Educational Research and Evaluation: An International Journal on Theory and Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/nere20

Socioeconomic and gender group differences in early literacy skills: a multiple-group confirmatory factor analysis approach

Julia Ai Cheng Lee[a] & Stephanie Al Otaiba[b]
[a] Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Sarawak, Malaysia
[b] Annette Caldwell Simmons School of Education & Human Development, Southern Methodist University, Dallas, TX, USA
Published online: 18 Feb 2015.

To cite this article: Julia Ai Cheng Lee & Stephanie Al Otaiba (2015) Socioeconomic and gender group differences in early literacy skills: a multiple-group confirmatory factor analysis approach, Educational Research and Evaluation: An International Journal on Theory and Practice, 21:1, 40-59, DOI: 10.1080/13803611.2015.1010545

To link to this article: http://dx.doi.org/10.1080/13803611.2015.1010545

Educational Research and Evaluation, 2015
Vol. 21, No. 1, 40–59, http://dx.doi.org/10.1080/13803611.2015.1010545

Socioeconomic and gender group differences in early literacy skills: a multiple-group confirmatory factor analysis approach

Julia Ai Cheng Lee[a*] and Stephanie Al Otaiba[b]

[a]*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Sarawak, Malaysia;* [b]*Annette Caldwell Simmons School of Education & Human Development, Southern Methodist University, Dallas, TX, USA*

(*Received 28 April 2014; final version received 7 January 2015*)

Socioeconomic status and gender are important demographic variables that strongly relate to academic achievement. This study examined the early literacy skills differences between 4 sociodemographic groups, namely, boys ineligible for free or reduced-price lunch (FRL), girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL. Data on kindergarteners (N = 462) were analysed using multiple-group confirmatory factory analysis. Early literacy skill differences between boys and girls are more nuanced than previously reported; subsidy status and gender interact. Both boys and girls from high-poverty households performed significantly lower than the girls from low-poverty households in alphabet knowledge, phonological awareness, and spelling. There were gender gaps, with a female advantage, among children from high-poverty households in alphabet knowledge and spelling and among children from low-poverty households in alphabet knowledge. These results highlight the importance of employing methodologically sound techniques to ascertain group differences in componential early literacy skills.

Keywords: early literacy skills; gender; socioeconomic status; multiple-group confirmatory factor analysis; kindergarten

Introduction

One of the most pressing issues in the United States education system is the achievement gap between socioeconomically disadvantaged students and their counterparts (e.g., Brooks-Gunn & Duncan, 1997; Entwisle, Alexander, & Olson, 2007). For example, the nation's report card on reading reveals substantial disparities between the socioeconomic status (SES) groups where a higher percentage of students from low-SES backgrounds in Grades 4 and 8 scored lower in reading than their counterparts (Lee, Grigg, & Donahue, 2007). Additionally, the achievement gap has also been linked to gender differences (Entwisle et al., 2007).

The achievement gap begins early (Arnold & Doctoroff, 2003; Chatterji, 2006; West, Denton, & Germino-Hausken, 2000). Mounting evidence from the national Early Childhood Longitudinal Study Kindergarten (ECLS-K) data suggests that SES predicts the entry-level literacy skills and early reading growth (Chatterji, 2006; McCoach, O'Connell,

*Corresponding author. Email: aclee@fcs.unimas.my

Reis, & Levitt, 2006). For instance, Chatterji's (2006) analysis of the ECLS-K data demonstrates that children from low-SES homes scored .5 standard deviation (*SD*) units lower than children from high-SES homes in kindergarten; by the end of first grade, the magnitude of this gap increased to between .61 *SD* and 1 *SD* units. This is supported by the latest ECLS-2011 data (Mulligan, Hastedt, & McCarroll, 2012). In the United States, a frequently used proxy for SES is whether students receive federally funded free or reduced price lunch. Families with low incomes (under $21,600) can apply for this service, and their children receive breakfast and lunch at free or significantly reduced prices.

For many such children, literacy-related difficulties are due to experiential-instructional inadequacies such as the lack of exposure to print or instructional resources and/or poor quality teaching (e.g., Fletcher et al., 2002; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Students who begin their academic careers as poor readers lag behind their peers (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). Consequently, the reading gap widens over time (e.g., McCoach et al., 2006), spawning other accompanying problems such as reading difficulties, poor motivation, frustration, dropping out of school, and restricted employment opportunities (Arnold & Doctoroff, 2003; Brooks-Gunn & Duncan, 1997; Fletcher et al., 2002).

Contrary to the achievement gap related to SES, the findings regarding the onset of gender gap in literacy achievement are inconclusive. Some studies reported that young girls and boys do not differ significantly in early literacy skills (e.g., Entwisle et al., 2007; Harper & Pelletier, 2008; Matthews, Ponitz, & Morrison, 2009). For instance, no gender differences on letter-word identification, expressive vocabulary, and sound awareness was found in Matthews et al.'s (2009) study, which was comprised of proportionate male-female kindergarteners (48% males), predominantly White (83%) children, and parents (i.e., 40%) with master's degrees.

Conversely, analyses based on national data (i.e., ECLS-K) report that girls outperformed boys in reading at kindergarten entry, learned marginally more than boys during the academic year (Chatterji, 2006; Ready, LoGerfo, Burkam, & Lee, 2005; West et al., 2000), and grew more rapidly than boys (McCoach et al., 2006). One ECLS-K analysis demonstrated that the gender gap, with a female advantage, increased from .17 *SD* units in kindergarten to .31 *SD* units in first grade reading (Chatterji, 2006). Noteworthy is Chatterji's (2006) finding that early literacy skills in kindergarten (i.e., print familiarity, letter recognition, initial and final sounds, rhyming sounds, word recognition, receptive vocabulary, listening comprehension, and comprehension of words in context) were more strongly related with poverty than they were with ethnicity or gender. Furthermore, Chatterji (2006) and Entwisle et al. (2007) found no significant child-level interactions between poverty and gender among first graders, but the poverty-gender interaction emerges only in Grade 2 onwards (Entwisle et al., 2007).

One reason for the inconclusive results could be related to the issue of measurement non-invariance where indicators that measure the constructs between groups are dissimilar (Byrne & Watkins, 2003; Kline, 2011). For instance, instruments (e.g., adolescent depression inventory) that have similar outcomes when tested individually may not function equivalently across groups or cultures (Byrne & Watkins, 2003). Thus, the difference between groups may be due to the construct conceptualization rather than a true difference between groups. Conversely, when there is measurement invariance, the instrument is measuring one group similarly to the other (Kline, 2011). Hence, determining measurement invariance is important before making group mean comparisons. Despite its importance, only one early literacy study has tested for measurement invariance (Townsend & Konold, 2010). Townsend and Konold (2010) reported that the emergent literacy measures comprising alphabet knowledge, phonological awareness, and print concept were generally invariant for both male and female students ($N = 4{,}518$). However, group mean differences in early literacy skills were not examined.

