items were developed for each measure, yielding a total item pool of 44 items per measure.

The measures considered for application in Study 2 included Alliteration 2.0, Rhyming 2.0, and Sound Blending. Administration procedures for each task were not revised. As a result, assessors were provided with the same manual as in Study 1. The Phonological Awareness subtest of the TOPEL was administered as the phonological awareness criterion measure (Lonigan et al., 2007).

Procedures

During Study 2, participants were administered measures during three waves of data collection (fall, winter, and spring) throughout the academic year as part of a larger study (Greenwood et al., 2011). During each wave, participants were administered nine IGDI measures from the Oral Language, Alphabet Knowledge, and Phonological Awareness domains of early literacy. In Waves 1 and 3, each participant received one of three criterion measures being used in three different IGDI validation efforts. Administration of the criterion measures was spiraled across participants so that one third of the participants received a criterion from each early literacy domain. Measures were administered across two or three sessions,

each lasting 15 to 20 minutes. All measures were administered by trained graduate or undergraduate students. Assessment sessions were conducted onsite at each center, either in an empty classroom, conference room, or quiet hallway.

In order to collect sufficient item-level data to meet Rasch model requirements, data collection was structured such that each item would be administered to at least 100 children. Due to the large number of items per measure, a bundling procedure was created to ensure 100 responses per item. Each bundle had four sample cards and five common cards (i.e., cards that remained constant across bundles) and 15 timed administration items, for a total of 24 items in each bundle. The five common cards were selected to represent the full range of ability and were used to anchor items across multiple bundles on the same scale. Because assessments were timed, bundles were designed such that some items overlapped across bundles to account for variation in child performance within the given time frame (e.g., 1 to 2 minutes). In this way, items that did not receive responses (because the student was unable to receive the item due to time) were not counted as incorrect; instead, they were simply excluded from analysis. Because of the overlap in bundles and the assessment scheme, all items achieved at least 100 responses.

For the purposes of the analyses, the IGDIs were scored using the Rasch model (Rasch, 1960; Albano et al., 2011). Once cases with raw scores of zero were removed—since they provide no information about child ability or item function—and Rasch scores calculated, we computed descriptive data for Waves 1 and 3.

Results

Characteristics of Measures

More than half of all children received a raw score of zero on Sound Blending; therefore, this measure was dropped from further analyses. Descriptive statistics for Alliteration 2.0 and Rhyming 2.0 IGDIs and TOPEL are presented in Table 5. Descriptive results for the TOPEL Phonological Awareness subtest were also computed. Overall, participants' scores on Rhyming 2.0 tended to vary more than scores on Alliteration 2.0.

Relations among measures. Correlations between measures are included in Table 6. Correlations between the IGDI measures and the TOPEL Phonological Awareness subtest were moderate.

Discussion

This study presented a large-scale field test of IGDI 2.0 phonological awareness measures, conducted to capture

Table 5. Study 2 Mean Rasch Scores, Standard Deviations, Skew, and Kurtosis by Measure and Wave.

Measure	<i>n</i>	<i>M</i>	<i>SD</i>	Skew	Kurtosis
Wave 1					
Alliteration 2.0	740	−0.79	1.51	0.33	−0.01
Rhyming 2.0	802	−0.47	1.75	0.52	−0.32
TOPEL	199	12.9	5.63		
Wave 3					
Alliteration 2.0	633	0.17	1.22	1.35	2.06
Rhyming 2.0	653	0.82	1.59	0.50	−0.57
TOPEL	198	16.0	5.80		

Note. Test of Preschool Early Literacy (TOPEL) scores are represented as raw scores.

Table 6. Study 2 Correlations Between Measures.

	Alliteration 2.0	Rhyming 2.0
Rhyming 2.0	.51**	—
TOPEL PA	.52**	.45**

Note. TOPEL PA = Test of Preschool Early Literacy, Phonological Awareness.

***p* < .01.

validity evidence to support further development and application of the IGDI 2.0 measures, with implications for use within an RTI model. Descriptive statistics and criterion-related validity coefficients between measures and with the TOPEL were evaluated to determine the feasibility, utility, and validity of the measures. This study represented a diverse sample of students, representing four geographic regions across the continental United States. Students included typically developing and special education students, as well as English Language Learners (ELLs) and students enrolled in programs primarily serving low-income families (e.g., Head Start). By evaluating student performance within the larger sample, greater confidence can be vested in the descriptive properties of the IGDI 2.0 measures, offering data to support application with differing populations. Similarly, by evaluating the relation between and among measures and the TOPEL information about the utility of the phonological awareness, IGDI measures as appropriate measures of the construct of phonological awareness can be evaluated, thus answering the research questions.

