
MEASURING YOUNG CHILDREN'S ALPHABET KNOWLEDGE

Development and Validation of Brief Letter-Sound Knowledge Assessments

ABSTRACT

Early childhood teachers are increasingly encouraged to support children's development of letter-sound abilities. Assessment of letter-sound knowledge is key in planning for effective instruction, yet the letter-sound knowledge assessments currently available and suitable for pre-school-age children demonstrate significant limitations. The purpose of the current study was to use item response theory to create short-form letter-sound assessments that are psychometrically sound, quick and easy to administer, and appropriate for researcher and practitioner use. Letter-sound knowledge data from 940 children attending pre-school centers were analyzed. Item response theory analyses indicated that a two-parameter logistic model best fit the data. Further analyses yielded a series of parallel six- and eight-letter forms with adequate test information, reliability, and theta recovery. Implications for assessment and instruction are discussed.

Shayne B. Piasta

THE OHIO STATE
UNIVERSITY

Beth M. Phillips

FLORIDA STATE
UNIVERSITY

Jeffrey M. Williams

UNIVERSITY OF TEXAS
HEALTH SCIENCES
CENTER

Ryan P. Bowles

MICHIGAN STATE
UNIVERSITY

Jason L. Anthony

UNIVERSITY OF TEXAS
HEALTH SCIENCES
CENTER

ACCUMULATED research makes it clearer than ever that critical elements of reading success begin developing well before children arrive at the kindergarten door. Converging evidence indicates, however, that there is wide diversity in the emergent literacy skills and experiences that children bring to kindergarten (Invernizzi, Justice, Landrum, & Booker, 2004; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). To support the continued learning of all children, early childhood educators require emergent literacy assessments that help them understand children's current skill levels and thereby plan appropriate instruction. Such assessments must be psychometrically sound and easily used, scored, and interpreted. The present study focuses on assessing children's acquisition of one particular emergent literacy skill, letter-sound knowledge, by addressing its psychometric characteristics and developing a measure with multiple short forms that allows for easy, efficient, and accurate assessment.

Significance of Letter-Sound Knowledge

Alphabet knowledge is well recognized as perhaps the most robust predictor of future decoding ability among children 3 to 5 years old (Torppa, Poikkeus, Laakso, Eklund, & Lyytinen, 2006). Longitudinal studies predicting from preschool to kindergarten (e.g., Lonigan, Burgess, & Anthony, 2000), preschool to later elementary school (e.g., Puolakanaho et al., 2007), and kindergarten to later elementary school (e.g., Leppänen, Aunola, Niemi, & Nurmi, 2008; Schatschneider et al., 2004) converge in noting that substantial variance in reading achievement is uniquely attributed to early measures of alphabet knowledge. Furthermore, research indicates that alphabet knowledge is a significant predictor of spelling and reading comprehension, over and above the contributions of oral language, phonological awareness, and demographic characteristics (e.g., Parrila, Kirby, McQuarrie, 2004; Torppa et al., 2010).

Alphabet knowledge includes several aspects, including letter-name knowledge and letter-sound knowledge. Both are typically conceptualized in terms of the total number of letter names or sounds known (i.e., sums of 0 to 26) or, more recently (e.g., Bowles, Skibbe, & Justice, 2011; Drouin, Horner, & Sondergeld, 2012; Lonigan et al., 2000), as latent constructs given evidence of high internal consistency and unidimensionality (e.g., Invernizzi et al., 2004; Phillips, Piasta, Anthony, Lonigan, & Francis, 2012), technical inaccuracies evident in sum scores (Bowles et al., 2011), and similar predictive utility to later reading for sum and latent scores (e.g., Lonigan et al., 2000; Schatschneider et al., 2004). Recent studies differ on whether letter-name and letter-sound skills are characterized as representing a unitary underlying capability (e.g., Drouin et al., 2012) or overlapping but distinctive constructs (e.g., Huang & Invernizzi, 2012; Kim, Petscher, Foorman, & Zhou, 2010). Letter-sound knowledge is a more difficult and later-developing skill than letter-name knowledge for children educated in the United States (Drouin et al., 2012; Ellefson, Treiman, & Kessler, 2009). In prior studies in which both letter-sound and letter-name knowledge have been entered into predictive models, the strongest predictor appears to depend in part on when the skills are assessed (Ritchey & Speece, 2006; Schatschneider et al., 2004; Speece, Mills, Ritchey, & Hillman,

2002). Arguably, letter-sound knowledge is a better predictor because it is more closely aligned with the process of learning the alphabetic code (Ehri, 1998; Treiman, Tincoff, Rodriguez, Mouzaki, & Francis, 1998).

A review of state early learning standards clearly indicates that attention to the development of alphabet knowledge is widely deemed a critical task for early childhood educators (Piasta, Petscher, & Justice 2012; Scott-Little, Kagan, & Frelow, 2006). Although more states address letter-name knowledge than letter-sound knowledge, most do include acquisition of some letter-sound associations within their emergent literacy goals for 4-year-old children. Specifically, as of 2011, 27 states and the District of Columbia include letter-sound acquisition as a learning target (Piasta, 2011). Whereas most of these states provide general goal statements regarding learning letter sounds (e.g., know “some” or “begin to [learn]”), some states include specific numeric targets ranging from 3 (Alaska) to 20 (Texas). In addition, the Head Start Readiness Framework includes a focus on learning letter-sound relationships (i.e., “identifies letters and associates correct sounds with letters”; U.S. Department of Health and Human Services, 2010, p. 15). Such emphasis on letter-sound knowledge also aligns with the new Common Core State Standards, which indicate that kindergarten students should know all common consonant sounds as well as both long and short vowel sounds (National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010).

Alphabet knowledge, and letter-sound knowledge in particular, is a compelling assessment focus not only because of its strong predictive efficacy but also because it is malleable as an early instructional target. Evidence from high-quality smaller-scale (e.g., Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2012; Piasta, Purpura, & Wagner, 2010) and larger-scale (e.g., Jackson et al., 2006; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011) preschool studies indicates that targeted instruction assists even children at risk for later reading difficulties in making significant alphabet knowledge gains. Therefore, access to high-quality criterion assessments for measuring individual children’s progress in acquiring letter-sound knowledge could be a distinct asset for teachers across varied instructional settings. Notably, most preschool curricula do not provide psychometrically sound progress monitoring of letter-sound knowledge (Spencer, Spencer, Goldstein, & Schneider, 2013).

The Need for Formative Assessments

Given increased recognition of the importance of developing letter-sound knowledge as part of a comprehensive set of emergent literacy capabilities, preschool teachers have followed the lead of elementary teachers (e.g., Roehrig et al., 2008; Stecker, Fuchs, & Fuchs, 2005) and increasingly recognized the potential benefits of formative measures of instructionally relevant content (Gettinger & Stoiber, 2012; Invernizzi, Landrum, Teichman, & Townsend, 2010; VanDerHeyden, Snyder, Broussard, & Ramsdell, 2008). Formative assessments can, when associated with appropriate professional development and aligned instructional materials, assist teachers in making data-based decisions regarding instructional pacing, grouping, and content coverage (Fuchs et al., 2004; Walker, Carta, Greenwood, & Buzhardt, 2008) and support teachers’ individualization of instruction by providing real-

time, predictively accurate feedback regarding which children have not yet mastered particular content and require further instruction (Petscher, Kim, & Foorman, 2011; Solari, Petscher, & Folsom, 2012; Schatschneider, Petscher, & Williams, 2008). Our development of easy-to-administer yet sophisticated letter-sound assessments is built on the assumption that the ability to track students' attainment in this skill area would be a good fit to virtually all instructional systems. Furthermore, we assume that feasibility is a key aspect of high-quality assessments (Deno, 2003; Fewster & Macmillan, 2002). Specifically, teachers will be more likely to regularly use assessments that are cost and time effective and result in minimal diversion of time from instruction. In the preschool arena, it is also important that measures can be administered reliably by teachers with a wide variety of credentials, given that many such educators do not have formal teaching certificates or bachelor's degrees (Early et al., 2007; Torquati, Raikes, & Huddleston-Casas, 2007).

Emerging evidence suggests that educators with an array of educational backgrounds can reliably administer measures related to alphabet knowledge (and other emergent literacy skills) when presented with a simple-to-administer assessment and clear guidelines (Invernizzi et al., 2010; Lonigan, Allan, & Lerner, 2011). For example, results from the Florida Voluntary Prekindergarten Assessment Measures field trial indicated interrater reliability of .80 or better between trained research assessors and classroom teachers for print knowledge assessment (Lonigan, 2011). Thus, if high-quality formative assessments for alphabet knowledge were made available to preschool educators, there is reason to believe that they could obtain valid and useful data from these assessments and use this information to inform their instructional decision making. We turn now to a brief discussion of existing assessment options.

