Studies in Educational Evaluation 55 (2017) 9–18

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Kindergarten standardized testing and reading achievement in the U.S.: Evidence from the early childhood longitudinal study☆

Haesung Im

Ewha Womans University, Banpo Raemian Firstage 104-602, 275 Banpo-daero, Seocho-gu, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Standardized testing
Kindergarten
Reading achievement & reading instruction
Early childhood longitudinal study-kindergarten (ECLS-K) 2010–2011

ABSTRACT

Drawing from data use theory (i.e., a theory for making data-driven educational decisions), the present study sought to understand how frequency of standardized testing is related to student learning, mediated by reading instruction, after controlling for child-level (e.g., gender, race/ethnicity) and school-level covariates (e.g., private/public, proportion of students eligible for free lunch). Using data from the Early Childhood Longitudinal Study Kindergarten Cohort of 2010–2011, the sample included 12,241 children attending 1067 kindergartens in the U.S. findings from a multilevel structural equation mediation model suggest that the frequency of state/local standardized testing in kindergarten did not have a direct effect on reading achievement near the end of kindergarten, after controlling for covariates. However, the amount and type of reading instruction mediated the relationship between the frequency of testing and reading achievement, after controlling for covariates. The implications for policy and practice on the use of standardized tests in kindergarten are discussed.

1. Introduction

After the inception of the No Child Left Behind Act (NCLB, 2001), assessment-driven accountability began to play an increasingly critical role in shaping curricular and instructional practice in the United States (Ravitch, 2011). The common elements of accountability included standards, assessments, and public reporting policies to hold schools and teachers accountable for raising student performance (Goertz & Duffy, 2000). The NCLB utilized standardized tests as a catalyst to improve instruction (Hanushek & Raymond, 2005). By definition, standardized tests are testing instruments that are administered, scored, and interpreted in a uniform manner (Martella, 2010). While some standardized tests are designed based on state content standards (e.g., benchmark tests), other types of standardized tests are developed without referencing the state content standards (Goldstein & Flake, 2016; McMillan, 2013). Typically, those tests include end of year standardized tests that are designed based on state content standards, high-stakes standardized tests, annual statewide accountability tests, interim tests developed by districts, and commercially produced tests (Goldstein & Flake, 2016).

Standardized testing has increasingly become a key instructional instrument in the field of early childhood education (Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009; National Research Council, 2008). More than 70% of young children in the U.S. completed standardized tests in kindergarten at least once in the 2010–2011 academic year (Bassok, Latham, & Rorem, 2016). Increasing numbers of states are developing pre-kindergarten standardized assessments for school readiness. As such, the use of standardized tests during early childhood has been at the center of the educational debate over the last two decades. Proponents of standardized testing believe that these scores can be used for monitoring and evaluating teaching effectiveness and students' learning outcomes (Hutchinson, & Young, 2011). They believe testing will raise student performance by making teachers more accountable for their teaching (Crocker, 2005; The National Early Childhood Accountability Task Force, 2007). Furthermore, advocates of standardized tests also believe that test scores can be used to improve teaching effectiveness through targeted professional development (Crocker, 2005).

Conversely, scholars and experts have published a substantial amount of criticisms, warnings, and guidelines to inform the direction for the use of standardized tests during early childhood (Gullo & Hughes, 2011; Wilson, 2009). In 2003, the National Association for the Education of Young Children (NAEYC) announced a position statement that argued that standardized testing is inappropriate for young children, due to their distinct nature, developmental stage, and rapid growth. Essentially, this statement purports that young children are not cognitively ready to understand the goals of standardized tests and their process (Goldstein & Flake, 2016; Meisels, 2007; Schultz et al., 2007). A specific level of language skills, possibly beyond the reach of many young children, is required to successfully

☆ The data provided in the current manuscript are part of a larger doctoral dissertation study.

E-mail address: comet815@gmail.com.

http://dx.doi.org/10.1016/j.stueduc.2017.05.001
Received 31 October 2016; Accepted 11 May 2017
Available online 23 May 2017

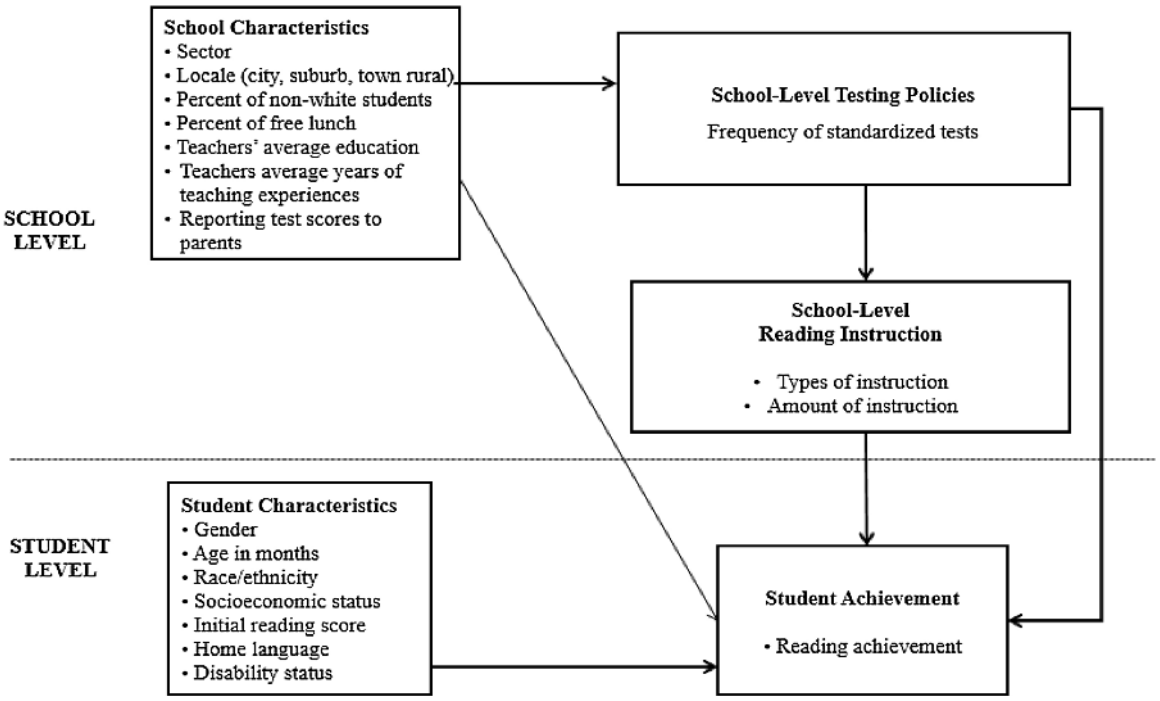

Fig. 1. A Conceptual Model Relating Standardized Testing to Student Achievement.

complete standardized tests. Further, younger children are more easily distracted and influenced by their emotional status or physical needs such as hunger or fatigue (Charlesworth, Fleege, & Weitman, 1994; NAEYC, 2003). Hence, several national organizations and scholars argue that young children should not complete standardized testing before the end of third grade (NAEYC, 2003; National Association of Early Childhood Specialists in State Department of Education, 2000; Schultz et al., 2007; Solley, 2007).