Descriptive Analysis and Item-Level Performance

For Rhyming 2.0 and Alliteration 2.0 mean scores suggest students' performance was below the ability required for the average item (located at 0) at Wave 1 and above the ability required for the average item at Wave 3. While mean performance of IGDI 2.0 measures is not comparable between

studies, Study 2 indicates items for each measure demonstrate utility and have implications for use within an RTI model in that the items represent student abilities that are appropriate for preschool-age children, however measures still contain too few items that may have greater utility for low-performing preschool students as indicated by the percentage of zero scores received (10% for Rhyming 2.0 and 9% for Alliteration 2.0, Wave 3). Although only Wave 3 scores are reported here, it should be noted that the percentage of zero scores decreased over time, with the largest percentage obtained at Wave 1 followed by Wave 2.

Validity Evidence

Validity was examined by evaluating the relation between IGDIs within the phonological awareness domain. Intermesure correlations between Rhyming 2.0 and Alliteration 2.0 suggest a moderate relation between measures. This correlation potentially illustrates common contributions to the phonological awareness domain, but also unique contributions of each measure. Compared to Study 1 correlations, Rhyming 2.0 and Alliteration 2.0 remained the same (.51), however sample sizes differed dramatically, from 47 in Study 1 to 653 in Study 2.

Relations between performance on IGDIs and TOPEL were also examined to evaluate external criterion-related validity evidence. Correlations between the TOPEL and Alliteration 2.0 and Rhyming 2.0 suggest moderate and generally equivalent relations between the established criterion test and the current IGDI measures.

Taken together, Study 2 findings suggest Rhyming 2.0 and Alliteration 2.0 perform adequately with preschool-age students who demonstrate higher levels of phonological awareness ability; however, for students with lower levels of phonological awareness ability, who as a result may be at risk for later reading difficulties, the phonological awareness IGDIs had less utility. Therefore, item sets included in Rhyming 2.0 and Alliteration 2.0 have improved but are in need of further item-level revisions and development of additional items for optimal use within an RTI paradigm. As such, a second level of revisions, including improving the Rhyming 2.0 and Alliteration 2.0 tasks to a two-choice selection (rather than three), further examining potential construct irrelevant features, and providing simplified instructions for students, was considered and was examined in a revision study during the 2010–2011 academic year (Study 3).

Study 3

Based on the results of Studies 1 and 2, two IGDI 2.0 measures, Rhyming 2.0 and Alliteration 2.0, demonstrated improved effects for high-achieving students, but warranted revisions to demonstrate utility with low-achieving students, for appropriate use in an RTI model. The authors determined

a third study would be appropriate to evaluate if the issues identified in Study 2 could be remedied. In this study, the same measurement model was utilized (i.e., Rasch); however, to more authentically employ the Rasch model, the timing of measures was removed and the procedures and item-level features were modified with the intention of appropriately capturing performance of low-ability students. Study 3 featured two research questions: (a) To what extent do the newly revised Rhyming 2.0 and Alliteration 2.0 measures show improved concurrent criterion validity than those established in Study 2? And (b) to what degree are the item locations representative of student ability level for Rhyming 2.0 and Alliteration 2.0 on the Rasch scale? That is, are the items more likely to represent low ability levels and reduce ceiling effects?

Method

Participants and Setting

A total of 278 children participated in two seasonal assessments: winter and spring of the 2010–2011 academic year. Four- and 5-year-old children (48 and 71 months) were recruited from early care and educational setting classrooms. Parental consent forms were sent home with all eligible participants, yielding a consented sample of 151 males (55%) and 127 females (45%). The distribution of race/ethnicity was as follows: 36% White, 30% African American, 5% Hispanic 2%, Asian, and 1% Other. Of the 241 students who reported disability status and ELL status, 38 (16%) had an Individualized Education Program (IEP) and 17 (7%) were considered ELLs.