Letter Knowledge Assessments

Existing measures of letter-sound knowledge fall into three primary categories. First, there are formative fluency-based measures, such as DIBELS letter-sound fluency (Good, Kaminski, Smith, Laimon, & Dill, 2001), and comparable curriculum-based letter-sound fluency measures developed by Fuchs and Fuchs (2001) and others (Alonzo & Tindal, 2007; Betts, Pickart, & Heistad, 2009; see also *Indicators of Early Progress—Early Reading*; Istation, 2014). Whereas most or all of these assessments have adequate or better psychometric properties, timed letter-sound measures for young children may pose a number of challenges. Many young children have slow or tenuous access to letter sounds in memory or may not understand the nature of timed assessments, resulting in floor effects (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009) and potential underestimates of children's abilities. Moreover, alphabet "rate" and "accuracy" may be only modestly correlated (e.g., .5; Speece et al., 2002), and measures of accuracy more directly map onto the alphabet-learning standards that teachers are currently asked to meet (i.e., progress toward knowing all letter-sound correspondences). In addition, current fluency-based measures have not been evaluated with preschool-age populations and, similar to many of the other available letter-sound assessments described subsequently, do not take into account interletter differences in how readily names and sounds are acquired (Phillips et al.,

2012; Piasta & Wagner, 2010b; Treiman et al., 1998, 2006; Treiman, Pennington, Shriberg, & Boada, 2008).

Second, there are screening measures such as *Get Ready to Read!* (Whitehurst, 2001), *EARLI* probes (Reid, DiPerna, Morgan, & Lei, 2009), and *PALS-PreK* (Invernizzi, Sullivan, & Meier, 2001) that, although valid and readily accessible by teachers, include only a single form, making them less amenable to repeated use within a formative assessment system. A limitation of some measures (e.g., *PALS*) is that they include all 26 letters, often in both uppercase and lowercase forms, which can be quite time consuming to administer to all children in a classroom. Furthermore, since these particular measures do not have ceiling rules, the low ability of many children at the beginning of preschool may lead to unintended frustration during the assessment process. A limitation of other measures is the inclusion of only a few letter sounds that are combined with other, non-letter-sound items to form a broader scale (e.g., *Florida Voluntary Prekindergarten Assessment*; Lonigan, 2011; *Get Ready to Read!*). Many published assessments (Lonigan, 2011), and certainly most created by teachers, use arbitrary or imprecise methods for choosing letter-sound items, which fails to account for interletter differences, limits their validity, and renders less meaningful comparisons between scores on different measures or at different assessment points, an issue comparable to nonequivalence of oral reading fluency passages (Christ & Ardoin, 2009; Francis et al., 2008).

Multimeasure diagnostic assessments such as the *Test of Preschool Early Literacy* (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), the *Woodcock Reading Mastery Test* (Woodcock, 1998), and the *Test of Early Reading Ability* (TERA; Reid, Hresko, & Hammill, 2001) constitute the final measure type, all of which include letter-sound items as part of larger assessment measures of print knowledge or decoding sight words. Whereas these measures have the advantage of norm-referenced scores, their need for brevity and content coverage means that only a small subset of letter sounds are assessed, and it is unknown whether the letters selected for inclusion represent an optimized subset of letter sounds for this age group.

We also note that none of the existing letter-sound measures attend to the phenomenon referred to as the first-initial advantage, in which children tend to show greater knowledge of the first letter of their first names. The first-initial advantage is well established with respect to children's letter-name knowledge (Justice, Pence, Bowles, & Wiggins, 2006; Treiman & Broderick, 1998; Treiman, Kessler, & Pollo, 2006), with mixed evidence as to whether the advantage also applies to letter-sound knowledge (Levin & Aram, 2005; Treiman & Broderick, 1998). A first-initial advantage could unfairly bias scores of letter-sound assessments. For example, scores could be affected by whether or not a child's first initial was included in the subset of letters assessed or how early the first initial is assessed in timed measures, such that this issue requires direct attention during measurement development.

The Current Study

Our purpose was to fill a need left by these extant assessment types with a set of formative assessments of letter-sound knowledge developed specifically for the

preschool period. The letter-sound short forms developed within this project are untimed and require no more than 3 to 5 minutes per child per administration to quickly and accurately ascertain a child's performance under age-appropriate conditions. Moreover, we used item response theory (IRT) to create brief assessments that overcame psychometric problems evidenced in extant measures (Bowles et al., 2011). The IRT framework, in which letter-sound knowledge is conceptualized as a latent ability, allowed us to address the need to model interletter differences, account for a potential first-initial advantage, and, perhaps most importantly, ensure equivalence across short forms during measure development. Notably, by using IRT, we take a representational approach in which the attribute of interest, letter-sound knowledge, is independent of the assessment method (i.e., not defined by how it is measured), including the specific letter sounds assessed on each short form (Borsboom, Mellenbergh, & van Heerden, 2003; see also Bowles et al., 2011). Using IRT to ensure form equivalence, these measures can be used for formative progress monitoring requiring repeated assessments or summative pre- and post-alternate-form assessment within the context of research studies. Moreover, the availability of multiple forms alleviates the potential bias in administering a form containing a child's first initial.

Method

Participants

Participants in the current study were drawn from two projects conducted in a large city in east Texas. In each project, letter-sound knowledge was assessed in multiple assessment waves. The first project was a longitudinal study of preschool children's emergent literacy development (Anthony, Williams, McDonald, & Francis, 2007); the second was an experimental evaluation of a program that trained parents in shared reading strategies (Anthony, Williams, Zhang, Landry, & Dunkelberger, 2014). Combined, these projects included 940 children from center-based preschool programs, including Head Start, public prekindergarten, and private child care. Children were between 3 and 6 years of age upon study entry (mean age = 52 months, $SD = 7.34$). Gender was equally represented in the sample, and the sample was ethnically diverse: 9% White non-Hispanic/Latino, 45% African American non-Hispanic/Latino, 43% Hispanic/Latino, 3% multiracial or other, and 1% unreported. Children with known or obvious sensory, physical, or cognitive impairments were excluded. All children were either native English speakers or passed an English language screening measure. Most children were assessed multiple times across the school year, with 11% of children assessed only once, 55% assessed twice, 4% assessed three times, and 30% assessed four times.

We divided the resulting 2,375 individual data points into two datasets for purposes of the present study—a calibration dataset and a validation dataset. For children who were assessed more than once, we randomly selected one wave of data to be included in the calibration dataset and a different wave of data to be included in the validation dataset; children with missing letter-sound data at the randomly selected wave were not included in the present analyses (four for calibration dataset and two for validation dataset). Children who were assessed only once were ran-

domly assigned with equal probability to either the calibration or validation dataset. The final calibration dataset included data from 884 children, and the final validation dataset included data from 885 children. Age ranges and demographic information for the two datasets mirrored those described for the full sample. Moreover, there were no differences in the average number of letter sounds known across the two datasets, $t(1767) = -0.19$, $p = .849$, and both datasets displayed the full range (0 to 26) of letter-sound knowledge.

Procedures

After receiving parental consent, children completed a number of emergent literacy, language, and cognitive assessments as per the protocols of the larger studies. Children were assessed individually at their preschools by trained research assistants in multiple sessions. Sessions lasted no longer than 45 minutes each, with only one session completed per day.

Both of the larger studies included an assessment of children's letter-sound knowledge, which constituted the focal assessment for the current study. The letter-sound assessment involved asking children to respond to the prompt of "What sound does this letter make?" when shown the uppercase and lowercase versions of a letter, shown side by side (e.g., N n) for all 26 letter pairs. Letters were printed in Arial 30-point font down the center of an $8\frac{1}{2} \times 11$ -inch piece of paper and presented one pair at a time by sliding a card down the page and exposing the next letter pair. Letter pairs were presented in the same random order to all children. Each individual letter pair was scored as correct (1) or incorrect (0). Because most English letters are associated with multiple sounds, any sound commonly associated with a given letter, including long vowel sounds, was accepted as correct. Other than vowels, if a letter name was given, the child was prompted, "That's the name of this letter. Tell me the sound this letter makes."

Analytic Strategy

The aim of the present study was accomplished via application of IRT, a flexible and informative measurement framework that considers latent characteristics of both the child and the item in modeling the likelihood of a correct response (de Ayala, 2009; Embretson & Reise, 2000). To accomplish our aim of developing letter-sound short forms, we used IRT analysis to determine an appropriate measurement model (Phase 1) following procedures similar to those reported in Phillips and colleagues (2012). This included use of calibration and validation datasets to analyze the assumptions of IRT, select an appropriate IRT model, generate and verify item parameters and fit, and examine theta recovery. We also confirmed that our measurement model was not biased by differential item functioning due to letters' inclusion in children's names. All Phase 1 analyses were conducted using Mplus software (Muthén & Muthén, 2008) and maximum-likelihood estimation, with the 26 letter items modeled as binary, categorical data.