Despite the debate on the use of standardized tests during early childhood, the link between these tests and learning outcomes of young children has been tenuous at best, due to the lack of studies examining early childhood education (Bauml, 2016; Pyle & DeLuca, 2013; Solley, 2007). Specifically, while some studies have investigated the connection between standardized testing and student achievement during middle childhood and early adolescence (Amrein & Berliner, 2002; Dee & Jacobs, 2011; Rosenshine, 2003), fewer empirical studies have explored this relationship during early childhood (Bauml, 2016; Boat et al., 2005; Charlesworth et al., 1994; Hatch & Grieshaber, 2002; Rous, McCormick, Gooden, & Townley, 2007). Additionally, these early childhood studies have exhibited limited generalizability, due to their small sample sizes and qualitative research designs.

Even though the administration of standardized tests in kindergarten can result in a significant change in reading instruction (Au, 2007; Gullo & Hughes, 2011; Hirsh-Pasek et al., 2009; Merki, 2011; Miller & Almon, 2009), the existing body of research has primarily focused on direct relationships between standardized testing and students' academic outcomes, without investigating the mediating role of instructional practices. Thus, the current study examined both the direct and indirect effects of standardized tests on kindergarten children's reading achievement and the mediating role of the amount and types of reading instruction.

1.1. Theoretical and conceptual framework

Drawing from data use theory, the current study sought to investigate the connection between the frequency of standardized testing and reading achievement in kindergarteners, as mediated by reading instructional practice (Hamilton et al., 2009; Marsh et al., 2006; Spillane, 2012). Data use theory offers insight into why certain kindergartens may or may not utilize standardized test scores for student learning (Brunner et al., 2005; Young, 2006). According to this theory, there are three stages of data-driven decision making for student learning (Mandinach, Honey, & Light, 2006). Here, data refers to all types, including information about specific programs, family backgrounds of students, student behavior, attendance, and assessment scores. At the first stage, data exists in a raw state without meaning. At this point, teachers do not use assessment data to improve instruction. At the second stage, teachers utilize data as "information" to understand the educational context and their students. However, they do not use this information to guide their instruction. Finally, at the third stage, teachers utilize data as "knowledge" to modify, or refine their instruction. This continuum of data utilization provides insight into why the frequency of testing may or may not be related to student learning mediated by teachers' use of data (Hamilton et al., 2009; Marsh et al., 2006).

Data use theory helps us to understand the complexity of data-driven decision making practices at multiple levels (Coburn & Talbert, 2006; Kerr et al., 2006). According to Spillane (2012), data-based decision-making is influenced not only by *individual teachers' cognitions,* but is also affected by organizational norms at the school level. At an individual level, there are multiple factors that influence teachers' utilization of assessment data. For example, if kindergarten teachers did not regard standardized tests as valid and reliable instruments for assessing young children, those kindergarten teachers would be more likely to disregard test scores as a means for making instructional decisions (Bauml, 2016; Brown & Goldstein, 2013; Pyle & DeLuca, 2013). Another important factor that impacts kindergarten teachers' use of data is called "assessment literacy," which denotes teachers' ability to collect, analyze, and utilize all types of data for student learning (Hamilton et al., 2009). Teachers' pedagogical content knowledge also plays a key role in making instructional decision based on the data (Mandinach & Gummer, 2015). For example, when kindergarten teachers believe in a "whole language" approach to reading instruction, these teachers are more likely to allocate more instruction time for whole language reading instruction, when their students do not perform well on reading assessments. Similarly, advocates of "phonics" would

increase instruction time with phonics instruction based on the analysis of the same assessment data.

At the school-level, when a school encourages their teachers to use data for student learning, these teachers are more likely to use assessment data to collectively modify their instruction (Kerr et al., 2006; Means et al., 2009; Spillane, 2012). In contrast, when a school focuses on accountability, teachers at that school are less likely to use data to facilitate student learning (Schildkamp & Kuiper, 2010; Shen & Cooley, 2008: Young, 2006). In light of data use theory, the conceptual framework outlined in Fig. 1 describes how the frequency of standardized testing in kindergarten impacts children's reading achievement both directly and indirectly through reading instruction.

1.2. Role of standardized testing in children's reading achievement

The existing body of literature that has sought to investigate the links between standardized testing and reading achievement in middle childhood is laden with mixed findings. Some studies suggest that states with high-stakes testing (i.e., externally mandated standardized tests that are attached with serious consequences for teachers, schools, and students) perform better in fourth grade reading achievement compared to states without this testing policy (Hanushek & Raymond, 2005; Kober et al., 2008; Rosenshine, 2003).

Other researchers have found that accountability policy was not related to greater fourth grade reading achievement (Amrein & Berliner, 2002; Dee & Jacob, 2011; Lee & Reeves, 2012; Nichols et al., 2006). According to comparative, interrupted time-series analyses of 1990–2009 NAEP state assessment data, no improvement was noted in average fourth grade reading achievement after the implementation of NCLB (Lee & Reeves, 2012). In fact, investment in statewide educational resources (e.g., investment in qualified teachers and small class size) was found to be more effective in promoting student achievement. Amrein and Berliner (2002) also found that 46% of the states with high-stakes testing exhibited fourth grade reading gains. This may have been because many states intentionally excluded English Language Learners and students with special needs. However, the application of empirical findings from middle childhood toward young children requires caution because of the distinct characteristics of early childhood (NAEYC (2003); Schultz et al., 2007; Solley, 2007).

1.3. Associations among standardized testing, reading instruction, and reading achievements

Studies have indicated that the use of standardized tests in kindergarten can impact the landscape of reading instruction in significant ways (Hirsh-Pasek et al., 2009; Kontovourki, 2009). Generally, many scholars argue that testing policy has a tremendous impact on both the types and amount of reading instruction. Studies have indicated that administering standardized tests in kindergarten is associated with increased reading instruction time (Bassok et al., 2016; McMurrer & Kober, 2007; Miller & Almon, 2009). This increase in reading instruction time has been shown to be devoted primarily to test preparation, that is, teaching to the test with decontextualized instruction (Gallant, 2009; Kontovourki, 2009). Not surprisingly, research has indicated that increased reading instruction time is associated with greater gains in reading achievement (Cavanaugh, Kim, Wanzek, & Vaughn, 2004; Chatterji, 2005; Harn, Linan-Thompson, & Roberts, 2008; Simmons et al., 2007; Sonnenschein, Stapleton, & Benson, 2010). However, relatively little research has examined how the effect of testing policy on students' reading achievement is mediated by reading instruction time.

Generally, there is a consensus that the types of instruction make a significant difference in supporting children to read and write (Afflerbach, 2011; Pearson, 2004). In this study, reading instruction was categorized into three types, by its orientation towards meaning versus decoding skills: (1) phonics, (2) whole language, and (3) a balanced approach. First, phonics is explicit instruction in the form and sounds of letters used to decode written language. The benefits of explicit phonics instruction, particularly on children with low literacy, have been well-documented (Juel & Minden-Cupp, 2000; Morrison & Connnor, & Bachman, 2006; Xue & Meisels, 2004). Next, whole language highlights the importance of learning language as a whole, in a meaningful context, without the need for explicit instruction (Goodman & Goodman, 2009; Pearson, 2004). Children learn to read and write by engaging in self-selected projects, such as writing, retelling stories, and performing plays and skits (Sonnenschein et al., 2010; Xue & Meisels, 2004). Research generally suggests that whole language is generally associated with improved reading achievement (Krashen, 2002; Sonnenschein et al., 2010), except in low-achieving children (Morrison & Connor, 2002; Morrison et al., 2006; Sonnenschein et al., 2010; Xue & Meisels, 2004). Finally, a balanced approach posits that phonics can be fostered in context by reading predictable books, stories, rhymes, and songs (Adams, 1990; Dahl et al., 1999; Farris, Fuhler, & Walther, 2004; Hornsby & Wilson, 2010; Roberts & Meiring, 2006). Going beyond the dichotomy of phonics versus whole language, an experimental study indicated that children who learned phonemes within contextualized instruction displayed higher reading achievement when compared to children in a control group who learned phonemes in isolation (Bitter, O'Day, Gubbins, & Socias, 2009; Dahl et al., 1999; Donat, 2006).