Given the limited research on the differential early literacy achievements between SES and gender groups among kindergarteners, the protracted stability of literacy skills in individuals (Francis et al., 1996), and the importance of research on the achievement gaps (Gamoran, 2007), the primary aim of the present study was to ascertain whether measurement equivalence of the early literacy measures between groups exist; after establishing measurement equivalence, the second aim was to examine the group mean differences namely between four groups: boys ineligible for free or reduced-price lunch (FRL), girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL. Data analysis was conducted using multiple-group confirmatory factor analysis (MGCFA; see Data analysis section for further details) to determine the (un)biased effects of the measures between groups.

Theoretical framework

Componential skills of early literacy

According to Snow (2006), literacy is "the product of an array of componential skills, all of which are necessary to high-level performance" (p. 277). A recent meta-analysis (National Early Literacy Panel [NELP], 2008) reported that the following six components of early literacy skills consistently predict later conventional literacy skills in moderate or strong correlations: alphabet knowledge, phonological awareness, rapid automatized naming of letters or digits, rapid automatized naming of objects or colors, writing and name writing, and phonological memory. Similarly, Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) reported that letter-name and letter-sound knowledge, phonological awareness, and naming speed are the most important kindergarten predictors of reading skills. For the purposes of developing a theoretical framework of early literacy skills, we borrowed Snow's (2006) definition of componential skills and the work of the NELP and Schatschneider et al. on the early literacy predictors of conventional literacy skills. We selected the following early literacy constructs for this study: alphabet knowledge (i.e., letter-name fluency and letter-sound fluency), phonological awareness (i.e., blending and elision), and spelling (i.e., real words and pseudowords). To date, these componential early literacy skills have not been examined in tandem in a multiple-group analysis comprising both SES and gender groups.

Alphabet knowledge

Adams (1990) suggests that naming the alphabetic letters with accuracy and speed is an index of automatic letter recognition. Accuracy and speed characterize the fluency construct (Speece, Mills, Ritchey, & Hillman, 2003). Letter-name fluency measured in kindergarten uniquely contributes to word reading ($r = .69$); it is able to identify 86% of the children who would eventually end up being poor readers in Grade 1 (Speece et al., 2003). Speece et al.'s (2003) finding corroborates with O'Connor and Jenkins' (1999) finding that the timed letter-knowledge task is a strong discriminator of good readers versus poor readers. Given that the predictive validity of letter-name fluency measured at the end of kindergarten is most optimal in relation to the substantial reduction in floor effects from the fall to the spring of kindergarten (Catts, Petscher, Schatschneider, Bridges, & Mendoza 2009), letter-name fluency is a more viable measure than letter-name knowledge at the end of kindergarten.

Contrary to letter-name fluency, there are fewer studies that examine the contribution of letter-sound fluency in early literacy acquisition (Al Otaiba et al., 2010; Fuchs et al., 2001; Ritchey & Speece, 2006). Ritchey and Speece (2006) demonstrated that letter-sound fluency contributes significant unique variance in word reading and spelling measures at the end of kindergarten above and beyond letter-sound accuracy. In addition, both accuracy and fluency measures of letter sound accounted for 53.9% unique variance in predicting spelling, while the accuracy and fluency measures of letter name contributed 40.6% (Ritchey & Speece, 2006). Similarly, Al Otaiba et al. (2010) and Ritchey (2008) have shown that letter-sound fluency is a strong predictor of kindergarten spelling with $r = .65$ and $r = .81$, respectively. Ritchey and Speece also reported that higher growth rates in letter-sound fluency were related to better performance in reading.

Phonological awareness

Phonological awareness encompasses the ability to detect, manipulate, or analyse sounds in spoken language in varying complexities such as words, syllables, and phonemes (Blachman, 2000; NELP, 2008). Numerous research studies (e.g., Byrne & Fielding-Barnsley, 1993; Schatschneider et al., 2004; Wagner et al., 1997), reviews (e.g., Blachman, 2000), and policy-motivated syntheses (NELP, 2008; National Reading Panel [NRP], 2000) have shown that children's phonological awareness is crucial for the acquisition of subsequent literacy skills. NELP's (2008) synthesis suggests that phonological awareness in kindergarten or earlier is moderately related to decoding ($r = .40$), spelling ($r = .40$), and reading comprehension ($r = .44$). Problems in acquiring phonological awareness in the early stages of literacy acquisition can result in reading difficulties or reading disabilities, which in turn has a cumulative effect in subsequent grades (Blachman, 2000; MacDonald & Cornwall, 1995; Wagner et al., 1997).

Spelling

The relations between phonological awareness, alphabet knowledge, and spelling are well established (e.g., NRP, 2000; Santoro, Coyne, & Simmons, 2006). Letter-name knowledge, letter-sound knowledge, and phonological awareness impact children's ability to write letters and to spell (Ritchey, 2008). Alphabet knowledge has been reported to be the strongest predictor of spelling ($r = .54$; NELP, 2008). In addition, the relation between spelling and later reading success is also well supported (e.g., Ehri & Wilce, 1987; NELP, 2008). For instance, NELP (2008) reported that spelling measured in kindergarten or earlier is strongly correlated with conventional literacy skills such as decoding ($r = .60$) and spelling ($r = .78$).

The present study

SES is a persistent correlate of early literacy outcome (e.g., Chatterji, 2006). However, findings on gender gap in literacy skills are more inconclusive (e.g., Entwisle et al., 2007; Matthews et al., 2009). Additionally, few studies have examined these two sociodemographic groups simultaneously in the same model. Given these reasons and the importance of establishing measurement invariance in early literacy skill differences between sociodemographic groups (Byrne & Watkins, 2003), we determined the measurement invariance in early literacy skill differences between boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL. We used multiple-group confirmatory factor analysis to achieve this purpose. We also examined the latent means differences once the measurement invariance was established. In summary, we addressed the following questions:

(1) Are the indicators of alphabet knowledge, phonological awareness, and spelling measured in the same way for boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL? Are the early literacy latent constructs generalizable across these groups?
(2) Do the interrelations of the latent factors of early literacy skills vary across the four groups namely, boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL? Are the factor structures of the latent factors more strongly correlated in one group than another?
(3) What are the factor mean differences in early literacy skills between boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL?