In Phase 2, we developed and validated short forms to assess letter-sound knowledge. Specifically, we began by generating a series of possible short forms based on Phase 1 results and estimated test information for all such forms. We then

selected optimal forms based on test information across the full range of theta and examined these forms with respect to reliability and theta recovery. Finally, we developed a scoring system that not only provides theta estimates for scoring the short forms but also maps these theta scores to raw scores on a 0 to 26 scale. The Phase 2 analyses were conducted using SAS (2012) and the calibration dataset.

Results

Table 1 presents descriptive information concerning the letter sounds known by children participating in the study, with similar results for the calibration and validation datasets.

Phase 1: Appropriate Measurement Model Determination

Step 1: Dimensionality. We first performed an item-level exploratory factor analysis with the calibration dataset, which indicated a single, dominant factor

Table 1. Percentage (by Letter) and Total Number of Correct Letter-Sound Responses in the Calibration and Validation Datasets

Letter	Calibration Dataset		Validation Dataset	
	Correct Sound Given (%)	<i>n</i>	Correct Sound Given (%)	<i>n</i>
A	46.09	883	47.96	884
B	44.28	883	44.73	883
C	46.43	883	46.10	885
D	37.07	882	38.35	884
E	33.30	883	32.24	884
F	37.22	884	38.42	885
G	29.90	883	32.24	884
H	27.38	884	26.89	885
I	24.12	883	23.87	884
J	37.60	883	37.63	885
K	38.05	883	39.25	884
L	27.75	883	28.05	884
M	34.43	883	33.37	884
N	26.08	882	25.93	883
O	38.39	883	38.46	884
P	41.06	884	42.76	884
Q	25.48	883	26.70	880
R	25.54	881	26.13	884
S	40.48	882	42.58	883
T	42.36	883	43.78	884
U	15.99	882	16.08	883
V	32.54	882	31.48	883
W	21.83	884	22.20	883
X	22.42	883	20.25	884
Y	13.59	883	12.46	883
Z	37.60	883	37.33	884
	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>
Total no. of letter sounds correct	8.46 (9.07)	884	8.54 (8.99)	885

(first eigenvalue = 21.016; subsequent eigenvalues ranged from 0.693 to 0.012) and thus the appropriateness of the data for IRT analysis. Unidimensionality was also verified via confirmatory factor analysis, by imposing a one-factor model on the validation dataset and achieving an excellent fit ($CFI = .993$, $TLI = .999$, $RMSEA = .040$).

Step 2: Selection of appropriate unidimensional IRT model. We next tested the fit of the data to two standard IRT models (one-parameter logistic model [1PL] vs. two-parameter logistic model [2PL]). Results supported the 2PL model for both the calibration dataset ($\Delta\chi^2 = 194.72$, $\Delta df = 25$, $p < .001$) and the validation dataset ($\Delta\chi^2 = 237.24$, $\Delta df = 25$, $p < .001$). The 2PL model was thus used in all subsequent analyses.

Step 3: Item parameter estimation. Our third step involved using the 2PL model to estimate item parameters. We first estimated these parameters using the calibration dataset. These results are presented in the first two columns of numbers in Table 2. Item discriminations ranged from 1.478 for the least discriminating letter (X) to 3.405 for the most discriminating letter (P). Most consonants evidenced very good discrimination (i.e., above 2). In contrast, vowels and the consonants X and C evidenced only moderate discrimination. Difficulties ranged from 0.136 for the least difficult letter (C) to 1.234 for the most difficult letter (Y). Examination of individual item fit (Orlando & Thissen, 2000) revealed two letters (H, O) that demonstrated significant lack of fit with the overall model (see third column of Table 2). Notably, given that theta is assumed to have a mean of 0 and standard deviation of 1, all of the letters are relatively difficult for this sample, and the item difficulty range is only approximately 1 standard deviation. Although the range of theta covered was somewhat limited, measurement of the full scale across that ability range was acceptable, as indicated by the conditional standard error of measurement at different levels of ability (see Table 3).

We next examined the consistency of item parameters by freely estimating these parameters using the validation dataset and comparing them to those freely estimated for the calibration dataset (i.e., those described above). The item parameters generated using the validation dataset are presented in the third and fourth columns of numbers in Table 2. Item parameters were largely consistent across datasets. Although some slight fluctuations in individual letter discriminations and difficulties were noted, parameter differences were small, with the absolute value of the difference between calibration and validation estimates ranging from .009 to 0.357 for discriminations ($M = 0.133$) and .001 to .096 for difficulties ($M = .032$). Moreover, parameters estimated with calibration versus validation datasets were highly correlated ($rs = .956$ and $.995$ for discriminations and difficulties, respectively). We were thus confident that the item parameters were consistent across datasets.

Step 4: Theta recovery. As an additional step in cross-validating the model, we next examined the thetas, or estimates of children's latent letter-sound knowledge abilities, generated by the 2PL IRT models described above. Specifically, we compared the theta estimates generated for children in the validation dataset using two different parameterizations. The first utilized the freely estimated item parameters listed in the third and fourth columns of numbers in Table 2. The second constrained item parameters to the values generated with the calibration dataset (see first and second columns of numbers in Table 2). If the model is accurately

Table 2. Item Parameters for IRT Models

Letter	Calibration Dataset 2PL Model (Selected as Final Model)			Validation Dataset 2PL Model (Freely Estimated)		Calibration Dataset DIF Model				Calibration Dataset, First Initial as 27th Letter Model	
	Discrim	Diff	S- χ^2 (df)	Discrim	Diff	Not First Letter of Name		First Letter of Name			
						Discrim	Diff	Discrim	Diff		p^a
A	1.78	.15	26.9 (24)	1.65	.09	1.80	.15	.03	.25	1.76	.16
B	2.50	.20	17.4 (24)	2.35	.19	2.54	.21	-.14	.03	2.47	.21
C	1.80	.14	21.7 (23)	1.69	.15	1.83	.15	-.21	.02	1.84	.15
D	2.94	.40	29.8 (23)	2.96	.37	2.96	.40	.31	.33	2.97	.41
E	1.80	.53	33.2 (22)	1.78	.57	1.82	.54	.18	.03	1.82	.54
F	2.74	.39	26.7 (24)	2.63	.37	2.75	.39	.08	.28	2.76	.4
G	2.63	.60	33.8 (23)	2.79	.54	2.64	.60	.50	.56	2.6	.61
H	2.31	.69	37.7* (24)	2.21	.71	2.34	.69	.26	.03	2.32	.7
I	1.55	.86	21.5 (23)	1.25	.93	1.55	.86	.41	.13	1.55	.87
J	2.45	.38	33.2 (24)	2.46	.39	2.51	.43	.15	.000**	2.53	.43
K	2.79	.37	29.4 (24)	2.97	.34	2.80	.38	.19	.06	2.81	.39
L	2.88	.66	26.9 (21)	2.52	.67	2.99	.67	.33	.004	3.03	.68
M	2.45	.48	21.4 (24)	2.39	.51	2.48	.49	.31	.08	2.52	.49

N	2.69	.72	30.1 (22)	2.64	.73	2.69	.71	.78	.73	2.63	.72
O	1.62	.38	31.0* (19)	1.61	.38	1.62	.37	.53	.74	1.62	.38
P	3.41	.29	33.9 (22)	3.18	.24	3.46	.29	-.09	.06	3.45	.29
Q	2.15	.76	26.6 (23)	2.04	.73	2.15	.750 ^b			2.15	.76
R	2.58	.74	26.3 (22)	2.27	.74	2.59	.74	.61	.38	2.56	.75
S	2.08	.31	24.5 (22)	2.07	.25	2.12	.32	-.07	.01	2.15	.32
T	2.52	.25	32.3 (24)	2.71	.22	2.52	.25	.18	.62	2.49	.26
U	1.54	1.20	27.1 (22)	1.46	1.21	1.53	1.196 ^c			1.53	1.21
V	2.71	.53	23.7 (16)	2.97	.56	2.73	.52	.62	.73	2.70	.53
W	2.29	.87	25.4 (24)	2.32	.86	2.31	.859 ^c			2.30	.87
X	1.48	.94	24.2 (24)	1.58	1.01	1.48	.93	.54	.41	1.47	.94
Y	1.96	1.23	29.7 (23)	1.74	1.33	1.96	1.228 ^c			1.95	1.24
Z	2.47	.39	33.0 (24)	2.29	.40	2.47	.38	.11	.33	2.45	.39
First initial ^d										2.09	.21

Note.—DIF = differential item functioning; Discrim = discrimination item parameter; Diff = difficulty item parameter; S- χ^2 = Orlando and Thissen's (2000) S- χ^2 item fit.