A plethora of studies have indicated that curricular decisions made within kindergartens are constrained by standardized tests (Bassok et al., 2016; Bauml, 2016; Gallant, 2009; Gullo & Hughes, 2011; Hirsh-Pasek et al., 2009). Although there is a general consensus that testing policy has a tremendous impact on the types of early reading instruction, relatively few studies have provided empirical evidence based on investigating the mediating role of reading instruction. On the one hand, studies have reported that testing policy places greater emphasis on phonics at the expense of whole language (Afflerbach, 2011; Pearson, 2004). In fact, kindergarten teachers in 2009 reported more emphasis on explicit instruction (e.g., worksheets and phonics workbooks), to prepare for state standardized tests compared to reading instruction in 1994 (Gallant, 2009; Lipson, Goldhaber, Daniels, & Sortino, 1994). On the other hand, researchers have also reported that some teachers utilize whole language despite the pressure to teach to the test with direct instruction (Brown & Goldstein, 2013). Given that the frequency of standardized testing may have a serious impact on kindergarten reading instruction, there is a need for evidence of the mechanism through which school-level testing policy affects young children's learning outcomes mediated by reading instructional practices.

1.4. Present study

To fill the gap within the literature, the present study investigated the direct and indirect effects of kindergarten standardized testing on reading achievement, with nationally representative data. To the best of the author's knowledge, no research has investigated both direct and indirect effects of school-level standardized tests on young children's learning outcomes in kindergarten. With data use theory, the present study will provide empirical evidence of the process through which the frequency of kindergarten standardized testing impacts reading achievement, as well as the mediation effects of the amount and types of reading instruction. The research questions are as follows:

- Is the frequency of standardized testing in kindergarten directly associated with children's reading achievement at the end of kindergarten, after controlling for student-level and school-level covariates?
- Does the amount and types of reading instruction in kindergarten mediate the relationship between the frequency of standardized testing and children's reading achievement near the end of kinder-

garten?

2. Methods

2.1. Data and sample

In this study, data were selected from the Early Childhood Longitudinal Study-Kindergarten Cohort of 2010–2011 (ECLS-K), which is a nationally representative dataset, collected by the National Center for Education Statistics (NCES, 2009) in the United States. The database has a sample of approximately 20,000 U.S. children who attended kindergarten in the academic year of 2010–2011. With multi-stage stratified sampling strategies, 23 children were selected from each of 1240 kindergartens that were sampled from 50 states. Among those kindergartens, about 35% of the kindergartens were sampled from large cities, 37% from suburban or urban fringe areas, 7% from towns, and approximately 21% from rural areas. In addition, roughly 85% were public, 4% were private, and approximately 10% were Catholic or other religious kindergartens.

Children who did not move schools from the fall of kindergarten to the spring of first grade were included in the analyses. In addition, children were excluded when there was a discrepancy between two sets of school-level testing policies, and when cases had missing values or variables that were weighted as zero. The final un-weighted sample consisted of 12,241 children, attending 1067 kindergarten classrooms, in both public and private schools nationwide (see Table 1 for further description).

2.2. Measures

2.2.1. Frequency of kindergarten standardized tests

The present study had a two-level nested structure: Level 1 being the student-level and Level 2 being the school-level. Thus, the student-level represents characteristics of the kindergarten children, while the school-level characterizes kindergarten as a whole in the analyses. The mean frequency of state/local standardized tests in the spring of kindergarten was used as a predictor variable. This was measured by aggregating the frequency of standardized tests at the classroom-level, from kindergarten teacher questionnaires administered in the spring. While the frequency of state/local standardized tests was measured using a 4-point Likert scale (1 = *never*, 2 = *one or two times a year*, 3 = *one or two times a month*, 4 = *three or more times a week*), it was treated as a continuous variable. In this study, 67.5% of children at kindergarten took state and local standardized tests at least once a year while 32.5% did not take standardized tests at all. Among those kindergarteners who took standardized tests, 20.7% completed state/local standardized tests more than one time a month.

Table 1
Description of the Full and Final Samples.

	Full Sample	Final Sample	
	Unweighted (N = 18,174)	Unweighted (N = 12,241)	Weighted[b]
Child characteristics			
Male (%)	51.1	51	51
White (%)	46.7	40.3	53.6
Black (%)	13.2	15.3	11.9
Hispanic (%)	25.2	26.4	23.6
Asian (%)	8.5	10.8	5.2
Other (%)	6.1	7.2	5.7
Non-English household (%)	17.2	23.1	20.8
SES	−0.05(.81)	−0.07	−0.07
Reading Achievement (K)	37.31(9.66)	37.98 (9.85)	37.73 (9.56)

[b] Weighted by W4C4P_20.

2.2.2. Types of kindergarten reading instruction

Kindergarten teachers were asked what types of reading instruction they utilized via a 17-item teacher questionnaire in the spring of 2011. These 17 items included, "how often do children in this class do each of the following reading activities, such as matching letters to sound, identifying the main idea and parts of a story, communicating complete ideas orally?" Kindergarten teachers responded to all survey items using a 6-point Likert scale (1 = *never*, 2 = *once a month or less*, 3 = *two or three times a month*, 4 = *once or twice a week*, 5 = *three or four times a week*, 6 = *daily*). The types of reading instruction at the school level were created based on the factor structures previously reported by researchers, using ECLS-K data (Sonnenschein et al., 2010; Xue & Meisels, 2004). The three-factor model from a confirmatory factor analysis (CFA) included whole language (6 items), phonics (3 items), and a balanced approach (3 items). Detailed information about the three factors are documented in the result section and Fig. 2.

2.2.3. Amount of kindergarten reading instruction time

Weekly school-level reading instruction time was measured by teacher questionnaire completed in the spring. The amount of time spent in reading instruction was constructed by multiplying the frequency of weekly instruction time by the duration of each instruction time. The frequency of school-level reading instruction time was assessed on a 5-point scale (1 = *never*, 2 = *less than once a week*, 3 = *1-2 times a week*, 4 = *3-4 times a week*, and 5 = *daily*). The duration of daily instruction time was rated with a 4-point scale (1 = *1–30* min, 2 = *31–60* min, 3 = *61–90* min, and 4 = *more than 90* min).

2.2.4. Child reading achievement

Reading outcomes were assessed using each child's Item Response Theory (IRT) score at the completion of kindergarten (i.e., spring semester). Reading competency comprises a holistic measure of reading comprehension, vocabulary knowledge, and children's basic skills (e.g., print familiarity, letter recognition, beginning and ending sounds, rhyming words, and word recognition). For reading comprehension questions, kindergarten children were asked to identify information specifically stated within a text. For example, young children were required to state definitions, facts, and supporting details from a segment of text and to make complex inferences within and across texts (Tourangeau et al., 2015). The reading assessment consisted of selected items from the Test of Preschool Early Literacy (TOPEL), the Peabody Individual Achievement Test − Revised (PIAT-R), the Peabody Picture Vocabulary Test − 3rd Edition (PPVT-III), the Preschool Language Assessment Scale (*pre*Las 2000) Form C, and the Test of Early Reading Ability − 3rd edition (TERA-3). Special accommodations were made for English Language Learners. Across the four waves of data collection, weighted mean reading IRT scores were 37.73 (*SD* = 9.56) for fall of kindergarten and 50.06 (*SD* = 11.45) for spring of kindergarten, respectively.