Method

Participants

This study used existing data collected in a larger study of response to early literacy instruction that was conducted with funding from the National Institute of Child Health and Human Development. In this larger study, 10 schools in one school district in a southeastern US city were recruited with help from district personnel to include a sample of diverse sample of students. Specifically, the study focused on improving reading outcomes for students who were considered at risk for reading difficulties because of their socioeconomic status or minority status. The present study involved a sample of 462 kindergarteners from 10 public elementary schools. Parental consents had been obtained during a larger study (Al Otaiba et al., 2011). The demographics of the participants were consistent with the recruitment for the larger study in terms of demographics: mean age at initial testing (M = 5.56 years, SD = .35); male (54.5%); Black (58.9%); White (31.4%), American Indian, Alaska Native, Asian, Native Hawaiian or other Pacific Islander (1.9%); Multiracial or other (7.8%); and FRL eligibility (69%). The Verbal IQ (M = 90.83, SD = 14.41) and non-verbal IQ (M = 91.61, SD = 11.41) was assessed using the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2004). The normal range of IQ has a mean of 100 and standard deviation of 15. About 6.9% of the sample had been retained in kindergarten; student absenteeism was an average of 11 days in the academic year. All the students with special needs (speech/language impairment, 2.4%; specific learning disabilities, 1.8%; and developmental delay, 1.8%) were included in the present study. There were 2.6% of the children (n = 12) with limited English proficiency.

Context of the larger study

Within the larger study, schools were matched on the percent of children who participated in FRL and in whether schools received Title 1 funding (which is a federally funded programme to support schools that serve a high proportion of students with low SES) and were randomized into either treatment or wait-list control conditions during the 1st year of the study (2007–2008). Teachers in the treatment group received training on a classwide early literacy instructional programme that helped teachers better differentiate instruction.

The programme was Individualized Student Instruction for Kindergarten (ISI-K), which included Assessment to Instruction (A2i) software, ongoing teacher professional development, and classroom support for the teachers (Al Otaiba et al., 2011). Teachers also received professional development and bi-weekly in-class support. The wait-list control group received more limited professional development, which involved a summer workshop on response to intervention, individualized instruction, and materials from the Florida Center for Reading Research (http://www.fcrr.org). In the 2nd year of the study (2008–2009), all kindergarten teachers from both conditions received the ISI-K training. Thus, all kindergarten teachers in the present study had received treatment; some received 2 years of treatment and others received 1 year of treatment.

Measures

The measures used in the larger study were also used for the present study and included two alphabetic fluency measures, two phonological awareness measures, and two spelling measures.

Alphabetic fluency measure

The Letter Naming Fluency task of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measure (LNF; Good & Kaminski, 2002) assessed the student's ability to name letters in 1 minute. Each probe consists of 26 randomly ordered uppercase and lowercase letters ordered in an array of 10 by 11 items. The number of correctly named letters was scored. No points were allocated for unnamed letter names after 3 seconds. If the letter sound rather than the letter name was provided, the test administrator requested for the letter name instead of the letter sound. The possible range for the LNF is 0–110. The alternate-form reliability is .99.

The AIMSweb Letter-Sound Fluency (LSF; Shinn & Shinn, 2004) subtest assessed the students' ability to say letter sounds in 1 minute. Each probe consisted of an array of 10 by 10 lower-case letters. Testing was discontinued if the child could not produce any correct sounds for the first 10 letters. The possible range for the LSF is 0–100. The alternate-form reliability is .90.

Phonological awareness

Two measures from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) were used. The Blending Words subtest assessed the students' ability to orally blend larger to smaller units of words (i.e., compound words, syllables, onset rimes, and phonemes). For example, students were required to blend /c/ /at/ and to respond "cat." The possible range for the Blending Words subtest is 0–20. The Elision subtest assessed the students' ability to delete words, syllables, or phoneme(s) from orally presented words. For example, students were required to say "meat" without /m/. The possible range for the Elision subtest is 0–20. The test-retest reliability for both subtests is .88.

Spelling

The spelling measure included six decodable real words (*dog*, *man*, *plug*, *went*, *limp*, and *tree*) and four decodable pseudowords (*ig*, *sut*, *frot*, and *yilt;* Byrne & Fielding-Barnsley, 1993). Research assistants provided lined answer sheets and instructions saying:

> I would like you to spell some words. Some are real, and some are made-up words. If you don't know how to spell a word, sound it out and do your best. First, I am going to say the word, then I will use it in a sentence, and then I will say the word one more time. Remember to write the word next to the correct number on your answer sheet. Ready begin.

Each word to be spelled was read, then a sentence with the word was read, and finally, the word was repeated (e.g., "dog. I took my dog to the park. Dog."). The decodable pseudowords were repeated three times (e.g., "Next word: sut, sut, sut"). The possible range for the decodable words is 0–36, while the possible range for the pseudowords is 0–24. The correlation of the spelling scores between the decodable real words and the decodable pseudowords was high (r = .75).

Demographics: SES and gender

FRL was the proxy measure for SES in this study because only children whose families meet the US federal criteria participate. The four grouping variables between FRL and gender were coded as follows: (0 = boys ineligible for FRL, n = 78; 1 = girls ineligible for FRL, n = 65; 2 = boys eligible for FRL, n = 174; 3 = girls eligible for FRL, n = 145).

Procedure

This is secondary analysis. All measures, except the spelling task, had been individually administered. The spelling task was group administered. Scores from the spring of kindergarten were used in the present study.

The students' spelling was analysed using Tangel and Blachman's (1992) developmental spelling rubric (see Al Otaiba et al., 2010, for details on the scoring procedure). The first author coded the entire corpus of spelled words, and two research assistants who had been trained to code the spelling errors in a related study (Al Otaiba et al., 2010) each coded 50% of the dataset. The inter-rater agreement based on percent agreement on the individual words ranged from 85.1% to 93%; the mean percent agreement of the entire data set was 89.9%. The range of the Cohen's kappa coefficient on the individual words was .76–.90; the mean kappa was .84. Discrepancies were resolved through discussion; the final agreement was 100%.

Preliminary analysis

As a preliminary analysis, we compared the mean differences of the early literacy indicators assessed in spring for kindergarteners whose schools had been cluster-randomized to either 2 years of treatment or to 1 year of treatment using multiple t tests. The results were not significantly different after adjusting for multiple comparisons to control the Type 1 error rate (α = .05) (Benjamini & Hochberg, 1995).

Data analysis

Mplus 6.1 (Muthén & Muthén, 1998–2010) was employed for the MGCFA with mean structures to analyse the data with four FRL-gender sociodemographic groups. The indicators for the constructs were as follows: (a) alphabet knowledge: LNF and LSF; (b) phonological awareness: CTOPP Blending and Elision; and (c) spelling: decodable real words and decodable pseudowords. A stepwise approach was used to assess measurement invariance, structural invariance, and factor means (Brown, 2006; Thompson & Green, 2006). Figure 1 presents the final MGCFA model.