^a Type I error rate set at .05/22 = .002 for DIF analyses.

^b DIF analyses not possible as no child had Q as the first letter of his or her first name.

^c DIF analyses not possible due to insufficient variances in the joint distribution.

^d Parameters for children's first initials were estimated only for the final model.

* $p < .05$.

** $p < .002$.

Table 3. Summed Score to Scale Score (EAP)
Conversion and Conditional Standard Error of
Measurement for the Full Model

Summed Score	EAP[θ x]	SD[θ x]
0	−1.620	.492
1	−1.140	.296
2	−.918	.219
3	−.778	.182
4	−.671	.162
5	−.585	.142
6	−.520	.124
7	−.462	.124
8	−.399	.132
9	−.334	.123
10	−.285	.101
11	−.251	.088
12	−.218	.099
13	−.170	.120
14	−.108	.127
15	−.051	.110
16	−.012	.091
17	.021	.095
18	.064	.117
19	.127	.133
20	.197	.130
21	.261	.129
22	.334	.147
23	.432	.173
24	.569	.208
25	.790	.290
26	1.336	.528

specified, theta estimates should be similar regardless of which parameterization was used.

All theta estimates were derived directly from MPlus. Children’s theta estimates using the two different parameterizations were correlated at $r = 1.000$. The mean theta when freely estimated from the validation dataset was 0.001 ($SD = 0.934$). When estimated using the calibration dataset item parameters, the same children’s mean theta was 0.013 ($SD = 0.914$). The difference between a child’s pair of theta scores ranged in value from -0.089 to 0.085 . The small difference in theta scores indicates high theta recovery and further demonstrates the accuracy and consistency of the selected IRT model and its item parameters. This 2PL model, with the item parameters listed in the first two columns of Table 2, was selected as the most appropriate measurement model for estimating children’s letter-sound knowledge.

Step 5: Differential item functioning due to own-name advantage. As a precaution, we performed one final analysis in Phase 1 to rule out children’s names as a source of differential item functioning (DIF), given the potential for the first-initial advantage. We considered two forms of DIF analyses. The first utilized the procedures advocated by Muthén and colleagues (Muthén, 1989; Muthén, Kao, & Burstein, 1991) and previously applied to letter-name data by Phillips and col-

leagues (2012). This analysis augments the typical 2PL model by regressing item responses for the 26 individual letters on 26 dummy codes indicating whether or not the given letter was the first in a child's first name (i.e., first initial). This parameterization provides two estimates of item difficulty: one for children who have a given letter as the first in their names and one for children who do not have a given letter as the first in their names. DIF exists when these regression coefficients are significantly different after adjusting the Type I error rate for multiple tests. The second DIF analysis utilized a version of Lord's (1980) Wald test as implemented in IRTPRO (Cai, Thissen, & du Toit, 2011). This analysis augments the typical 2PL with parameters reflecting DIF in item difficulty and item discrimination, with a Wald test to test whether the DIF parameters are significantly different from 0. This test is performed separately for each item assuming the DIF parameters are 0 for all items other than the focal item, adjusting for the Type I error rate for multiple tests. Consistent with the approach implemented in IRTPRO, we focused on a 1 *df* Wald test for DIF in the item discrimination, followed by a 1 *df* Wald test for DIF in the item difficulty conditional on the item discrimination.

The results of our DIF analysis, as estimated with the calibration dataset, were substantively consistent with both approaches, with no evidence of DIF with the item discriminations, so we report only the results for the first approach. These results are presented in the fifth through eighth columns of numbers in Table 2. Note that no child had the letter Q as a first initial, and there was insufficient variance in the joint distributions to estimate separate difficulties or standard errors for the letters U, W, and Y. DIF was thus interpreted only with respect to the remaining 22 letters. As Table 2 indicates, only the letter J showed significant DIF once a Bonferroni adjustment was made to correct the Type I error rate. An identical pattern was noted when the same analysis was conducted on the validation dataset. Moreover, when an additional model (last two columns of Table 2) was estimated in which children's responses to the letter sound associated with their initial was removed as a source of bias (by coding as a separate, 27th letter; see Phillips et al. [2012] for further details regarding this type of analysis), (a) the correlation between the discrimination parameter estimates from this model and the original 2PL model (i.e., columns 1 and 9 in Table 2) was $r = .997$, (b) the correlation for difficulty parameters for these models was $r = 1.000$, and (c) theta recovery for the validation dataset was perfect (i.e., $r = 1.000$). We thus had no reason to adjust our final, selected model from Step 4, which was subsequently utilized for short-form creation in Phase 2.

Phase 2: Short-Form Creation

In Phase 2, we used the final 2PL model to create multiple, parallel forms that could be administered across the year without repeating letters and that would yield comparable (theta) scores. We considered dividing the letters into three forms of eight letters each, and, alternatively, four forms of six letters each. In either case, two letters would not be included on any form. Because letters H and O had relatively low discriminations, significant $S\text{-}\chi^2$ item-fit statistics, and similar difficulties to other letters, these letters were excluded from the letter-sound short forms.

Step 1: Assignment of letters to forms. We wanted each short form to (a) cover the full range of theta measured when assessing all 26 letters and (b) provide the same amount of test information (de Ayala, 2009) as the other short forms of the same length. To ensure the best coverage for all forms, we ordered the letters according to difficulty and used block randomization to assign letters to forms. For example, with three forms, there are eight blocks of three letters. The three letters with the lowest difficulty were assigned to the first block, the next three to the second block, and so on; letters within each block were then randomly assigned to a form.

To satisfy the second requirement that each form provide approximately the same amount of information, we repeated the randomization 10,000 times and calculated the maximal test information (i.e., information at the peak of the test information curve) for each form for every randomization. The standard deviation of the test information of all the forms was then calculated and used as an index of form comparability, in which lower standard deviations indicated higher comparability of forms. The randomization with the lowest standard deviation was selected as the best random assignment for both the three-form and four-form version, as this yielded the most comparable scores across the forms (see Table 4 for letters per form).

Step 2: Comparison of three-form and four-form versions. Properties of the three- and four-form versions were compared to each other and also to the full, 26-letter model in terms of test information, reliability, and theta recovery. The properties of the three- and four-form versions were highly similar. Test information for both versions is presented in Figures 1 and 2. Because the amount of information provided by a test is the sum of the information of each item, the three eight-letter forms by definition provided more information than the four six-letter forms. The average reliability of the different forms was calculated according to Raykov, Dimitrov, and Asparouhov (2010) and is presented in Table 4. Although reliability of the short forms was substantially lower than that of the full model, it remained high for all versions of the short forms ($\rho = .92-.93$ for the three-form version; $\rho = .89-.91$ for the four-form version). Finally, we generated theta estimates for each child using the full 26-letter model finalized in Phase 1 and

Table 4. Reliabilities for the Full Model (All 26 Letters) and Final Short Forms, and Correlations between Short Forms and Theta Scores from the Full Model

Form	Letters	Reliability	Correlation with Theta Score
Full	All 26	.98	n/a
24 letters	All except H and O	.97	n/a
3 Forms-1	A, P, Z, M, E, L, W, X	.92	.75
3 Forms-2	B, S, K, F, G, N, I, Y	.93	.75
3 Forms-3	C, T, J, D, V, R, Q, U	.93	.77
4 Forms-1	C, P, F, G, I, U	.89	.75
4 Forms-2	B, S, Z, V, N, Y	.91	.75
4 Forms-3	A, K, M, L, R, X	.90	.75
4 Forms-4	T, J, D, E, Q, W	.91	.73

Note.—Reliabilities calculated from IRT parameters (not converted scores) according to Raykov et al. (2010). Correlation with theta scores from the full model with all 26 items.

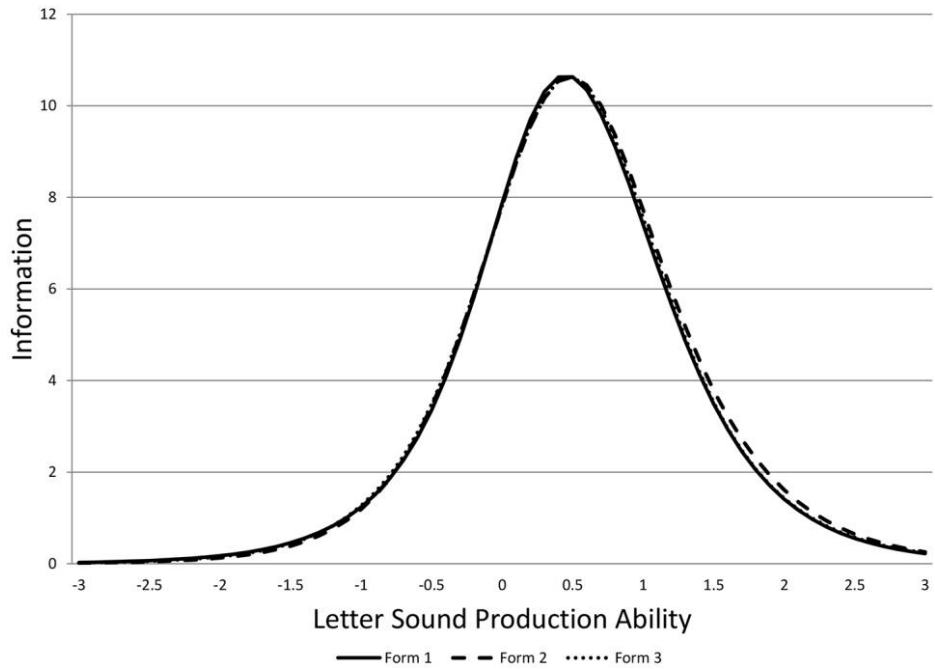


Figure 1. Test information curves for the three-form version. These were derived from the 2PL model for the calibration dataset, which was also selected as the final measurement model.