2.2.5. Control variables

Multiple control variables were included at both the child and school levels. At the child level, socioeconomic status (SES), child's initial reading score, race, gender, age in months, child home language, and whether or not children had special needs were used as covariates. At the school level, school sector (public/private), proportion of children with free and reduced lunch, proportion of minority students, location, average teacher education degree, and average years of teaching experiences were included. Additionally, a variable was included specifying reporting of test scores. This variable was binary and indicated whether the school principal provided a state/local standardized test scores to parents or not.

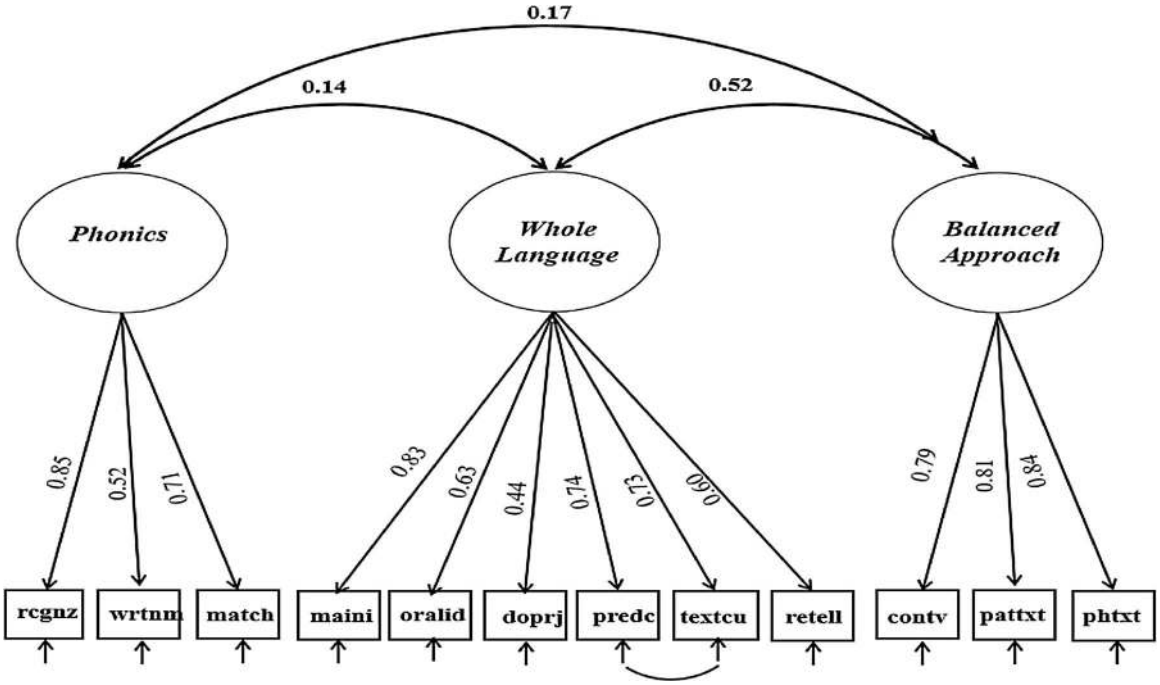

Fig. 2. Types of Reading Instruction Extracted from Confirmative Factor Analysis. RCGNZE = Recognizing alphabet and letters; WRTME = Writing own name; MATCH = Matching letters to sounds; MAINID = Identifying the main idea and parts of a story; ORALID = Communicating complete ideas orally; DOPROJ = Doing an activity or project related to a book or story; PREDIC = Making predictions based on text; TEXTCU = Using context cues for comprehension; RETELL = Retelling stories; CONVOC = Controlled vocabulary; PATTXT = Patterned text; PHOTXT = Phonetic text.

2.3. Analytic strategies

First, descriptive statistics for the frequency of standardized tests, reporting test scores to parents, reading instruction time, types of reading instruction, and concurrent reading achievement in the spring of kindergarten were obtained. All analyses were performed after applying the proper weight variable for estimating unbiased population parameters of nationally representative of the population. Second, a CFA was conducted to investigate the internal structure of the measured items, using *Mplus* 7 (Muthén & Muthén, 2012). Because the measured items in the scale were ordinal, a robust mean and variance adjusted weighted least squares (WLSMV) estimator was utilized (Muthén, Du Toit, & Spisic, 1997). Finally, a multilevel structural equation model was conducted by using the latent variables as the mediator (i.e., types of reading instruction from the 3-factor model) with *Mplus* 7 (Muthén & Muthén, 2012). To handle missing data, a Full Information Maximum Likelihood (FIML), with auxiliary variables, was used so that additional variables could increase the accuracy of the estimating model parameters and augment statistical power (Enders, 2010). These auxiliary variables included the number of books the children had at home and the frequency of books parents read to their children.

The current study modeled the effect of school-level standardized testing policy on children's reading achievement through the effect of school-level reading instruction (see Figs. 1 and 2). The multilevel structural equation modeling (MSEM) approach was employed, as the mediator occurred at a level of analysis higher than the preceding predictor (Preacher, Zyphur, & Zhang, 2010; Preacher, Zhang, & Zyphur, 2011). Specifically, while the antecedent (X) and mediator (M) were at the school level, the outcome variable (Y) was at the student level (Zhang, Zyphur, & Preacher, 2009). The equations below represent the model in this study.

$$\text{Level 1: } Reading\ Ach._{ij} = \beta_{0j} + Covariates + e_{ij} \quad (1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}InstTime_j + \gamma_{02}Phonics_j + \gamma_{03}Whole\ language_j + \gamma_{04}Balanced\ approach_j + \gamma_{05}Test_j + Covariates + u_{0j} \quad (2)$$

$$\text{Level 2: } InstTime_j = \gamma_{10} + \gamma_{11}Test_j + u_{1j} \quad (3)$$

$$\text{Level 2: } Phonics_j = \gamma_{20} + \gamma_{21}Test_j + u_{2j} \quad (4)$$

$$\text{Level 2: } Whole\ language_j = \gamma_{30} + \gamma_{31}Test_j + u_{3j} \quad (5)$$

$$\text{Level 2: } Balanced\ approach_j = \gamma_{40} + \gamma_{41}Test_j + u_{4j} \quad (6)$$

$Reading\ Ach._{ij}$ denotes child *i*'s score, in school *j*, as a function of child level (Level 1) and school level (Level 2) predictors and covariates. β_{0j} is the Level 1 intercept, and e_{ij} is the Level 1 residual. The Level 1 intercept, β_{0j}, is a function of the following Level 2 predictors and covariates: the Level 2 intercept for each school (γ_{00}); the effects of school-level mediators on student-level reading achievement in the spring of kindergarten, adjusted for all other mediators and school-level testing policy (γ_{01}–γ_{04}); the effect of school-level testing policy on child-level reading achievement in the spring of kindergarten, adjusted for the effects of the mediators (γ_{05}); and the Level 2 residual for reading achievement (u_{0j}). The Level 2 intercepts for each school in reading instruction time, phonics, whole language, and a balanced approach were represented by γ_{10}–γ_{40}, respectively. The effects of school-level testing policy to school-level mediators (i.e., amount and type of reading instruction) were represented by γ_{11}–γ_{41}, respectively. Additionally, $u_{1j} - u_{4j}$ are Level 2 residuals for these respective variables. All paths were adjusted for child-level covariates (SES, child's initial reading score, race, gender, age in months, home language of child, and whether or not children had special needs) and school-level covariates (sector, location, percent of non-white students, percent of free lunch, mean education level of teachers, mean years of teaching experiences).