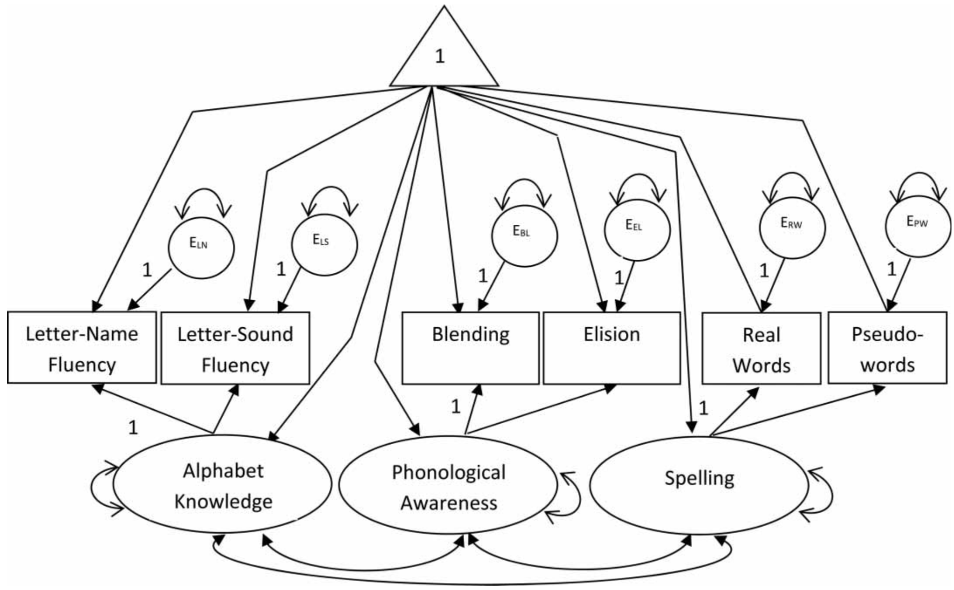

Figure 1. The final measurement model of alphabet knowledge, phonological awareness, and spelling with mean structures evaluated across samples of boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL.

Evaluation of measurement invariance

Step 1 of the test of measurement invariance establishes a general confirmatory factor analysis (CFA) model that fits the individual groups (i.e., boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL). If the model-data fit is adequate for all the groups, then Step 2, which establishes a baseline model without cross-group constraints, is conducted, or else no further analysis is pursued. If the model-data fit is adequate, then Step 3, which is a test of complete measurement invariance and equality of factor loadings and intercepts for the overall group, is conducted. If the model-data fit is inadequate for the baseline model, then there will be no further analysis. A chi-square difference test and a comparative fit index (CFI) difference test are conducted to compare the measurement invariance model (i.e., constrained model) and the baseline model (i.e., unconstrained model). If the chi-square difference test between the two models is not significant, then the invariance of factor loadings and intercepts (i.e., measurement invariance) is supported. However, if the chi-square difference test is significant, the invariance of loadings and intercepts is not supported. A cutpoint of CFI difference test of less than .01 was chosen to decide whether there was a substantial decrease in model fit between the baseline model and the constrained model (Cheung & Rensvold, 2002; Hirschfeld & von Brachel, 2014).

Evaluation of structural invariance and latent means

In this study, we established cross-group constraints between the four groups by fixing the factor covariance to be the same. Because equality of both factor loadings and intercepts was found, the latent means between groups could be compared (Brown, 2006). Next, an omnibus test of null hypothesis was conducted by fixing each group's factor means to zero (Thompson & Green, 2006). If the null hypothesis is not rejected, then the factor means between groups are equivalent; no further analysis is required. If the null hypothesis is rejected, one or more factor means differ between groups. One of the groups is fixed as the reference group (i.e., the factor means for that group is fixed to zero), while other factor means are freely estimated. Subsequently, multiple steps are undertaken to fix another factor means to be zero while other factor means are freely estimated to obtain the mean differences between groups per construct (Thompson & Green, 2006).

Results

Descriptive statistics and selection of estimation methods

The descriptive statistics and correlations for the early literacy outcome measures are presented in Table 1. Across all measures, the means of the groups ineligible for FRL were higher than the means of groups eligible for FRL. Similarly, the means of girls were also higher than the means of boys for both subsidized and unsubsidized groups. The standard deviations for the groups ineligible for FRL were smaller than the standard deviations of the groups eligible for FRL on the LNF, LSF, and CTOPP Blending Word measures for both boys and girls. In general, there were stronger correlations for within-construct indicators than there were for between-construct indicators. For instance, across the four sociodemographic groups, measures representing alphabet knowledge were more strongly correlated with each other (i.e., an average of r = .74) than they were with other measures; spelling decodable real words and spelling decodable pseudowords were more strongly correlated (i.e., an average of r = .80) than they were with other measures. Within each sociodemographic group, all the measures were significantly correlated, except the correlation between LNF and CTOPP Elision for the girls ineligible for FRL.

We also screened the dataset for non-normality, an assumption of maximum likelihood (ML), by visually inspecting the frequency histograms and the SPSS skewness and kurtoses indexes. There was a moderate departure from normality (i.e., skew < 2, kurtosis < 7; Curran, West, & Finch, 1996; Finney & Distefano, 2006) for the following measures: CTOPP Blending and Elision, spelling real words, and spelling pseudowords. Based on both the moderately non-normal distribution and small sample size per group (i.e., $n \leq 250$), we chose the Satorra-Bentler (S-B) scaling method as the estimator (Curran et al., 1996; Finney & DiStefano, 2006). Missing data was coded as "999". There were a total of nine missing data patterns across all four groups. The missing data patterns across groups were proportionately similar, which suggests that the missing data were missing completely at random. Thus, full information maximum likelihood was used (Kline, 2011).

Multiple-group confirmatory factor analysis

We answered Research question 1 by evaluating the measurement invariance of the latent variables and their respective observed variables; Research question 2 by evaluating the structural invariance; and Research question 3 by evaluating the equality of latent means.

Evaluation of measurement invariance

As a first step in evaluating measurement invariance, the model for each of the four groups (i.e., boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL) was individually fitted. Our judgment regarding good model-data fit indices were indicated by the following: (a) non-significant χ^2 values; (b) comparative fit index

Table 1. Descriptive statistics and correlations of the measures in each group and the total dataset.