Measures

As suggested in Study 2, IGDI Rhyming 2.0 and Alliteration 2.0 measures were selected for use in Study 3. During Study 3 a series of revisions were made to each of the measures to improve item-level functioning. The procedures described in Study 2 to remove construct-irrelevant features were again employed. In addition, for each task we reduced the number of choice responses available within each item from three to two to further reduce the cognitive load, such that children would be required to remember less information before making a choice response. Additional items were written explicitly to sample lower ability content for each measure. After revisions and new item construction, a total item pool of 60 items per measure was developed.

Administration procedures were also revised to reduce the cognitive load of each task by providing scaffolding during administration such that the administrator paired each target and response choice together for the child (e.g., “Toy, boy, mask. Which two rhyme? Is it toy, boy (insert pause) or toy, mask?” for Rhyming and “Tree, duck. Which one starts with /d/?” for Alliteration). In addition, the timing

Table 7. Study 3 Mean Raw Scores, Mean Rasch Score (Ability), Standard Deviations, Minimum, Maximum, Skew, and Kurtosis by Measure by Wave.

Measure	Wave	<i>n</i>	<i>M</i> (Rasch)	<i>M</i> (Raw)	<i>SD</i>	Min.	Max.	Skew	Kurtosis
Alliteration 2.0	Winter	276	2.31	48.00	11.45	24	60	-0.46	-1.29
	Spring	82	3.08	49.95	12.08	25	60	-0.73	-1.19
Rhyming 2.0	Winter	271	0.9	45.15	12.17	21	60	-0.22	-1.55
	Spring	115	1.91	47.64	11.66	24	60	-0.51	-1.34
TOPEL PA	Spring	59	N/A	14.54	6.22	4	27	0.26	-0.97

Note. All statistics are provided for raw scores with the exception of the *M* (Rasch). TOPEL PA = Test of Preschool Early Literacy, Phonological Awareness.

of the IGDI measures was removed, and the test was redesigned to be a fixed length interaction to more authentically employ the assumptions of the Rasch model.

Assessors were trained on the revised procedures and obtained 90% fidelity of implementation prior to data collection efforts. Similar to Study 2, the Phonological Awareness subtest of the TOPEL was administered as the criterion measure (Lonigan et al., 2007).

Procedures

During Study 3, participants were administered measures during two waves of data collection (winter and spring) in 2011. During each wave, participants were administered six IGDI 2.0 measures from the Oral Language, Alphabet Knowledge, Phonological Awareness, and Comprehension domains of early literacy. Sixty participants were randomly selected for standardized criterion assessments during the second wave, with 57 standardized assessments (i.e., TOPEL) completed (three students were absent on the day of assessment). Measures were administered across three sessions, each lasting 10 to 15 minutes, such that each student saw a total of 60 items per Rhyming 2.0 and Alliteration 2.0 measure. All measures were administered by trained graduate or undergraduate students. Assessment sessions were conducted onsite at each center, either in an empty classroom, conference room, or quiet hallway.

As noted in Study 2, Rasch modeling was used to evaluate each measure. First, to better understand the structure of the measures of phonological awareness and to provide evidence of unidimensionality to support the use of the Rasch model, we conducted two forms of confirmatory factor analysis (CFA). The first tests the fit of the data to a unidimensional model for Alliteration 2.0 and Rhyming 2.0 independently, and the second tests a two-factor model allowing the factors of Alliteration 2.0 and Rhyming 2.0 to correlate. For each model, two fit indices are reported, including the comparative fit index (CFI), where good fit is found with values greater than .95, and the root mean squared error of approximation (RMSEA), where good fit is found with values less than .08 (Brown, 2006). The first independent unidimensional models fit very well. For Alliteration 2.0, the

CFI was .986 and RMSEA was .026. For Rhyming 2.0, the CFI was .922 and RMSEA was .072. The combined model allowing the two measures of Phonological Awareness to correlate yielded a CFI of .961 and RMSEA of .038, with a correlation between the factor scores of Alliteration 2.0 and Rhyming 2.0 (removing measurement error) of .75 (56% common variance between the constructs of Alliteration 2.0 and Rhyming 2.0). The CFA results indicate adequate to excellent fit.

Second, Rasch assumptions, including local independence—or that the response on one item does not depend on a response to other items—item fit, and item discrimination, were tested within the model. Results indicate that assumptions were confirmed with empirically robust item-level statistics (infit and outfit less than a value of 2; discrimination was uniformly moderate to high).