RMediation (Tofighi & MacKinnon, 2011) was used to estimate asymmetric confidence intervals for the indirect effects of interest. The asymmetric confidence interval estimation in RMediation provides the most accurate way of assessing the significance of indirect effects because it considers the fact that the product of the *a* and *b* paths is not normally distributed. Therefore, the asymmetric confidence interval estimation provides the most accurate Type 1 error rates and higher

Table 2
Weighted Pearson Product Moment Correlations for Study Variables.

Level 1 (Unweighted N = 12,241)

	Reading (s)	Reading (f)	English at Home	Age	Gender	SES	Hispanic	Asian	Black	White	Disability
Reading (s)	–										
Reading (f)	.84**	–									
English at Home	−0.17**	−0.16**	–								
Age	.15**	.19**	−0.10**	–							
Gender	−0.08**	−0.05**	.01**	.07**	–						
SES	.40**	.41**	−0.30**	.03**	.01**	–					
Hispanic	−0.20**	−0.19**	.49**	−0.09**	.01**	−0.33**	–				
Asian	.15**	.16**	.22**	−0.08**	−0.02**	.13**	−0.14**	–			
Black	−0.08**	−0.06**	−0.12**	−0.02**	.01**	−0.14**	−0.15**	−0.12**	–		
White	.01**	−0.01**	−0.05**	.06**	.02**	.07**	.22**	−0.42**	−0.66**	–	
Disability	−0.09**	−0.06**	−0.10**	.09**	.13**	−0.01**	−0.04**	−0.07**	−0.03**	.06**	–

Level 2 (Unweighted N = 1067)

	Minorities(%)	Sector	Reporting	Frequency of ST	Teachers' Education	Teachers' Years of Experiences	Free & Reduced Lunch (%)	Location Type of School
Minorities (%)	–							
Sector	−0.14**	–						
Reporting	.18**	−0.27**	–					
Frequency of ST	0.13	−0.20**	.43**	–				
Teachers' Education	0.01	−0.21**	−0.03**	−0.02**	–			
Teachers' Years of Experience	−0.11	.11**	−0.07**	−0.08**	.14**	–		
Free & Reduced Lunch	0.61	−0.47**	.27**	.20**	.08**	−0.12**	–	
Location Type of School	−0.5	−0.04**	−0.07**	−0.06**	−0.06**	.04**	−0.19**	–

Note: Weight is W4C4P_20 Weighted coefficients are below the diagonal.

statistical power than methods that do not take this non-normal distribution into account (MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, & Williams, 2004; Tofighi & MacKinnon, 2011).

3. Results

Table 2 shows the bivariate correlations calculated among study variables. Children's reading scores in kindergarten were positively associated with their age. Children who came from higher SES families, and had English as a home language, were more likely to have higher scores. Among school-level variables, public schools were more likely to administer frequent standardized tests. Schools that frequently administered standardized tests tended to have teachers with less teaching experience than schools that did not frequently administer standardized tests (see Table 2).

Based on the literature on the types of reading instruction from the ECLS-K 1998–1999 (Sonnenschein et al., 2010; Xue & Meisels, 2004), three-factors were extracted; phonics, whole language, and a balanced approach. The model fit indices for the three-factor model were χ^2 (49) = 620.02, $p < 0.001$, CFI = 0.96, TLI = 0.94, RMSEA = 0.03, and SRMR = 0.03, suggesting that the revised model fit the data well (See Fig. 2). Studies indicate that models fit the data well when the Comparative Fit Index (CFI) exceeds 0.95 (Hu & Bentler, 1999), when the Tucker-Lewis Index (TLI) exceeds 0.90 (Hu & Bentler, 1999), and when the Root-Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) are less than 0.08 (Browne & Cudeck, 1992; Hu & Bentler, 1999). The two-factor model (e.g., whole language versus phonics) from previous literature that utilized the ECLS-K 1998-199 (Sonnenschein et al., 2010; Xue & Meisle, 2004) did not fit the data well (χ^2 (274) = 35627.41, $p < 0.001$. CFI = 0.66, TLI = 0.62, RMSEA = 0.10, SRMR = 0.11) (Table 3).

The phonics factor included items such as recognizing the alphabet and letters, matching letters to sounds, and writing one's own name. The whole language factor included items such as identifying the main idea and parts of a story, communicating complete ideas orally, doing an activity or project related to a book or story, making predictions based on text, using context cues for comprehension, and retelling stories. The balanced approach factor included items such as reading books with phonetic text, reading books with patterned text, and reading books with controlled vocabulary. The subscale means ranged from 3.04 to 4.80, with the highest scoring subscales being recognition of alphabet and letters and the lowest scoring subscales found in doing an activity or project related to a book or story. The standardized factor loadings ranged from 0.52 to 0.85 for phonics, from 0.44 to 0.83 for whole language, and from 0.79 to 0.84 for a balanced approach. Phonics was positively, but weakly correlated with a balanced approach, ($r = 0.17$, $p < 0.001$), whole language was positively correlated with a balanced approach, ($r = 0.52$, $p < 0.001$), and positively but minimally correlated with phonics, ($r = 0.14$, $p < 0.001$). The graphic description of the reading instruction CFA is presented in Fig. 2.

3.1. Direct effects of frequency of standardized tests on reading achievement

The model used here is presented in Fig. 3. The fit indices of the model predicting children's reading achievement from the frequency of kindergarten standardized tests were, χ^2 (201) = 1465.71, $p < 0.001$, CFI = 0.95, TLI = 0.92, RMSEA = 0.02, and SRMR < 0.001 (within) and 0.11 (between), suggesting that the model fit the data well. The Intraclass Correlation Coefficient (ICC), which is the proportion of variance accounted for by school-level variables, among the total variance in kindergarten reading achievement, was 7.9%. The direct effect of the frequency of standardized tests was not associated with children's reading scores after controlling for the covariates ($\gamma_{05} = 0.32$, $SE = 0.28$, $p < 0.05$; see Fig. 3). Stated differently, children who were enrolled in kindergarten with frequent standardized tests did not

Table 3
Predicting Reading Achievement at Kindergarten (Unstandardized Coefficient).

Predictor	Reading Instructional Time (a_1)	Phonics (a_2)	Whole Language (a_3)	Balanced Approach (a_4)	Reading Achievement (Kindergarten spring)
	Estimate (SE)	*Estimate (SE)*	*Estimate (SE)*	*Estimate (SE)*	*Estimate (SE)*
Child Level					
Initial Reading Score					.94 (.01)***
Sex (female)					−0.69 (.14)***
SES					.90 (.13)***
Hispanic					−0.45 (.24)
Asian					.61 (.39)
Black					−0.68 (.34) *
White					.07 (.34)
English at Home					−0.64 (.25)*
Student Age in Months					−0.03 (.02)
Child with Disability					−1.22 (.21)***
School Level					
% of Minority Student					−0.02 (.01)***
% of Free Lunch					.02 (.01)***
Private School					1.15 (.52)*
Reporting ST					.32 (.29)
Frequency of ST	42.46 (9.73)***	.06 (.02)***	.13 (.03)***	.16(.03)***	.32(.29)
Teachers' Education					−0.46(.34)
Teachers' Experience					−0.06 (.02)**
Location of School					.19 (.14)
Reading Instructional Time					.01 (.01)*
Phonics					−1.09 (.52)*
Whole Language					.58 (.28)
Balanced Approach					.61 (.28)*
Residual variance	32540.40(1637.85)	.14(.04)	.25(.04)	.38(.04)	–

Note. ST[a] = Standardized Tests; Phonics = recognizing alphabet and letters, matching letters to sounds, writing own name; Whole Language = Identifying the main idea and parts of a story, communicating complete ideas orally, doing an activity or project related to a book or story, making predictions based on text, using context cues for comprehension, retelling stories; Balanced Approach = reading book with phonetic text, reading book with patterned text, reading book with controlled vocabulary. Weighted by W4C4P_20.