Observed variables	1.	2.	3.	4.	5.	6.		
	Boys ineligible for FRL and boys eligible for FRL						Boys ineligible for FRL[a]	
							M	*SD*
1. Letter-name fluency	–	.66	.53	.40	.50	.47	45.26	14.40
2. Letter-sound fluency	.75	–	.60	.49	.53	.54	40.45	16.74
3. Blending words	.52	.55	–	.56	.52	.50	10.91	4.27
4. Elision	.37	.40	.67	–	.53	.56	6.51	4.48
5. Spelling (real)	.49	.54	.67	.60	–	.83	23.88	8.58
6. Spelling (pseudo)	.54	.54	.60	.56	.77	–	14.12	6.66
Boys eligible for FRL [c] *M*	41.32	34.88	9.08	4.45	19.24	11.07		
SD	17.20	18.00	4.36	3.17	8.02	6.22		
	Girls ineligible for FRL and girls eligible for FRL						Girls ineligible for FRL[b]	
							M	*SD*
1. Letter-name fluency	–	.63	.41	.24[n]	.38	.33	52.65	15.18
2. Letter-sound fluency	.77	–	.54	.34	.38	.38	45.57	14.20
3. Blending words	.47	.49	–	.53	.58	.61	12.06	3.67
4. Elision	.49	.42	.61	–	.60	.67	6.55	3.97
5. Spelling (real)	.56	.54	.61	.55	–	.82	25.00	7.33
6. Spelling (pseudo)	.47	.51	.63	.56	.73	–	15.60	6.06
Girls eligible for FRL [d] *M*	47.56	39.81	9.55	4.87	21.30	12.28		
SD	19.93	17.78	4.03	3.53	6.63	5.40		
	Total[e] – boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL							
							M	*SD*
1. Letter-name fluency	–						45.31	17.89
2. Letter-sound fluency	.74	–					38.79	17.57
3. Blending words	.50	.55	–				9.96	4.24
4. Elision	.40	.43	.62	–			5.20	3.74
5. Spelling (real)	.50	.53	.63	.59	–		21.51	7.95
6. Spelling (pseudo)	.48	.52	.61	.60	.80	–	12.64	6.24

Note: All values were significant at $p < .01$ except those marked [n]; FRL = free or reduced-price lunch; [a]$n = 78$; [b]$n = 65$; [c]$n = 174$; [d]$n = 145$; [e]total dataset $N = 462$; values above the diagonal are for children who were ineligible for FRL and the values below the diagonal are for children who were eligible for FRL.

(CFI) values greater than .90 (Hu & Bentler, 1999); (c) Tucker-Lewis index (TLI; Tucker & Lewis, 1973) approaching 1.0; (d) root mean square error of approximation (RMSEA) less than or equal to .08 (Browne & Cudeck, 1993); and (e) standardized root mean square residual (SRMR) values less than .08 or .10 (Hu & Bentler, 1999).

The models of each of the four groups converged to an admissible solution. The model-data fits were not statistically significant across all four groups. Table 2 presents the model-data fit indices for the stepwise approach undertaken. The model-data fit for the boys ineligible for FRL was: χ^2 (6) = 4.94, $p = .55$, CFI = 1.000, TLI = 1.000, RMSEA = 0, 90% confidence interval (CI) = .00 to .13, and SRMR = .02; the model-data fit for the girls ineligible for FRL was: χ^2 (6) = 6.76, *p* = .34, CFI = .996, TLI = 0.990, RMSEA = .04, 90% CI = .00 to .17, and SRMR = .04; the model-data fit for boys eligible for FRL was: χ^2 (6) = 9.30, *p* = .16, CFI = .994, TLI = 0.985, RMSEA = .06, 90% CI = .00 to .12, and SRMR = .02; and the model-data fit for the girls eligible for FRL: χ^2 (6) = 6.79, *p* = .34, CFI = .998, TLI = 0.995, RMSEA = .03, 90% CI = .00 to .12, and SRMR = .02. These results suggest that the model-data fits for all four groups were adequate.

Subsequently, a baseline model comprising all the groups with no constraints on the loadings was derived in the second step of the stepwise approach. The baseline model converged to an admissible solution. The model-data fit was not significant: χ^2 (33) = 36.21, *p* = .32, CFI = .998, TLI = 0.996, RMSEA = .03, 90% CI = .00 to .08, and SRMR = .04. The fit indices revealed very good model-data fit.

In the third step, a test of complete measurement invariance on factor loadings (i.e., constrained model) for the overall group was conducted. The model-data fit for the measurement invariance model was excellent: χ^2 (42) = 52.41, *p* = .13, CFI = .993, TLI = 0.990, RMSEA = .05, 90% CI = .00 to .08, and SRMR = .05. The Satorra-Bentler scaled chi-square difference test between the final measurement model and the baseline model was not significant (Δ S-B χ^2 = 16.12, Δ *df* = 9), which suggests that complete measurement invariance was established. Similarly, the CFI difference test (Δ CFI) between the baseline model and the constrained model was less than .01. The parameter estimates for the complete measurement invariance model with constrained factor loadings are reported in Table 3.

For the measurement invariance model with constrained factor loadings, we report the unstandardized estimates when comparing across groups because the unstandardized factor loadings are set to equality across groups (Kline, 2011). Conversely, for within-group comparison, we report the standardized estimates (Kline, 2011). The factor variances of alphabet knowledge (i.e., LNF and LSF) across both gender groups were substantially larger than the participants who were not eligible for FRL. For example, the mean unstandardized variance for the boys and girls eligible for FRL was 238.62 but 142.04 for the boys and girls ineligible for FRL. The variance for the boys ineligible for FRL on spelling was larger than the other three groups.

The standardized loadings on most of the factors were very high (i.e., above .80) except for the factor loading on CTOPP Elision for two groups who were not eligible for

Table 2. Model-data fit indices of the comparisons between models.

Model	χ^2	*p*	*df*	CFI	TLI	RMSEA (90% CI)	SRMR
1. Single group solutions							
Boys ineligible for FRL (*n* = 78)	4.94	.55	6	1.000	1.000	.00 (.00–.13)	.02
Girls ineligible for FRL (*n* = 65)	6.76	.34	6	.996	0.990	.04 (.00–.17)	.04
Boys eligible for FRL (*n* = 174)	9.30	.16	6	.994	0.985	.06 (.00–.12)	.02
Girls eligible for FRL (*n* = 145)	6.79	.34	6	.998	0.995	.03 (.00–.12)	.02
2. Baseline model	36.21	.32	33	.998	0.996	.03 (.00–.08)	.04
3. Constrained Model (factor loadings)	52.41	.13	42	.993	0.990	.05 (.00–.08)	.05
4. Cross-group Constraint Model (factor loadings and covariance)	61.92	.14	51	.992	0.991	.04 (.00–.08)	.09

Note: *N* = 462; All models were not statistically significant, *p* > .05; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; CI = confidence interval for RMSEA; SRMR = standardized root mean square residual.

Table 3. Robust maximum likelihood parameter estimates for the complete measurement invariance model across the four groups.