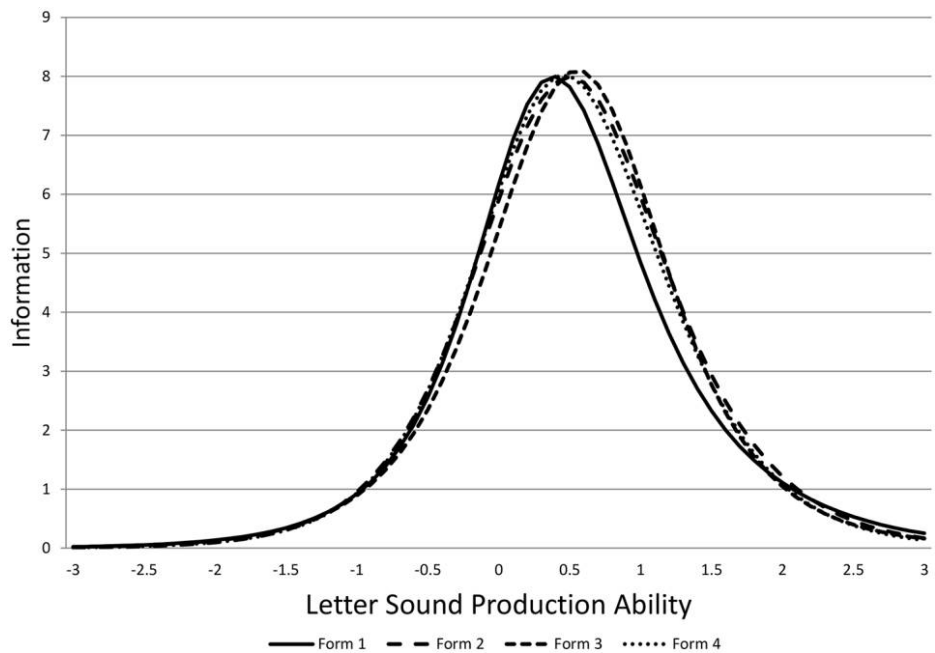


Figure 2. Test information curves for the four-form version. These were derived from the 2PL model for the calibration dataset, which was also selected as the final measurement model.

compared these to the theta estimates derived using the short forms. Correlations were similar across all forms (see Table 4). We conclude that either three- or four-form versions could be used to reliably assess children's letter-sound knowledge.

Step 3: Scoring. Finally, because the short forms are intended to be used in the field by teachers or other school personnel, we developed a scoring system that capitalizes on the results of the IRT analysis yet yields scores that require no theta estimation from the pattern of item responses. We employed the technique of Thissen and Orlando (2001), which uses sum scores to yield maximum-likelihood theta estimates that are close approximations of the theta score estimated using the full pattern of responses. Conceptually, the technique averages the theta estimate for all possible patterns of item responses for a particular sum score, weighted by the probability of the pattern, using a computational algorithm adapted from Lord and Wingersky (1984). We converted these estimates to a scale with a mean of 20 and standard deviation of 2 for easier interpretation. We also calculated the predicted number of correct responses if the child had responded to all 26 letters, akin to the 0 to 26 sum scores that most researchers and practitioners utilize, as we expect that such sum scores would be more meaningful and useful to teachers than the IRT-based scaled scores. Results are presented in Table 5. The first column of the table presents the raw score, or total number of letters for which the child pro-

Table 5. Scaled and Sum Scores for Short-Form Letter-Sound Assessments

Three-Form Version (8 Letters per Form)								
No. Correct	Form 1: APZMELWX		Form 2: BSKFGNIY		Form 3: CTJDVRQU			
	Scaled Score (SE)	Sum Score	Scaled Score (SE)	Sum Score	Scaled Score (SE)	Sum Score		
0	17.89 (1.27)	.87	17.95 (1.26)	.92	17.90 (1.26)	.88		
1	19.09 (.95)	2.76	19.22 (.90)	3.11	19.14 (.92)	2.88		
2	19.87 (.76)	5.59	19.97 (.72)	6.09	19.90 (.74)	5.74		
3	20.45 (.67)	8.84	20.53 (.65)	9.33	20.47 (.66)	8.94		
4	20.94 (.64)	12.17	21.01 (.63)	12.63	20.95 (.63)	12.22		
5	21.43 (.66)	15.47	21.49 (.65)	15.91	21.43 (.65)	15.50		
6	21.96 (.73)	18.66	22.03 (.72)	19.05	21.97 (.72)	18.70		
7	22.62 (.86)	21.58	22.70 (.85)	21.85	22.64 (.86)	21.67		
8	23.52 (1.07)	23.90	23.61 (1.07)	24.05	23.53 (1.07)	23.93		
Four-Form Version (6 Letters per Form)								
	Form 1: CPFGIU		Form 2: BSZVNY		Form 3: AKMLRX		Form 4: TJDEQW	
	Scaled Score (SE)	Sum Score	Scaled Score (SE)	Sum Score	Scaled Score (SE)	Sum Score	Scaled Score (SE)	Sum Score
0	18.06 (1.30)	1.02	18.09 (1.29)	1.06	18.10 (1.31)	1.06	18.10 (1.29)	1.06
1	19.34 (.99)	3.49	19.48 (.92)	3.94	19.40 (.99)	3.68	19.45 (.93)	3.86
2	20.23 (.81)	7.50	20.30 (.76)	7.89	20.28 (.80)	7.76	20.28 (.77)	7.75
3	20.94 (.77)	12.14	20.95 (.71)	12.19	20.28 (.80)	12.21	20.92 (.72)	11.98
4	21.63 (.82)	16.77	21.58 (.74)	16.49	21.59 (.76)	16.53	21.54 (.75)	16.17
5	22.40 (.95)	20.72	22.34 (.86)	20.50	22.34 (.88)	20.50	22.25 (.86)	20.09
6	23.33 (1.15)	23.55	23.36 (1.09)	23.60	23.30 (1.10)	23.48	23.27 (1.10)	23.42

Note.—Scaled score is the theta estimate converted to a scale with mean 20 and SD 2. Sum score refers the predicted number of correct responses for all 26 items.

duced a correct sound. Subsequent columns give the scaled and sum scores for each short form. Assessors can use these as look-up tables. For example, a child with a raw score of 7 letters correct on Form 2 of the eight-letter form has a scaled score of 22.65, which is just more than 1 *SD* above the mean. Moreover, if this child had been assessed with all 26 items, he or she would have been expected to correctly produce sounds for 21 or 22 letters. Copies of each short form, along with administration instructions and scoring information, are available at <http://ccec.ehe.osu.edu/resources/assessments>.

Discussion

The present study fills an important gap in the literature, in creating empirically derived letter-sound short forms that can be used for formative as well as summative purposes. The study not only responds to calls for more sophisticated measurement of alphabet knowledge (Bowles et al., 2011; Drouin et al., 2012; Paris, 2005), but also to increased emphasis on alphabet learning and instruction (Piasta & Wagner, 2010a; Piasta et al., 2012; Scott-Little et al., 2006) and data-based decision making (Gettinger & Stoiber, 2012; Invernizzi et al., 2010; Spencer et al., 2013). Our findings make important contributions to research and, potentially, to educational practice.

Contributions to Research and Theory

From a research standpoint, the current study and its assessments address some of the psychometric limitations evident in previous letter-sound measurement. The letter-sound assessments are shown to provide reliable results with preschool-age children and avoid the construct confusion inherent in composite measures of alphabet knowledge or broader early literacy skills. Most importantly, the assessments account for interletter variability. Previous studies have documented that children are more or less likely to know the sounds of particular letters (e.g., Evans, Bell, Shaw, Moretti, & Page, 2006; McBride-Chang, 1999; Treiman et al., 1998), with 12% to 26% of the variance in children's letter-sound knowledge attributable to interletter differences (Kim et al., 2010; Piasta, 2006; Piasta & Wagner, 2010b). The current study complements and extends these findings in two key ways.