* $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

display higher reading scores after controlling for covariates. In the current model, 71.4% of the variance in children's kindergarten reading achievement was explained by child-level variables, whereas only 18.3% of the variance in children's reading achievement at kindergarten was explained by school-level variables.

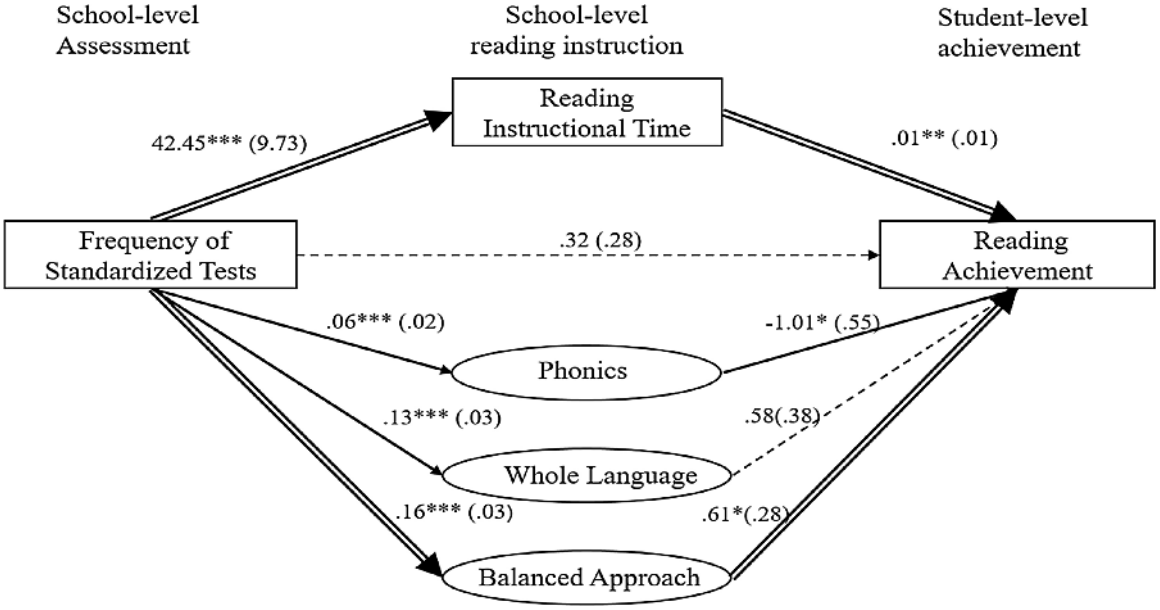

Fig. 3. Fixed-effects Model Depicting Relations Between Frequency of Standardized Testing Policy and Teacher's Reading Instruction and Reading Achievement at Kindergarten. This model fit the data well on most indices: χ^2 (Unweighted N = 12,241, df = 201, MLR scaling correction factor = 1.99) = 1465.71, p < 0.0001; CFI = 0.95; TLI = 0.92; RMSEA = 0.02; SRMR = 0.01(within), 0.11(between). 12,241 students nested in 1067 schools. Solid lines represent significant relations, whereas dashed lines represent non-significant relations. The double dashed line represents, pathway of mediated effect. The coefficients are unstandardized estimates and the values in parentheses denote the standard errors. *p < 0.05; **p < 0.01, ***p < 0.001.

3.2. The indirect effects: mediating role of reading instruction

The frequency of kindergarten standardized tests was significantly related to increased time on all types of reading instruction at the school level (See Fig. 3). For example, when the kindergartens took state/local standardized tests more frequently, these kindergartens tended to spend more instruction time on phonics ($\gamma_{21} = 0.06$, $SE = 0.02$, $p < 0.001$), on a balanced approach ($\gamma_{41} = 0.16$, $SE = 0.03$, $p < 0.001$), and on whole language, ($\gamma_{31} = 0.13$, $SE = 0.03$, $p < 0.001$). Results from a multilevel structural mediation model indicate that there was a mediated effect of frequency of kindergarten standardized tests on children's reading scores through reading instruction time ($\gamma_{11}\ \gamma_{01} = 0.09$, $SE = 0.04$, $p < 0.05$). The 99% asymmetric confidence interval from RMediation also supported this mediated effect of reading instruction time by not including zero in the confidence interval (Lower confidence limit = 0.01; Upper confidence limit = 0.19). Thus, kindergarten reading instruction time partially mediated the relationship between the frequency of kindergarten standardized testing and reading achievement at the end of kindergarten.

Additionally, there was a mediated effect of the frequency of standardized tests on reading achievement through a balanced approach ($\gamma_{41}\ \gamma_{01} = 0.10$, $SE = 0.05$, $p < 0.05$), but not through whole language ($\gamma_{31}\ \gamma_{01} = 0.07$, $SE = 0.05$, $p > 0.05$), nor through phonics ($\gamma_{21}\ \gamma_{01} = -0.06$, $SE = 0.03$, $p > 0.05$). The results from RMediation were consistent, in that the 99% confidence interval of a balanced approach did not include zero (Lower confidence limit = 0.01; Upper confidence limit = 0.20). Therefore, students who are enrolled in kindergartens with a higher frequency of standardized tests had significantly higher reading scores, mediated by an increase in a balanced approach toward reading instruction. However, the mediating effect of the frequency of kindergarten standardized testing was small, given the small coefficients and large sample size

4. Discussion

Drawing from data use theory, this study examined the direct and indirect effects of kindergarten standardized tests on children's concurrent reading achievement, mediated by reading instructional practices with a multilevel structural mediation model. While the existing body of research has primarily focused on the direct effects of state-level standardized testing on older children's academic outcomes (i.e., Grade 4), this is the first study, to the author's knowledge, that offers new insights on the mediational role of reading instruction that links the relationship between the frequency of kindergarten standardized testing and concurrent reading achievement during early childhood with nationally representative data.

4.1. Direct effect of the frequency of standardized tests on reading achievement

A major finding of this study is that children who went to kindergartens that frequently administered standardized tests did not perform any better, after controlling for child-level covariates and school-level covariates. This finding is inconsistent with previous studies that found strong accountability is associated with increased reading achievement during elementary school (Hanushek & Raymond, 2005; Kober et al., 2008; Rosenshine, 2003). There are two factors that may account for this discrepancies. First, it may be that the frequency of standardized tests may not motivate kindergarteners to learn because they do not understand the purpose of standardized tests (Goldstein & Flake, 2016; Meisels, 2007; Schultz et al., 2007). Second, it is possible that the effect of the frequency of standardized testing may be accurately captured in the current study as it included children with disabilities and English Language Learners with a multilevel approach (Lee & Reeves, 2012).