	Boys ineligible for FRL			Girls ineligible for FRL			Boys eligible for FRL			Girls eligible for FRL		
Parameter	Unstd.	*SE*	Std.	Unstd.	*SE*	Std.	Unstd.	*SE*	Std.	Unstd.	*SE*	Std.
						Equality-constrained estimates						
Factor variances and covariances												
AK	152.07	34.94	1.00	132.02	28.12	1.00	223.69	34.26	1.00	253.54	43.68	1.00
PA	13.00	2.76	1.00	8.84	2.53	1.00	12.63	1.73	1.00	11.31	1.90	1.00
SP	61.17	10.29	1.00	45.98	9.66	1.00	49.62	5.71	1.00	33.96	5.42	1.00
AK with PA	38.02	7.91	.86	24.61	6.61	.72	35.38	5.73	.67	35.20	6.24	.66
PA with SP	21.76	4.59	.77	18.71	4.44	.93	21.23	2.64	.85	17.27	2.71	.88
AK with SP	68.80	16.66	.71	37.56	12.65	.48	75.88	10.72	.72	64.24	11.73	.69
Factor loadings												
AK→ LN	1.00	.00	.82	1.00	.00	.74	1.00	.00	.86	1.00	.00	.83
AK → LS	1.04	.06	.80	1.04	.06	.86	1.04	.06	.87	1.04	.06	.92
PA→ BL	1.00	.00	.82	1.00	.00	.78	1.00	.00	.84	1.00	.00	.83
PA → EL	.75	.55	.63	.75	.55	.61	.75	.55	.80	.75	.55	.73
SP → RW	1.00	.00	.91	1.00	.00	.90	1.00	.00	.89	1.00	.00	.87
SP → PW	.78	.03	.91	.78	.03	.91	.78	.03	.87	.78	.03	.85
Measurement error variances												
E_{LN}	76.40	21.95	.33	106.95	22.87	.45	78.32	14.43	.26	117.18	21.43	.32
E_{LS}	90.69	20.52	.36	50.03	24.90	.26	78.16	17.77	.24	48.72	18.17	.15
E_{BL}	6.33	2.20	.33	5.74	1.42	.39	5.21	.76	.29	5.16	1.14	.31
E_{EL}	11.01	2.73	.60	8.56	1.94	.63	3.97	.77	.36	5.68	1.29	.47
E_{RW}	12.40	4.76	.17	10.71	3.93	.19	13.42	2.68	.21	11.20	2.05	.25
E_{PW}	7.60	2.84	.17	6.14	1.79	.18	9.73	1.66	.24	8.09	1.40	.28

Note: AK = alphabet knowledge; LN = DIBELS Letter Name Fluency; LS = AimsWeb Letter Sound Fluency; PA = phonological awareness; BL = CTOPP Blending; EL = CTOPP Elision; SP = spelling; RW = decodable real words; PW = pseudowords; FRL = free or reduced-price lunch; *SE* = standard error; unstd. = unstandardized; std. = standardized. All parameter estimates, $p < .01$.

FRL (i.e., boys ineligible for FRL and girls ineligible for FRL), which was between .61 and .63. The lower factor loadings for CTOPP Elision were further confirmed by the large measurement error variances for CTOPP Elision. The measurement error variances of spelling real words and spelling pseudowords were generally smaller than the other measures, suggesting that these two measures provided larger proportions of variance explained for the groups. All the factor loadings were statistically significant.

Evaluation of structural invariance

The final step in the stepwise approach is to establish cross-group constraints across all groups. The model-data fit of the cross-group constraint model was good: χ^2 (51) = 61.92, p = .14, CFI = .992, TLI = 0.991, RMSEA = .04, 90% CI = 00 to .08, and SRMR = .09. The Satorra-Bentler scaled chi-square difference test between the final measurement model and the cross-group constrain model was not significant (Δ S-B χ^2 = .09, Δ df = 9). Similarly, the CFI difference test (Δ CFI) between the constrained model and the cross-group constrained model was less than .01. For reasons of parsimony, the simpler model (i.e., model with cross-group constraints) was retained as the final model. Thus, the factor structures between the four groups (i.e., boys ineligible for FRL, girls ineligible for FRL, boys eligible for FRL, and girls eligible for FRL) were the same.

Evaluation of latent means

To evaluate the equality of latent means, first, we conducted an omnibus test of the null hypothesis by fixing every group's factor means to zero (Thompson & Green, 2006). The null hypothesis was rejected (p < .01), suggesting that one or more factor means differ between groups. Table 4 presents the latent means and the standardized effect sizes for the latent factors of the four sociodemographic groups. To derive the standardized effect size (d) of the differences between two latent factor means, we divided the latent factor means difference by the square root of the latent factor variance (Hancock, 2001; Thompson & Green, 2006).

First, we fixed the boys ineligible for FRL as the reference group. The results revealed that the girls ineligible for FRL significantly outperformed the boys ineligible for FRL in alphabet knowledge by 6.26 points (p < .01, d = +.54). The former had higher scores in phonological awareness and spelling than the latter, but these results were not significantly different. In comparison to the boys ineligible for FRL, the boys eligible for FRL scored 4.57 points lower in alphabet knowledge (p < .05, d = –.31), 2.05 points lower in phonological awareness (p < .001, d = –.58), and 4.41 points lower in spelling (p < .001, d = –.63). These mean differences were statistically significant. In contrast to the boys ineligible for FRL, the girls eligible for FRL scored 1.48 points lower in phonological awareness (p < .05, d = –.44) and 2.62 points lower in spelling (p < .05, d = –.45).

Next, we fixed the girls ineligible for FRL as the reference group. When comparing the girls from the two SES groups, the girls ineligible for FRL significantly outperformed the girls eligible for FRL in all three measures: 5.38 points higher in alphabet knowledge (p < .01, d = .34), 2.43 points higher in phonological awareness (p < .001, d = .72), and 4.35 points higher in spelling (p < .001, d = .75). The girls ineligible for FRL also significantly outperformed the boys eligible for FRL in all three measures: alphabet knowledge by 10.83 points (p < .001, d = .72), phonological awareness by 2.99 points (p < .001, d = .84), and spelling by 6.13 points (p < .001, d = .87).

Table 4. Parameter estimates for the mean structure of a three-factor model with factor mean differences across four sociodemographic groups.

	Boys ineligible for FRL		Girls ineligible for FRL			Boys eligible for FRL			Girls eligible for FRL		
Factor means	Unstd.	*SE*	Unstd.	*SE*	*d*	Unstd.	*SE*	*d*	Unstd.	*SE*	*d*
AK	0	–	6.26**	2.26	+.54	–4.57*	2.01	–.31	0.88	2.14	+.06
PA	0	–	.94	.66	+.32	–2.05***	.57	–.58	–1.48*	.58	–.44
SP	0	–	1.72	1.30	+.25	–4.41***	1.11	–.63	–2.62*	1.08	–.45
AK			0	–		–10.83***	2.04	–.72	–5.38**	2.12	–.34
PA			0	–		–2.99***	.54	–.84	–2.43***	.55	–.72
SP			0	–		–6.13***	1.08	–.87	–4.35***	1.05	–.75
AK						0	–		5.45**	1.91	+.34
PA						0	–		.57	.44	+.17
SP						0	–		1.78*	.80	+.31

Note: AK = alphabet knowledge; PA = phonological awareness; SP = spelling; FRL = free or reduced-price lunch; Unstd. = unstandardized latent means; *SE* = standard error; *d* = standardized latent factor effect size calculated based on guidelines by Hancock (2001).
*$p < .05$, **$p < .01$, ***$p < .001$. For all other unstandardized estimates, $p > .05$.