First and foremost, our IRT-based analyses provide additional evidence of variation in letter difficulties by directly estimating this item characteristic (see also Drouin et al., 2012) and also document that letters vary in a second characteristic, namely, the extent to which specific letters discriminate among children with varying levels of letter-sound knowledge. These findings confirm that not all letters contribute equally to measurement of children's letter-sound knowledge; assessments must take interletter differences into account to accurately estimate children's letter-sound knowledge. This is especially important when selecting subsets of letters for repeated administration and when comparing performance on letter subsets, as arbitrarily selected subsets may include an overabundance of particularly difficult or easy letters and thus vary in the precision and reliability of estimates of children's letter-sound abilities. The letter-sound short forms that we established

ensure that differences in scores truly represent differences in letter-sound knowledge and not form variability. Moreover, the provision of multiple forms addresses the first-initial bias (Levin & Aram, 2005) as a specific case of interletter differences. Our results suggest that such concern is particularly legitimate for the letter J. Other letters did not show large evidence of DIF, commensurate with extant studies showing no first-initial advantage for children's letter-sound knowledge (Piasta, 2006; Treiman & Broderick, 1998), although this could not be evaluated for letters U, W, and Y. In using our short forms, however, any bias due to first initials can be alleviated by only administering those forms that do not contain a child's first initial. As an additional advantage, the short forms also place children's letter-sound knowledge on an interval scale via the IRT theta estimates. Together, these strong psychometric characteristics greatly improve our ability to accurately measure children's growth in letter-sound knowledge over time, whether for formative classroom assessment or research-related pretest-posttest assessment. Future research might expand the use of the forms by empirically establishing benchmarks or norms.

Second, and extending beyond basic issues of measurement, our findings also increase understanding of letter-sound knowledge development. For example, our results corroborate previous evidence suggesting that letter-sound knowledge acquisition is challenging for preschool children: the difficulties for all letters were greater than children's average ability level. In comparison, all difficulties but one were less than children's average ability in the IRT analysis of letter-name knowledge conducted by Phillips et al. (2012). Drouin et al. (2012) similarly found that letter sounds were more difficult than letter names when simultaneously examining both in the same IRT model. We note, however, that whereas Drouin et al. concluded that letter sounds were "too difficult" (p. 551) for their preschool sample, the preschoolers in the current study exhibited a fair amount of knowledge concerning letter sounds.

Our findings, based on the largest investigation of preschoolers' knowledge of individual letter sounds to date, also point toward a developmental sequence of letter-sound acquisition. Ordering letters by difficulty, most children seem to acquire the sounds for C and A first, followed by B, and then acquire sounds for T, P, and other letters. Difficulties were fairly clustered for letters subsequent to B, suggesting that children may not follow a clear sequence in acquiring these sounds. However, children did show a pronounced sequence in those letter sounds acquired last, concluding their learning with the sounds of I and W, then X, and finally U and Y. This pattern of results, exhibiting a common sequence for first and last letter sounds, is similar to IRT results when examining letter names (Phillips et al., 2012). In fact, the same three letters (A, B, C) were the most likely to be known in the Phillips et al. study as in the current study of letter sounds, which may reflect the tendency for children to learn the sounds of letters whose names have already been acquired (Evans et al., 2006; Kim et al., 2010; Treiman et al., 1998) as well as children's greater exposure to these letters due to the "ABC song" or other emphasis on these initial three letters (Justice et al., 2006; McBride-Chang, 1999; Piasta, 2006).

This developmental sequence is also consistent with a number of other findings in the literature. All acrophonic letters or letters whose names begin with their corresponding sound (e.g., B, T; also known as consonant-vowel letters) occur within the first half of the developmental sequence, whereas letters with sounds at the

ends of their names (e.g., S, M; also known as vowel-consonant letters) are distributed throughout the sequence, and letters whose names and sounds are unassociated (e.g., H, Y) fall in the second half of the sequence. This is consistent with prior work suggesting that children use letter names as cues in letter-sound learning, particularly for acrophonic letters, and may also account for mixed findings concerning acquisition of vowel-consonant letters (Kim et al., 2010; McBride-Chang, 1999; Piasta & Wagner, 2010b; Treiman et al., 1998, 2008). Also consistent with previous work, the present results corroborate previous evidence indicating that the sounds for letters such as W, X, U, and Y are particularly challenging for young children (e.g., Treiman, Weatherston, & Berch, 1994).

Contributions to Educational Practice

A key goal for this project was to develop reliable and easy-to-administer assessments of letter-sound knowledge appropriate for the preschool period. Preschool teachers are increasingly asked to take a more proactive role in facilitating students' emergent literacy skill development and to teach to early childhood standards that include targets for alphabet knowledge (e.g., Jones & Reutzel, 2012). Given these motivations, teachers can benefit from access to psychometrically sound but simple to use letter-sound assessments appropriate for this age group. The speed and ease of the short forms can facilitate rapid assessment of an entire classroom of children without taking substantial instructional time. Furthermore, given the inclusion of at most eight letters per assessment wave, children who may perform poorly will not experience frustration in being asked to persist at length on a task that may in that moment be quite challenging for them. Another benefit of using a short-form assessment is that its brevity makes it ideal for embedding these items within a larger assessment system that also includes assessments of phonological awareness, language skills, mathematics skills, or general cognitive abilities, all skill areas frequently assessed in early childhood classrooms for either screening or formative purposes (e.g., Hymel, LeMare, & McKee, 2011; Lonigan, Allan, et al., 2011; Panter & Bracken, 2009). Finally, given its brevity and ease of use, we expect that the assessment could be used not only by preschool teachers, but also classroom support personnel (e.g., aides, paraprofessionals, volunteers) as well as others involved in supporting young children's emergent literacy (e.g., community volunteers, children's librarians, physicians).

Our hope is that teachers will utilize letter-sound assessment data, such as those generated by the letter-sound short forms described in this study, to inform instructional decision making. This might include using the short forms to screen for children who are in need of letter-sound instruction and to inform small-group membership and instructional pacing decisions (Gettinger & Stoiber, 2012) to best target individual students' needs. Moreover, teachers might utilize the short-form results to determine whether additional assessment is necessary. Whereas the efficiency of the short-form administration has many desirable aspects, we recognize that there are certainly occasions when teachers will be better served by assessing children on all of their letter sounds. In particular, teachers of children whose scores begin and remain in the mid-range on the short-form scoring system may benefit from obtaining full alphabet assessment so as to best determine which

specific letters an individual student still has not mastered. Likewise, the short forms are not keyed to any specific instructional sequencing and therefore some teachers may also want to assess individual children on the particular letters just taught (e.g., with a more curriculum-linked assessment focus; see, e.g., Lonigan, Farver, et al., 2011).

We note, however, that it remains an open question as to how readily teachers will adopt these measures and how teachers will utilize the formative letter-sound knowledge data generated by our measures. Several prior studies (e.g., Capizzi & Fuchs, 2005; Fuchs, Fuchs, Hamlett, Phillips, & Karns, 1995; Graney & Shinn, 2005) have indicated that teachers do not always know how to optimally adjust and differentiate instruction based on assessment results. Graney and Shinn suggested that simply having the data is unlikely to lead to improvements in student learning in the absence of specific instructional recommendations and ongoing support in modifying their practices. This may be particularly the case in early childhood classrooms where there is a wide range of teacher credentials and expertise and also a range of in-service supports for professional development related to differentiating and modifying instruction (Schumacher, Ewen, & Hart, 2005; Spitler, 2001). With respect to our letter-sound short-form assessment, teachers may need additional guidance from either curricular materials or professional development to understand how to alter or intensify their instructional strategies to maximize student learning of letter sounds. Although beyond the scope of this article, we direct readers to Jones, Clarke, and Reutzel (2013), Phillips and Piasta (2013), and Lonigan et al. (2007) for discussions of evidence-based instructional strategies to promote alphabet knowledge. The few studies of assessment-focused professional development in preschool settings do suggest that teachers can be supported in making relevant instructional adaptations if provided with regular, individualized professional development (e.g., Gettinger & Stoiber, 2012; Landry et al., 2009).

Limitations and Conclusions

Three limitations of the current study are worth noting. First, letters were simultaneously presented to children in both cases. As such, we cannot differentiate the difficulty or discrimination of the letters for upper- and lowercase representations, and additional research is necessary to evaluate the impact of alternative modes of administration on item parameters and scores. Second, we accepted both short and long vowel sounds as correct responses, and we do not have independent information on the distinct phonemes. That long vowel sounds correspond to the names of vowels may explain why the vowel items provided less information about children's letter-sound knowledge than did consonant items. Both long and short vowel sounds are included as learning goals within the new Common Core standards (National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010). However, once children begin reading, their exposure to vowels conveying their long or short sound in printed words will differ, often substantially (e.g., letter *I* as a short sound is much more frequent than as a long sound; Fry, 2004). We might expect that, with time, specific vowel-phoneme associations are more and less difficult for children to acquire (Jones & Reutzel,

2012) and encourage others to further investigate this issue as it applies to letter-sound assessment.