Finally, Rasch modeling required 100 responses per item. As such, this study sampled items across students to achieve 100 responses for each of the 60 items in each measure. Items were then scored using descriptive methods and the Rasch model (Albano et al, 2011; Rasch, 1960).

Results

Characteristics of Measures

During Study 3, no child assessed received a score of zero on both Rhyming 2.0 and Alliteration 2.0. Descriptive statistics for Alliteration 2.0 and Rhyming 2.0 IGDI and TOPEL are presented in Table 7. Descriptive results for the TOPEL Phonological Awareness subtest were also computed.

Relations among measures. Correlations between measures are included in Table 8. Correlations between the IGDI measures and the TOPEL PA subtest were moderate, with IGDI correlations with TOPEL .50 to .61 (compared to $r = .45$ to $.52$ in Study 2).

Discussion

This study represented the third step in an iterative development process to field test two IGDI 2.0 Phonological

Table 8. Study 3 Correlations Between Measures.

	Alliteration 2.0	Rhyming 2.0
Rhyming 2.0	.67**	—
TOPEL PA	.61**	.50**

Note. $n = 57$. TOPEL PA = Test of Preschool Early Literacy, Phonological Awareness.

** $p < .01$.

Awareness measures: Rhyming 2.0 and Alliteration 2.0. This study was conducted to capture validity evidence to support the use of IGDI measures within an RTI model to identify students who may be in need of additional instructional support or intervention at the Tier 2 or Tier 3 level. Descriptive statistics and criterion-related validity coefficients between measures and with the TOPEL Phonological Awareness were examined to evaluate the utility and validity of the measures.

Descriptive Analysis and Item-Level Performance

For Rhyming 2.0 and Alliteration 2.0, mean Rasch scores suggest average student ability was above the ability required for the average item (located at 0). Although mean performance of IGDI 2.0 measures is not comparable between studies because the content of the items was revised, Study 3 indicates no floor effects were present in this sample, with a minimum raw score of 24 and 21 on the IGDI 2.0 measures. As a result, compared to Study 2, the revised Rhyming 2.0 and Alliteration 2.0 measures demonstrate utility and have implications for use within an RTI model in that the items represent student abilities that are appropriate for preschool-age children. More specifically, the items capture ability levels of students who may be appropriate candidates for Tier 2 or Tier 3 intervention, as illustrated by the lack of zero scores. These item analyses suggest the IGDI 2.0 Phonological Awareness measures may be appropriately used for identifying students who are candidates for Tier 2 and Tier 3 intervention, such that their level of performance can be accurately identified in the Rasch model.

Validity Evidence

In comparison to Study 2, the revised Rhyming 2.0 and Alliteration 2.0 measures demonstrate improved concurrent and criterion correlations, improving from 0.51 ($n = 633$) between Rhyming and Alliteration in Study 2 to 0.67 in Study 3 ($n = 57$). Similarly, relations between the standardized measure (TOPEL Phonological Awareness) and the revised IGDI measures were also improved, from 0.52 for Alliteration and 0.45 for Rhyming to 0.61 and 0.50, respectively, in Study 3.

Results from Study 3 indicate Rhyming 2.0 and Alliteration 2.0 perform adequately with preschool-age students across

ability levels of phonological awareness. Improvements to the measures, reduced cognitive load, and examination of item characteristics to write new items at lower ability levels contributed to an empirically validated scale of items used for seasonal identification of students in need of intervention at a Tier 2 or Tier 3 level.

General Discussion

Early literacy assessment models for RTI represent new directions in early childhood education, moving away from a “wait to fail” approach and toward a responsive and preventative approach to child intervention and assessment (Greenwood et al., 2011). The measures developed here (IGDIs 2.0) have been designed to be uniquely suited for use within an RTI model, positioning item sets to meet the ability levels of preschool-age children for use as the identification of students who may need additional intervention. As a result, IGDIs 2.0 demonstrate promise in an RTI framework and will be further developed through the iterative development process as identification measures—beginning with the process described within this article.