Last, we fixed the boys eligible for FRL as the reference group. The mean differences between the boys eligible for FRL and the girls eligible for FRL were statistically significant on two measures; the girls outperformed the boys in alphabet knowledge (5.45 points, $p < .01$, $d = .34$) and spelling (1.78 points, $p < .05$, $d = .31$). There was no statistically significant difference between these two groups in phonological awareness.

Discussion

The primary goal of the present study was to examine the SES-gender group differences in early literacy skills among kindergarteners who attended high needs schools in a southeastern US city. Another equally important goal was to determine whether the end-of-year kindergarten measures comprising alphabet knowledge, phonological awareness, and spelling constructs were measured in the same degree of accuracy for boys and girls who were (in) eligible to receive free/subsidized meals. We also evaluated the structural invariance and equality of latent means between these sociodemographic groups. To our knowledge, no other study has simultaneously examined the variables for measurement and structural equivalence, and the equality of latent means in early literacy skills across SES and gender groups. Additionally, we used more than one early literacy skill measure to compare SES and gender groups (cf. Entwisle et al., 2007).

Key findings

The results from this study support our theoretical framework that the latent factors were defined by their corresponding measured variables; alphabet knowledge (LNF and LSF); phonological awareness (CTOPP Blending and Elision); and spelling (decodable real words and decodable pseudowords). Previous research has established that these early literacy constructs and indicators are important determinants of subsequent literacy skills (e.g., NELP, 2008). Furthermore, this study not only confirms that the poverty and the gender gaps in literacy achievement begin early, but the poverty and gender gaps in early literacy skills are more nuanced than previously reported (Chatterji, 2006; Entwisle et al., 2007; Matthews et al., 2009; West et al., 2000). Using multiple measures and factors, we were able to highlight the specific gaps in componential literacy skills between boys and girls who were from varying SES backgrounds.

Measurement invariance across four sociodemographic groups

In addressing our first research question, we used a relatively sophisticated data-analytic method to rule out the potential measurement invariance in the relations between the measured variables and the latent factors across all sociodemographic groups. This study corroborates a past study by Townsend and Konold (2010), who found that the preschool measures comprising alphabet knowledge, phonological awareness, and print concepts were equivalently accurate across male and female students. Additionally, we found that there was measurement invariance for all measures across the four sociodemographic groups. Thus, the six observed variables measured their respective constructs in the same way across kindergarteners who were of different gender and socioeconomic status. In other words, the true differences in scores were due to the group differences. The present finding is promising because it demonstrates that the measures used did not differentially estimate the early literacy performance of students from different sociodemographic backgrounds.

Structural invariance across the four sociodemographic groups

The second research question addressed whether there was structural invariance (i.e., population heterogeneity) between the four groups. We tested for structural invariance by fixing the factor covariances to equality across groups. The results suggest that interrelations did not vary significantly across the four sociodemographic groups. The latent constructs were similarly correlated in one group as in other groups.

Latent means differences across the four sociodemographic groups

The final research question addressed whether there were differences in the latent means across the four sociodemographic groups. Our findings demonstrate that the nuanced gender differences in early literacy skills among kindergarteners relate to FRL. Importantly, our study found that both boys and girls eligible for FRL significantly underperformed girls ineligible for FRL in alphabet knowledge, phonological awareness, and spelling. A similar pattern was found when comparing between the children from high-poverty households and the boys from low-poverty households, except that the girls from high-poverty households were not significantly different from the boys from low-poverty households in alphabet knowledge. Among children from high-poverty households, girls eligible for FRL significantly outperformed boys eligible for FRL in two constructs: alphabet knowledge and spelling. Conversely, among children from low-poverty households, girls ineligible for FRL significantly outperformed boys ineligible for FRL in only alphabet knowledge. These findings demonstrate that the simple approach of comparing SES group differences without simultaneously examining gender differences vice versa misses out on the nuanced interaction between SES and gender. Early literacy achievement gap exists not only between children from low-SES and high-SES backgrounds, but the gender gap also exists for those children who are ineligible for meal subsidies. Thus, this study extends Entwisle and colleagues' findings (2007) that gender interacts with SES much earlier than Grade 2.

Instructional implications

The present study contributes to converging evidence that differences in early literacy skills begin early. The standardized effect sizes suggest that the boys and girls eligible for the meal subsidies are particularly weak in all three constructs: alphabet knowledge, phonological awareness, and spelling. Thus, componential literacy skills should be addressed early. That children from poor families are generally weak in phonological awareness is not surprising (Bowey, 1995; Raz & Bryant, 1990). The present study suggests that it is practically important for teachers to consider the vulnerability of boys eligible for FRL in early literacy skills. Furthermore, considering that the aforementioned constructs are important determinants for later literacy skills and based on what we already know about early identification, the protracted stability of literacy skills in individuals, and motivational issues related to later reading success, high-quality early preparation in literacy skills in preschool and at home may help reduce literacy-related difficulties, poverty gaps, and gender gaps.

Limitations and future research

There are several limitations related to this study. The findings are limited to SES-gender group differences in early literacy skills among kindergarteners. A further study is warranted to investigate how constructs that are measured generalize across groups longitudinally and importantly, whether the nuanced patterns in literacy skills in SES-gender groups found in the present study persist over time. Another limitation is the exclusion of letter writing, which was recommended by NELP (2008) as a strong predictor of later literacy skills. There was only one writing construct (i.e., letter writing fluency) from the larger study (Al Otaiba et al., 2010), and, thus, the two-indicator rule necessary for confirmatory factor analyses was not met (Kline, 2011). Additionally, there was insufficient data (n = 12) on children with limited English proficiency for group analysis. Future studies should include writing measures such as dictated letter writing (Ritchey, 2008). In addition, future studies should investigate how classroom management techniques and student's self-regulation interact with literacy skills. Self-regulation has been linked to gender differences; studies show that young boys in the lowest 10% in self-regulation ratings perform worse than girls in the same percentile (Matthews et al., 2009). Moreover, teachers are more likely to consider boys from low-SES backgrounds as having inattentive issues than girls from low-SES backgrounds (Entwisle et al., 2007). Finally, although the CFA model was theory driven (i.e., phonological awareness and alphabet knowledge are different but related factors; Hulme & Snowling, 2013; Townsend & Konold, 2010), the TLI for boys ineligible for FRL that was greater than 1 suggests a probable overfitting. Thus, a replication with larger sample size should be conducted in future studies.