Third, usefulness of the various short forms and the 26-item long form is limited to indexing individual differences in letter-sound knowledge that span a relatively narrow range of ability. Very young preschoolers may exhibit floor effects, and kindergartners would likely exhibit ceiling effects by the end of the school year. Separate assessment of upper- and lowercase letters may expand coverage to a degree. In addition, the upper end of coverage can be extended by adding items of graphophonemic knowledge (e.g., digraphs, consonant clusters), thereby creating a seamless test of graphophonemic knowledge and phonics skills that spans the preschool through early elementary years. Although beyond the scope of the current study, those interested in such assessment are referred to the School Readiness Curriculum Based Measurement System, which is currently being scaled and normed by the last author.

Despite the limitations noted above, the brief letter-sound forms presented in this article represent a substantial improvement over currently available methods for measuring children's letter-sound knowledge, and they address an important gap in emergent literacy assessment. The assessments are both psychometrically sound and easy to use. Together, these characteristics suggest great potential for the assessments to support preschool teachers and others in making data-based instructional decisions that best serve young children's alphabet development and, ultimately, place emergent readers on the path to reading success.

Note

This research was supported by grant P3004179 from the W. K. Kellogg Foundation (Anthony), a program evaluation contract from Gulf Coast Community Services Association Head Start (Anthony), and grant R305E100030 (Piasta) from the Institute of Education Sciences. The opinions expressed are those of the authors and do not necessarily represent views of the funding agencies. Shayne B. Piasta is assistant professor, Crane Center for Early Childhood Research and Policy and Department of Teaching and Learning, The Ohio State University; Beth M. Phillips is associate professor, Florida Center for Reading Research and Department of Educational Psychology and Learning Systems, Florida State University; Jeffrey M. Williams is assistant professor, Department of Pediatrics, University of Texas Health Sciences Center; Ryan P. Bowles is associate professor, Department of Human Development and Family Studies, Michigan State University; Jason L. Anthony is professor, Department of Pediatrics, University of Texas Health Sciences Center. Address all correspondence to Shayne B. Piasta, Crane Center for Early Childhood Research and Policy, The Ohio State University, Columbus, OH 43210; e-mail: piasta.1@osu.edu.

References

- Alonzo, J., & Tindal, G. (2007). *Examining the technical adequacy of early literacy measures in a progress monitoring assessment system: Letter names, letter sounds, and phoneme segmenting* (Technical Report No. 39). Eugene: Behavioral Research and Teaching, University of Oregon.
- 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, 113–137. PMID 18058023

- Anthony, J. L., Williams, J. M., Zhang, Z., Landry, S. H., & Dunkelberger, M. J. (2014). Evaluation of Raising a Reader and supplemental parent training in shared reading. *Early Education and Development, 25*, 493–514.
- Betts, J., Pickart, M., & Heistad, D. (2009). Construct and predictive validity evidence for curriculum-based measures of early literacy and numeracy skills in kindergarten. *Journal of Psychoeducational Assessment, 27*(2), 83–95. doi:10.1177/0734282908323398
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*, 203–219. doi:10.1037/0033-295X.110.2.203
- Bowles, R. P., Skibbe, L. E., & Justice, L. M. (2011). Analysis of letter name knowledge using Rasch measurement. *Journal of Applied Measurement, 12*, 387–399.
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software and manual]. Chicago: Scientific Software International.
- Capizzi, A. M., & Fuchs, L. S. (2005). Effects of curriculum-based measurement with and without diagnostic feedback on teacher planning. *Remedial and Special Education, 26*(3), 159–174.
- Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163–176. doi:10.1177/0022219408326219
- Christ, T. J., & Ardoyn, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology, 47*(1), 55–75. doi:10.1016/j.jsp.2008.09.004
- de Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. New York: Guilford.
- Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3–4), 3–12. doi:10.1177/073724770302800302
- Drouin, M., Horner, S. L., & Sondergeld, T. A. (2012). Alphabet knowledge in preschool: A Rasch model analysis. *Early Childhood Research Quarterly, 27*(3), 543–554. doi:10.1016/j.ecresq.2011.12.008
- Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., . . . Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558–580.
- Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential to learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.
- Ellefson, M. R., Treiman, R., & Kessler, B. (2009). Learning to label letters by sounds or names: A comparison of England and the United States. *Journal of Experimental Child Psychology, 102*, 323–341. doi:10.1016/j.jecp.2008.05.008
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Evans, M. A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds and phonological awareness: An examination of kindergarten children across letters and of letters across children. *Reading and Writing, 19*, 959–989.
- Fewster, S., & Macmillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149–156.
- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*(3), 315–342.
- Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revisited. *Journal of Literacy Research, 36*(1), 85–98.
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*, 216–227. doi:10.2307/1593674
- Fuchs, L. S., & Fuchs, D. (2001). Progress monitoring with letter-sound fluency: Technical data. In E. Kame'enui et al., *Reading First assessment committee report*. Unpublished document, available from author.

- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Karns, K. (1995). General educators' specialized adaptation for students with learning disabilities. *Exceptional Children*, *61*(5), 440–459.
- Gettinger, M., & Stoiber, K. C. (2012). Curriculum-based early literacy assessment and differentiated instruction with high-risk preschoolers. *Reading Psychology*, *33*(1–2), 11–46. doi:10.1080/02702711.2012.630605
- Good, R. H., Kaminski, R. A., Smith, S., Laimon, D., & Dill, S. (2001). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* (5th ed.). Eugene: Institute for Development of Educational Achievement, University of Oregon.
- Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review*, *34*(2), 184–201.
- Huang, F. L., & Invernizzi, M. A. (2012). The association of kindergarten entry age with early literacy outcomes. *Journal of Educational Research*, *105*(6), 431–441. doi:10.1080/00220671.2012.658456
- Hymel, S., LeMare, L., & McKee, W. (2011). The early development instrument: An examination of convergent and discriminant validity. *Social Indicators Research*, *103*(2), 267–282. doi:10.1007/s11205-011-9845-2
- Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research*, *36*(4), 479–500. doi:10.1207/s15548430jlr3604_3
- Invernizzi, M., Landrum, T. J., Teichman, A., & Townsend, M. (2010). Increased implementation of emergent literacy screening in pre-kindergarten. *Early Childhood Education Journal*, *37*(6), 437–446. doi:10.1007/s10643-009-03717
- Invernizzi, M., Sullivan, A., & Meier, J. D. (2001). *Phonological awareness literacy screening: Pre-kindergarten*. Charlottesville: University of Virginia.
- Istation. (2014). *ISEP Early Reading*. Retrieved from <http://www.istation.com/Assessment/ISIEarlyReading>
- Jackson, B., Larzelere, R., Clair, L. S., Corr, M., Fichter, C., & Egertson, H. (2006). The impact of HeadsUp! reading on early childhood educators' literacy practices and preschool children's literacy skills. *Early Childhood Research Quarterly*, *21*, 213–226.
- Jones, C., Clark, S., & Reutzel, D. R. (2013). Enhancing alphabet knowledge instruction: Research implications and practical strategies for early childhood educators. *Early Childhood Education Journal*, *41*, 81–89. doi:10.1007/s10643-012-0534-9
- Jones, C. D., & Reutzel, D. R. (2012). Enhanced alphabet knowledge instruction: Exploring a change of frequency, focus, and distributed cycles of review. *Reading Psychology*, *33*, 448–464. doi:10.1080/02702711.2010.545260
- Justice, L. M., Pence, K., Bowles, R. B., & Wiggins, A. (2006). An investigation of four hypotheses concerning the order by which 4-year-old children learn the alphabet letters. *Early Childhood Research Quarterly*, *21*, 374–389. doi:10.1016/j.ecresq.2006.07.010
- Kim, Y.-S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—a cross-classified multi-level model approach. *Journal of Educational Psychology*, *102*, 313–326. doi:10.1037/a0018449
- Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, *101*, 448–465.
- Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J. E. (2008). Letter knowledge predicts grade 4 reading fluency and reading comprehension. *Learning and Instruction*, *18*, 548–564.
- Levin, I., & Aram, D. (2005). Children's names contribute to early literacy: A linguistic and social perspective. In D. D. Ravid & H. B.-Z. Shyldkrot (Eds.), *Perspectives on language and language development* (pp. 219–239). Boston: Kluwer Academic.
- Lonigan, C. J. (2011). *Florida VPK Assessment Measures: Technical manual*. Tallahassee: Florida Department of Education Office of Early Learning.
- Lonigan, C. J., Allan, N. P., & Lerner, M. D. (2011). Assessment of preschool early literacy skills: Linking children's educational needs with empirically supported instructional activities. *Psychology in the Schools*, *48*, 488–501. doi:10.1002/pits.20569

- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*, 596–613. doi:10.1037//0012-1649.36.5.596
- Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, *24*, 305–337. doi:10.1007/s11145-009-9214-6
- Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2012). *Evaluating the components of an emergent literacy intervention for children at risk for reading difficulties*. Manuscript submitted for publication.
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy*. Austin, TX: Pro-Ed.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement*, *8*(4), 453–461.
- McBride-Chang, C. (1999). The abcs of the abcs: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly*, *45*, 285–308.
- Muthén, B. O. (1989). Using item-specific instructional information in achievement modeling. *Psychometrika*, *54*, 385–396. doi:10.1007/BF02294624
- Muthén, B. O., Kao, C.-f., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, *28*, 1–22. doi:10.1111/j.1745-3984.1991.tb00340.x
- Muthén, L. K., & Muthén, B. O. (2008). *Mplus* (version 5.2). Los Angeles: Author.
- National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC: Authors.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50–64.
- Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken School Readiness Assessment for predicting first grade readiness. *Psychology in the Schools*, *46*(5), 397–409. doi:10.1002/pits.20385
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*, 184–202.
- Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading*, *8*(1), 3–26. doi:10.1207/s1532799xssr0801_2
- Petscher, Y., Kim, Y.-S., & Foorman, B. R. (2011). The importance of predictive power in early screening assessments: Implications for placement in the response to intervention framework. *Assessment for Effective Intervention*, *36*(3), 158–166. doi:10.1177/1534508410396698
- Phillips, B. M., & Piasta, S. B. (2013). Phonological awareness and print knowledge: Key precursors and instructional targets to promote reading success. In T. Shanahan & C. J. Lonigan (Eds.), *Early childhood literacy: The National Early Literacy Panel and beyond* (pp. 95–116). Baltimore, MD: Brookes.
- Phillips, B. M., Piasta, S. B., Anthony, J. L., Lonigan, C. J., & Francis, D. J. (2012). IRTs of the ABCs: Children's letter name acquisition. *Journal of School Psychology*, *50*, 461–481. doi:10.1016/j.jsp.2012.05.002
- Piasta, S. B. (2006). *Acquisition of alphabetic knowledge: Examining letter- and child-level factors in a single, comprehensive model*. M.S. thesis, Florida State University, Tallahassee.
- Piasta, S. B. (2011). *Early learning standards relevant to alphabet knowledge*. [Data file compiled from all available state and national standards documents]. Columbus: The Ohio State University.
- Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks

- for predicting first-grade literacy achievement. *Journal of Educational Psychology*, **104**, 945–958. doi:10.1037/a0027757
- Piasta, S. B., Purpura, D. J., & Wagner, R. K. (2010). Fostering alphabet knowledge development: A comparison of two instructional approaches. *Reading & Writing*, **23**, 607–626. doi:10.1007/s11145-009-9174-x
- Piasta, S. B., & Wagner, R. K. (2010a). Developing emergent literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, **45**, 8–38.
- Piasta, S. B., & Wagner, R. K. (2010b). Learning letter names and sounds: Effects of instruction, letter type, and phonological processing skill. *Journal of Experimental Child Psychology*, **105**, 324–344. doi:10.1016/j.jecp.2009.12.008
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A. M., . . . Lyytinen, H. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, **48**, 923–931. doi:10.1111/j.1469-7610.2007.01763.x
- Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling: A multidisciplinary Journal*, **17**(2), 265–279. doi:10.1080/10705511003659417
- Reid, D., Hresko, W., & Hammill, D. (2001). *Test of Early Reading Ability* (3rd ed.). Austin, TX: PRO-ED.
- Reid, M. A., DiPerna, J. C., Morgan, P. L., & Lei, P. W. (2009). Reliability and validity evidence for the EARLI literacy probes. *Psychology in the Schools*, **46**(10), 1023–1035. doi:10.1002/pits.20441
- Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology*, **31**(3), 301–327. doi:10.1016/j.cedpsych.2005.10.001
- Roehrig, A. D., Duggar, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction identifying potential supports and challenges. *Remedial and Special Education*, **29**(6), 364–382.
- SAS. (2012). *SAS 9.3*. Cary, NC: SAS Institute.
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative study. *Journal of Educational Psychology*, **96**, 265–282. doi:10.1037/0022-0663.96.2.265
- Schatschneider, C., Petscher, Y., & Williams, K. M. (2008). How to evaluate a screening process: The vocabulary of screening and what educators need to know. In L. M. Justice & C. Vukelich (Eds.), *Achieving excellence in preschool literacy instruction* (pp. 304–316). New York: Guilford.
- Schumacher, R., Ewen, D., Hart, K., & Lombardi, J. (2005). *All together now: State experiences in using community-based child care to provide pre-kindergarten*. Washington, DC: Center for Law and Social Policy.
- Scott-Little, C., Kagan, S. L., & Frelow, V. S. (2006). Conceptualization of readiness and the content of early learning standards: The intersection of policy and research? *Early Childhood Research Quarterly*, **21**, 153–173. doi:10.1016/j.ecresq.2006.04.003
- Solari, E. J., Petscher, Y., & Folsom, J. S. (2012). Differentiating literacy growth of ELL students with LD from other high-risk subgroups and general education peers: Evidence from grades 3–10. *Journal of Learning Disabilities*. doi:10.1177/0022219412463435
- Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2002). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education*, **36**(4), 223–233. doi:10.1177/002246690303600403
- Spencer, E. J., Spencer, T. D., Goldstein, H., & Schneider, N. (2013). Identifying early literacy learning needs: Implications for child outcome standards and assessment systems. In T. Shanahan & C. J. Lonigan (Eds.), *Early childhood literacy: The National Early Literacy Panel and beyond* (pp. 45–70). Baltimore, MD: Brookes.
- Spitler, M. E. (2001). Life-long preparation of early childhood teachers: A professional development system for the coming millennium. *Journal of Early Childhood Teacher Education*, **22**(1), 21–28.

- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*(8), 795–819. doi:10.1002/pits.20113
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 73–140). Hillsdale, NJ: Erlbaum.
- Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities*, *43*, 308–321. doi:10.1177/0022219410369096
- Torppa, M., Poikkeus, A.-M., Laakso, M.-L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology*, *42*, 1128–1142. doi:10.1037/0012-1649.42.6.1128
- Torquati, J. C., Raikes, H., & Huddleston-Casas, C. A. (2007). Teacher education, motivation, compensation, workplace support, and links to quality of center-based child care and teachers' intention to stay in the early childhood profession. *Early Childhood Research Quarterly*, *22*(2), 261–275. doi:10.1016/j.ecresq.2007.03.004
- Treiman, R., & Broderick, V. (1998). What's in a name: Children's knowledge about the letters in their own names. *Journal of Experimental Child Psychology*, *70*, 97–116. doi:10.1006/jecp.1998.2448
- Treiman, R., Kessler, B., & Pollo, T. C. (2006). Learning about the letter name subset of the vocabulary: Evidence from U.S. and Brazilian preschoolers. *Applied Psycholinguistics*, *27*, 211–227. doi:10.1017/S0142716406060255
- Treiman, R., Pennington, B. F., Shriberg, L. D., & Boada, R. (2008). Which children benefit from letter names in learning letter sounds? *Cognition*, *106*, 1322–1338.
- Treiman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development*, *69*, 1524–1540. doi:10.2307/1132130
- Treiman, R., Weatherston, S., & Berch, D. (1994). The role of letter names in children's learning of phoneme-grapheme relations. *Applied Psycholinguistics*, *15*, 97–122.
- U.S. Department of Health and Human Services, Administration for Children and Families, Office of Head Start. (2010). *The Head Start child development and early learning framework: Promoting positive outcomes in early childhood programs serving children 3–5 years old*. Washington, DC: Author.
- VanDerHeyden, A. M., Snyder, P. A., Broussard, C., & Ramsdell, K. (2008). Measuring response to early literacy intervention with preschoolers at risk. *Topics in Early Childhood Special Education*, *27*(4), 232–249.
- Walker, D., Carta, J. J., Greenwood, C. R., & Buzhardt, J. F. (2008). The use of individual growth and developmental indicators for progress monitoring and intervention decision making in early education. *Exceptionality*, *16*(1), 33–47. doi:10.1080/09362830701796784
- Whitehurst, G. J. (2001). *The NCLD Get Ready to Read screening tool technical report*. Retrieved from www.getreadytoread.org/pdf/TechnicalReport.pdf
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—revised/NU*. Circle Pines, MN: American Guidance Service.