4.2. The mediating role of reading instruction

Another key finding of this study is that the association between the frequency of standardized testing and reading achievement was mediated by reading instruction time, after controlling for covariates. The finding that the frequency of testing was related to increased reading time was consistent with the previous literature (Center on Education Policy, 2008; Gallant, 2009; Miller & Almon, 2009). Further, as the intensity and duration of reading instruction time was related to improved reading achievement (Cavanaugh et al., 2004; Chatterji, 2005; Harn et al., 2008; Simmons et al., 2007; Sonnenschein et al., 2010), it may be that children had higher reading scores because children were exposed to contents and formats similar to the standardized tests, as reading instruction time was typically utilized for test preparation (Abu-Alhija, 2007; Crocker, 2005). However, it may also have been that teachers of kindergartens with frequent testing may have utilized test scores as "knowledge" to allocate reading instruction time, with the goal of improving student achievement, according to the data use theory (Hamilton et al., 2009; Mandinach & Gummer, 2015; Marsh & Farrell, 2015).

Consistent with previous literature, the frequency of standardized testing was significantly related to an increase in time for all types of reading instruction (i.e., phonics, whole language, and a balanced approach; Gullo & Hughes, 2011; Miller & Almon, 2009). What is notable is that kindergarten teachers responded to the frequency of standardized testing differently, specifically through variation in the types of instruction (e.g., phonics, whole language). This finding echoes data use theory, in that teachers' pedagogical content knowledge impacts the way teachers utilize assessment data (Mandinach & Gummer, 2015).

However, it is noteworthy that only one particular type of reading instruction (i.e., a balanced approach) mediated the connection between frequent standardized testing and reading achievement. This finding advances the current knowledge base because the balanced approach is a novel type of reading instruction in the newly released ECLS-K 2010–2011 data, and was not observed in the cohort of the ECLS-K 1998–1999 data (Sonnenschein et al., 2010; Xue & Meisels, 2004). It is possible that kindergarten teachers in the academic year of 2010–2011 may have chosen to teach with a balanced approach because the Common Core State Standards emphasize both decoding skills and reading comprehension (Caldwell, 2014). It may also be that a balanced approach may align well with the holistic measure of reading scores used in this current study.

Finally, the finding that no mediated effects were found in either phonics or whole language instruction requires further discussion. These findings suggest that merely implementing standardized tests frequently does not benefit children's learning when accompanied with ineffective reading instruction. On the one hand, phonics may not mediate the relationship between frequent testing and reading achievement in kindergarten, partly due to the nature of the measurement in this particular study. If the measurement of reading achievement focused on decoding skills, it is possible that phonics may have increased reading scores more (Goodman & Goodman, 2009). Conversely, it may be that whole language did not exhibit any mediating effect on young children, as whole language was more effective in promoting reading achievement, particularly for those children who have already acquired a certain level of literacy skills and knowledge (Morrison & Connor, 2002; Morrison et al., 2006; Sonnenschein et al., 2010; Xue & Meisels, 2004). Thus, it is possible those children who did not have sufficient decoding skills may not have received as much of a benefit from whole language instruction.

4.3. Limitations and suggestions for future studies

Due to the nature of secondary data analyses, there were several limitations in operationalizing critical concepts into study variables.

First, the teacher questionnaire did not specifically provide a definition of standardized tests. Thus, definitions of standardized tests were subjective and may have varied based on teachers' judgments. Second, the variations of teachers' instructional practices across classrooms were not well captured, as the current study utilized school-level reading instructional practices. Third, there might be a discrepancy between teachers' perceived practices and actual practices observed in the classroom, as the types of reading instruction were measured by teachers' self-report (Charlesworth et al., 1993; McMullen, 1999). Furthermore, the control variables in the current study only included the external school environment, without examining administrators' and teachers' pedagogical beliefs or conceptions regarding standardized testing.

4.4. Implications for policy and practices

The current study highlighted the mediational role of kindergarten reading instruction in the connection between the frequency of standardized testing and reading achievement. Consistent with data use theory, the findings suggest that frequent implementation of standardized tests alone does not benefit children's learning, unless mediated by effective reading instruction. The current findings underscore the importance of providing kindergarten teachers with ongoing and systematic professional development opportunities to help teachers utilize assessment data for student learning (Gullo, 2013; Marsh & Farrell, 2015; Means et al., 2009; Vanhoof & Van Petegem, 2007). As kindergartens have a unique place within the U. S. public school system for early childhood education (Gullo & Hughes, 2011; National Research Council, 2008), future studies should examine the long-term impact of testing policy on child outcomes holistically to determine appropriate implications for policy and practice.

References

Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation*, *33*(1), 50–68. http://dx.doi.org/10.1016/j.stueduc.2007.01.005.

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

Afflerbach, P. (2011). *Understanding and using reading assessment, K–12*. New York: International Reading Association.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, *10*(18), Retrieved 4/1/02 from http://epaa.asu.edu/epaa/v10n18.

Au, W. (2007). High-stakes testing and curricular control: A qualitative meta-synthesis. *Educational Researcher*, *36*(5), 258–267. http://dx.doi.org/10.3102/0013189X07306523.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, *2*(1), http://dx.doi.org/10.1177/2332858415616358 [2332858415616358].

Bauml, M. (2016). One size never fits all: Teachers' responses to standardized curriculum materials and implications for early childhood teacher educators. *Journal of Early Childhood Teacher Education*, *37*(1), 76–95.

Bitter, C., O'Day, J., Gubbins, P., & Socias, M. (2009). What works to improve student literacy achievement? An examination of instructional practices in a balanced literacy approach. *Journal of Education for Students Placed at Risk*, *14*(1), 17–44.

Boat, M., Zorn, D., & Austin, J. T. (2005). Assessment and program accountability in early childhood education: Lessons learned in Ohio. *NHSA Dialog*, *8*(1), 5–15. http://dx.doi.org/10.1207/s19309325nhsa0801_3.

Brown, K. D., & Goldstein, L. S. (2013). Preservice elementary teachers' understandings of competing notions of academic achievement coexisting in post-NCLB public schools. *Teachers College Record*, *115*(1), 1–37.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. http://dx.doi.org/10.1177/0049124192021002005.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mardinach, E., & Wexler, D. (2005). Linking data and learning: The grow network study. *Journal of Education for Students Placed at Risk*, *10*(3), 241–267.

Caldwell, J. S. (2014). *Reading assessment: A primer for teachers in the common core era*. New York: Guilford Publications.

Cavanaugh, C. L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal*, *2*(1), 9–21.

Center on Education Policy (2008). *NCLB year 5: instructional time in elementary schools: A closer look at changes for specific subjects*. Washington, DC: Center on Education Policy.

Charlesworth, R., Hart, C. H., Burts, D. C., Thomasson, R. H., Mosley, J., & Fleege, P. O. (1993). Measuring the developmental appropriateness of kindergarten teachers' beliefs and practices. *Early Childhood Research Quarterly*, *8*(3), 255–276.

Charlesworth, R., Fleege, P. O., & Weitman, C. J. (1994). Research on the effects of group standardized testing on instruction, pupils, and teachers: New directions for policy. *Early Education and Development*, *5*(3), 195–212.

Chatterji, M. (2005). Achievement gaps and correlates of early mathematics achievement: Evidence from the ECLS K—first grade sample. *Education Policy Analysis Archives*, *13*(46), 1–37. http://dx.doi.org/10.14507/epaa.v13n46.2005.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, *112*, 469–496.