Funding

This work was supported by grant P50HD052120 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

Notes on contributors

Julia Lee Ai Cheng, PhD, works as a Senior Lecturer at the Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak. She holds a master's degree in Educational Psychology from the University of Wisconsin-Madison and a doctoral degree in Special Education from Florida State University. While Dr. Lee was at Florida State University, she worked closely with renowned professors in her field such as Professor Dr. Stephanie Al Otaiba and Professor Dr. Barbara Foorman, co-chairs of her doctoral dissertation. Dr. Lee also underwent substantial training as a reading interventionist. She has also volunteered for America Reads, a national literacy campaign. Her research interests include early identification and remediation of reading difficulties and disabilities, language and literacy profiles among early elementary students, and technology use for the early identification and intervention of reading difficulties and disabilities. Dr. Lee's current research project, in her bid to contribute to the advancement of learning disabilities research in Malaysia, is entitled "Development of a comprehensive diagnostic instrument for the identification and classification of language and literacy skills in primary school children with reading disabilities". This research project, which runs from 2013–2015, is funded by the Malaysian Ministry of Higher Education. Dr. Lee enjoys teaching undergraduate courses in cognition and learning, human development psychology, and providing one-to-one intervention sessions to children with learning disabilities. She also teaches a graduate course entitled *Social and cultural dimensions in learning and instruction*.

Stephanie Al Otaiba, PhD, joined Southern Methodist University in January 2012 as Professor of Teaching and Learning in the Annette Caldwell Simmons School of Education & Human Development. Previously, she was an Associate Professor at the Florida State University (FSU) and was on faculty at the Florida Center for Reading Research. She received her PhD in 2000 in Human Development from Peabody College of Vanderbilt University, and prior to that, she was a special educator for over a decade in the United Arab Emirates. She enjoys teaching graduate courses in literacy, special education, assessment, response to intervention, and mentoring doctoral students.

Dr. Al Otaiba's research interests include school-based literacy interventions, response to intervention, learning disabilities, diverse learners, and teacher training. She has published over 80 journal articles and book chapters related to these interests. She has also developed reading curricular materials. Her line of research has been supported by several federally funded grants from the U.S. Department of Education Institute of Education Sciences, the Office of Special Education Programs, and from the National Institute of Child Health and Human Development. Her dissertation was awarded the 2001 Outstanding Dissertation Award from the International Reading Association, and in 2010 she was the recipient of The Council for Exceptional Children Division for Research Distinguished Early Career Research Award. She also received the Developing Scholar Award and the Graduate Faculty Mentor Award at FSU in 2010. She teaches graduate courses in the Reading and Writing and Special Education programs. Dr. Al Otaiba serves on review panels for grants for the Institute of Education Science and has reviewed for the Office of Special Education Programs. She is an Associate Editor of *Education Researcher* and the *Elementary School Journal* and also serves on numerous editorial boards for scholarly journals in the field of education. Dr. Al Otaiba has consulted nationally and internationally related to early literacy intervention and assessment.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*, *111*, 535–560.

Al Otaiba, S., Puranik, C. S., Rouby, D. A., Greulich, L., Sidler, J. F., & Lee, J. (2010). Predicting kindergarteners' end-of-year spelling ability based on their reading, alphabetic, vocabulary, and phonological awareness skills, as well as prior literacy experiences. *Learning Disability Quarterly*, *33*, 171–183.

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, *54*, 517–545.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *57*, 289–300.

Blachman, B. A. (2000). Phonological awareness. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 483–502). Mahwah, NJ: Erlbaum.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, *87*, 476–487.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children: Children and Poverty*, *7*, 55–71.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B., & Fielding-Barnsley, R. (1993). Recognition of phoneme invariance by beginning readers: Confounding effects of global similarity. *Reading and Writing*, *5*, 315–324.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*, 155–175.

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*, 163–176.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, *98*, 489–507.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*, 233–255.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification errors in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.

Ehri, L. C., & Wilce, L. S. (1987). Does learning to spell help beginners learn to read words? *Reading Research Quarterly*, *22*, 47–65.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, *80*, 114–138.

Finney, S. J., & DiStefano, C (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: Information Age.

Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., … Shaywitz, B. A. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to policy* (pp. 185–250). Mahwah, NJ: Erlbaum.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*, 3–17.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., … O'Connor, R. E. (2001). Is reading important in reading-readiness programs? A randomized controlled field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*, 251–267.

Gamoran, A. (Ed.). (2007). *Standards-based reform and the poverty gap: Lessons for "No Child Left Behind"* (pp. 3–16). Washington, DC: Brookings.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*, 373–388.

Harper, S. N., & Pelletier, J. P. (2008). Gender and language issues in assessing early literacy: Group differences in children's performance on the test of early reading ability. *Journal of Psychoeducational Assessment*, *26*, 185–194.

Hirschfeld, G., & von Brachel, R. (2014). Multiple-group confirmation factor analysis in R-A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, *19*(7). Retrieved from http://pareonline.net/getvn.asp?v=19&n=7

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Hulme, C., & Snowling, M. J. (2013). The interface between spoken and written language: Developmental disorders. *Philosophical Transactions of the Royal Society B*, 369: 20120395. Retrieved from http://rstb.royalsocietypublishing.org/content/369/1634/20120395

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test manual*, *2nd edition*. Circle Pines, MN: American Guidance Service Publishing.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007–496). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2007/2007496.pdf

MacDonald, G. W., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities*, *28*, 523–527.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, *101*, 689–704.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14–28.

Mulligan, G. M., Hastedt, S., & McCarroll, J. C. (2012). *First-time kindergartners in 2010–11: First findings from the kindergarten rounds of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011)* (NCES 2012-049). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Muthén, B., & Muthén, L. (1998-2010). Mplus (6.1) [Statistical software]. Los Angeles, CA: Authors.

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

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Bethesda, MD: National Institute of Child Health and Human Development.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*, 159–197.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology*, *8*, 209–225.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *The Elementary School Journal*, *106*, 21–38.

Ritchey, K. D. (2008). The building blocks of writing: Learning to write letters and spell words. *Reading and Writing*, *21*, 27–47.

Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology*, *31*, 301–327.

Santoro, L. E., Coyne, M. D., & Simmons, D. C. (2006). The reading-spelling connection: Developing and evaluating a beginning spelling intervention for children at risk of reading disability. *Learning Disabilities Research & Practice*, *21*, 122–133.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282.

Shinn, M. M., & Shinn, M. R. (2004). *AIMSweb*. Eden Prairie, MN: Edformation.

Snow, C. E. (2006). What counts as early literacy in early childhood? In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 274–294). Malden, MA: Blackwell.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education*, *36*, 223–233.

Tangel, D. M., & Blachman, B. A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior*, *24*, 233–261.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 119–170). Greenwich, CT: Information Age.

Townsend, M., & Konold, T. R. (2010). Measuring early literacy skills: A latent variable investigation of the phonological awareness literacy screening for preschool. *Journal of Psychoeducational Assessment*, *28*, 115–128.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

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*, 468–479.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergartners: Findings from the Early Childhood Longitudinal Study, kindergarten class of 1998-99, fall 1998.* Washington, DC: National Center for Education Statistics, U. S. Department of Education, Office of Educational Research and Improvement.