More specifically, IGDIs 2.0 feature phonological awareness tasks that capture a portion of the continuum of skills represented in current theories including syllable awareness (Rhyming 2.0) and onset-rime awareness (Alliteration 2.0; Anthony et al., 2003). Furthermore, consistent with Goswami (1990), the IGDI measures were developed focusing on initial early literacy skills capitalizing on rhyme and alliteration development. By developing tasks that represent the continuum of the phonological awareness construct, the research team intended to appropriately span the ability levels represented in preschool classrooms, further illustrating performance at the RTI tier-level divisions (i.e. Tier 1, Tier 2, Tier 3). It is relevant to note that these phonological awareness measures were not developed in exclusion of alphabet knowledge tasks that closely inform and link to phonological awareness skill development, such as letter sounds and letter names. Instead, complementary IGDI 2.0 measures, including a sound identification measure, were developed by domain with parallel research to support measure development within the domain of alphabet knowledge (see Bradfield, Wackerle-Hollman, & McConnell, 2011).

The data presented here indicate the measures have progressed through three iterative phases of development and have now reached standards for potential use within an RTI model. The measures are able to accurately capture the ability levels of all preschool-age students as demonstrated in Study 3. However, to identify those students in need of Tier 2 or Tier 3 intervention, measures that are sensitive to low ability levels are not enough; relevant cut-score criteria and predictive validity estimates are also needed.

Current work on the IGDI 2.0 measures is focused on creating empirically robust cut-scores for Tier 2 and Tier 3 candidacy or, stated another way, for screening or identification purposes. This identification set of IGDI 2.0 phonological awareness measures can be used to reliably and accurately detect those students with ability levels that may need support at the Tier 2 or Tier 3 level during season screening assessments. With robust measures for identification of students as candidates for intervention in hand, practitioner needs will move toward the next step of assessment in an RTI model: progress monitoring. However, the data presented in these studies are not useful for progress monitoring analysis. The studies presented here do not offer any parameters to evaluate expected growth rates and sensitivity to growth.

It is also important to recognize the IGDI 2.0 identification measures are not without limitations. Given the nature of development during the preschool years, the opportunities to examine “typical” emergence and mastery of early literacy skills are at best brief. It may be the case that the measures presented here have utility for only a brief period. In practice, complementing IGDI 2.0 with other measures or methods of evaluating performance (e.g., master monitoring tasks) may prove useful. Furthermore, because IGDI 2.0 are in their infancy, no data are yet available that examine the predictive validity of the measures. Without this information, end users cannot be confident in their ability to reliably predict academic success in the area of phonological awareness at later grades (kindergarten through third grade).

Nevertheless, even considering these limitations, there are currently no early literacy measures available that cater specifically to the unique needs of an early childhood RTI model and meet the psychometric criteria suggested in Studies 1 through 3. Furthermore, by implementing an iterative refinement and revision process, the IGDI 2.0 measures will ensure appropriate interpretability through reduced measurement error, strong construct representations, and specific item-level information to support task creation. This ongoing process represents an effort to maintain a research-to-practice transition, by ensuring both robust psychometric standards *and* practical utility, resulting in a superior set of early literacy assessment tools.

As the process of refinement and revision continues, including development of criterion-based cut scores and the demonstration of predictive validity, a simultaneous program of early childhood RTI model development is also occurring with interventions at the Tier 2 and Tier 3 level, fidelity of implementation procedures, and considerations for parents, assessors, and facilitators. As both the assessment and complementary RTI intervention and support come to fruition, there are tremendous opportunities for improved assessment and intervention and, as a result, dramatically improved student outcomes.

Acknowledgments

The authors would like to thank colleagues who assisted with this project including participating childcare centers and programs in the Minneapolis/St. Paul area. Additionally, 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. However, 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.

Declaration of Conflicting Interests

The author(s) 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 the 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 of Minnesota in accordance with its conflict of interest policies.

Funding

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

References

- 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., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*(1), 53–55.
- Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*(4), 470–487.
- 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.

- Bradfield, T., Wackerle-Hollman, A., & McConnell, S. (2011, January). *Using IGDIs to identify children for tiered intervention*. Presentation at the Center for Response to Intervention in Early Childhood Summit on Early Childhood RTI, Albuquerque, NM.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Carroll, J. M., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*(5), 913–923.
- Clay, M. M. (1985). *The early detection of reading difficulties*. Portsmouth, NH: Heinemann.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford.
- Dodd, B., Crosbie, S., McIntosh, B., Teitzel, T., & Ozanne, A. (2003). *Pre-Reading Inventory of Phonological Awareness*. San Antonio, TX: Psychological Corporation.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Circle Pines, MN: American Guidance Service.
- 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
- Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD). (1998). *Theoretical foundations of the early childhood research institute on measuring growth and development: An early childhood problem-solving model* (Tech. Rep. No. 6). Minneapolis: University of Minnesota, Center for Early Education and Development.
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488–499.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99.
- Fuchs, L. S., Fuchs, D., & Compton, D. (2012). Smart RTI: A next generation approach to multilevel prevention. *Exceptional Children, 78*(3), 263–279.
- Gombert, J. E. (1992). *Metalinguistic development*. New York: Harvester Wheatsheaf.
- Gorin, S., & Embretson, J. (2008). *Item response theory and Rasch Modeling*. In D. McKay (Ed.), *The handbook of research methods in abnormal and clinical psychology* (pp. 271–292). Thousand Oaks, CA: Sage Publications.
- Goswami, U. (1990). *Phonological skills and learning to read*. London: Lawrence Erlbaum.
- Goswami, U., & East, M. (2000). Rhyme and analogy in beginning reading: Conceptual and methodological issues. *Applied Psycholinguistics, 21*(1), 63–93.
- Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The Response to Intervention (RTI) approach in early childhood. *Focus on Exceptional Children, 43*(9), 1–22.
- Greenwood, C. R., Carta, J. J., McConnell, S. R., Goldstein, H., & Kaminski, R. A. (2009). *Center for Response to Intervention in Early Childhood*. Retrieved March 1, 2011, from <http://www.crtiec.org>
- Greenwood, C. R., Kratochwill, T. R., & Clements, M. (2008). *Schoolwide prevention models: Lessons learned in elementary schools*. New York: Guilford Press.
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy*. Austin, TX: PRO-ED.
- 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., McEvoy, M., & Priest, J. (2002). “Growing” measures for monitoring progress in early childhood education: A research and development process for individual growth and development indicators. *Assessment for Effective Intervention, 27*(4), 3–14.
- McConnell, S. R., & Missall, K. N. (2008). Best practices in monitoring progress in preschool children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., Vol.2, pp. 561–573). Bethesda, MD: National Association of School Psychologists.
- McConnell, S. R., & Wackerle-Hollman, A. (2013). *Can we measure the transition to reading? Relations among general outcome measures of language and early literacy development from preschool to early elementary grades*. Unpublished manuscript, University of Minnesota, Minneapolis.
- McConnell, S. R., Wackerle-Hollman, A., & Bradfield, T. A. (in press). Early childhood literacy screening. In R. Kettler, T. Glover, C. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Identification, implications, and interpretation*. Washington, DC: American Psychological Association.
- Missall, K. (2004). *Relations between general outcome measures of literacy*. Retrieved January 10, 2006, from www.dynamic-measurement.org/presentations/ds04Missall1.pdf
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665–681.
- National Early Literacy Panel. (2008). *Developing early literacy report of the National Early Literacy Panel*. Washington, DC: Author.
- Phillips, B. M., ClancyMenchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children: Lessons from the classroom. *Topics in Early Childhood Special Education, 28*(1), 3–17.
- Priest, J. S., Silbergitt, B., Hall, S., & Estrem, T. L. (2000). *Progress on preschool IGDIs for early literacy*. Paper presented at the Presentation at Heartland Area Education Association, Des Moines, IA.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Roseth, C. J., Missall, K. N., & McConnell, S. R. (in press). Early Literacy Individual Growth and Development Indicators (EL-IGDIs): Growth trajectories using a large, internet-based sample. *Journal of School Psychology*.
- Senechal, M., LeFevre, J., SmithChant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy: The role of empirical evidence. *Journal of School Psychology, 39*(5), 439–460.

- Snow, C. E., Burns, M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Torgesen, J. K., & Bryant, B. R. (2004). *Test of phonological awareness-second edition: PLUS*. Austin, TX: PRO-ED.
- Wackerle-Hollman, A. (2009). The effects of progress monitoring and consultation on emergent literacy performance as measured by the Individual Growth and Development Indicators. *Digital Dissertations*. Retrieved from http://conservancy.umn.edu/bitstream/54208/1/Hollman_umn_0130E_10460.pdf
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, 33(3), 468–479.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69(3), 848–872.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.