Crocker, L. (2005). Teaching for the test: How and why test preparation is appropriate. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 159–174). Mahwah, New Jergey: L. Erlbaum Associates.

Dahl, K. L., Scharer, P. L., Lawson, L. L., & Grogan, P. R. (1999). Phonics instruction and student achievement in whole language first-grade classrooms. *Reading Research Quarterly*, *34*(3), 312–341.

Dee, T. S., & Jacob, B. (2011). The impact of No child left behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446 [0.1002/pam.20586].

Donat, D. J. (2006). Reading their way: A balanced approach that increases achievement. *Reading & Writing Quarterly*, *22*(4), 305–323.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.

Farris, P. J., Fuhler, C. J., & Walther, M. P. (2004). *Teaching reading: A balanced approach for today's classrooms*. McGraw-Hill Humanities Social.

Gallant, P. A. (2009). Kindergarten teachers speak out: Too much, too soon, too fast!. *Reading Horizons*, *49*(3), 201–220.

Goertz, M. E., & Duffy, M. C. (2001). *Assessment and accountability systems in 50 states: 1999–2000 (CPRE research report No. RR-046)*Philadelphia, PA: Consortium for Policy Research in Education.

Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems Educational Assessment. *Evaluation and Accountability*, *28*(3), 273–293. http://dx.doi.org/10.1007/s11092-015-9231-8.

Goodman, K. S., & Goodman, Y. M. (2009). Helping readers make sense of print: Research that supports a whole language pedagogy. In S. E. Israel, & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 99–114). NY: Routledge.

Gullo, D. F., & Hughes, K. (2011). Reclaiming kindergarten: Part I. Questions about theory and practice. *Early Childhood Education Journal*, *38*(5), 323–328. http://dx.doi.org/10.1007/s10643-010-0430-0.

Gullo, D. F. (2013). Improving instructional practices, policies, and student outcomes for early childhood language and literacy through data-driven decision making. *Early Childhood Education Journal*, *41*(6), 413–421.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., Wayman, J. C., ... Steele, J. L. (2009). *Using student achievement data to support instructional decision making*.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297–327 [10.3386/w10591].

Harn, B. A., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction: Does additional instructional time make a difference for the most at-risk first graders? *Journal of Learning Disabilities*, *41*(2), 115–125. http://dx.doi.org/10.1177/0022219407313586.

Hatch, J. A., & Grieshaber, S. (2002). Child observation and accountability in early childhood education: Perspectives from Australia and the United States. *Early Childhood Education Journal*, *29*(4), 227–231.

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. (2009). *A mandate for playful learning in preschool: Applying the scientific evidence*. Oxford: Oxford University Press.

Hornsby, D., & Wilson, L. (2010). *Teaching phonics in context*. Illinois: National Council of Teachers of English.

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

Hutchinson, C., & Young, M. (2011). Assessment for learning in the accountability era: Empirical evidence from Scotland. *Studies in Educational Evaluation*, *37*(1), 62–70. http://dx.doi.org/10.1016/j.stueduc.2011.03.007.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*, *35*(4), 458–492. http://dx.doi.org/10.1598/rrq.35.4.2.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvements: Actions, outcomes, and lessions from three urban districts. *American Journal of Education*, *112*(496), 520.

Kober, N., Chudowsky, N., & Chudowsky, V. (2008). *Has student achievement increased since 2002? State test score trends through 2006-07*. Washington, DC: Center on Education Policy.

Kontovourki, S. (2009). *Becoming literate in assessment-saturated classrooms: Elementary students' literacy performances at the intersection of literacy curricula and testing (Unpublished doctoral dissertation)*. New York: Teachers College Columbia University.

Krashen, S. (2002). Defending whole language: The limits of instruction and the efficacy of whole language instruction. *Reading Improvement*, *39*(1), 32–42.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources state NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209–231. http://dx.doi.org/10.3102/0162373711431604.

Lipson, M. Y., Goldhaber, J., Daniels, P., & Sortino, S. (1994). Literacy practices in kindergarten: Competing interests and multiple practices. *Paper presented at the 39th annual convention of the international reading association*.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. http://dx.doi.org/10.1207/

s15327906mbr3901_4.
MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*(3), 384–389. http://dx.doi.org/10.3758/bf03193007.
Mandinach, E. B., & Gummer, E. (2015). Data-driven decision making: Components of the enculturation of data use in education. *Teachers College Record*, *117*(4) [1–50].
Mandinach, E. B., Honey, M., & Light, D. (2006). A theoretical framework for data-driven decision making. *Paper presented at the annual meeting of the american educational research association* [April].
Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making A framework for understanding capacity building. *Educational Management Administration & Leadership*, *43*(2), 269–289.
Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: The Rand Corporation. Retrieved from http://www.rand.org/pubs/occasional_papers/OP170. html.
Martella, J. (2010). The words we use: A glossary of terms for early childhood education standards and assessments. In D. Koralek (Ed.), *Spotlight on young children and assessment* (pp. 42–44). Washington: National Association for the Education of Young Children.
Sage Handbook of research on classroom assessment. In J. H. McMillan (Ed.), Thousand Oaks: Sage Publications. http://dx.doi.org/10.4135/9781452218649.
McMullen, M. B. (1999). Characteristics of teachers who talk the DAP talk and walk the DAP walk. *Journal of Research in Childhood Education*, *13*(2), 216–230.
McMurrer, J., & Kober, N. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington. DC: Centre on Education Policy.
Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-Informed decision making in schools: Teacher access, supports and use*. US Department of Education.
Meisels, S. J. (2007). Accountablity in early chidlhood: No easy answers. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness, early learning, and the transition to kindergarten* (pp. 31–48). Balitmore: Paul H. Brookes.
Merki, K. M. (2011). Special issue: Accountability systems and their effects on school processes and student learning. *Studies in Educational Evaluation*, *37*(4), 177–179. http://dx.doi.org/10.1016/j.stueduc.2012.03.002.
Miller, E., & Almon, J. (2009). *Crisis in the kindergarten: Why children need to play in school*. NJ: Alliance for Childhood. Retrieved from ERIC Database [ERIC Document 504839].
Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology*, *40*(6), 493–500. http://dx.doi.org/10.1016/s0022-4405(02)00127-9.
Morrison, F. J., Connor, C. M., & Bachman, H. J. (2006). The transition to school. In D. K. Dickinson, & S. B. Neuman (Eds.), *Handbook of early literacy research* (pp. 375–394). New York: Guilford Press.
Muthén, B. O., & Muthén, L. K. (2012). *Mplus version 7: user's guide*. CA: Muthén & Muthén.
Muthén, B., Du Toit, S. H., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*, *75*, 1–45.
National Association for the Education of Young Children (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective accountable system in programs for children birth through age 8*. Washington, DC: National Association for the Education of Young Children.
National Association of Early Childhood Specialists in State Department of Education (2000). *STILL unacceptable trends in kindergarten entry placement: A position statement developed by the National Association of Early Childhood Specialists in State Departments of Education*. Washington, DC: National Association of Early Childhood Specialists in State Department of Education.
National Center for Education Statistics (2009). *ECLS-K longitudinal kindergarten-eighth grade public-use child file, CD-ROM and user' manual*. Washington, DC: U.S . Department of Education.
National Early Childhood Accountability Task Force (2007). *Taking stock: Assessing and improving early childhood learning and program Quality*. [Retrieved from http] wwww.pewtrusts.org.
National Research Council (2008). *Early childhood assessment: Why, what and how*. Washington, D.C: The National Academies Press.
Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, *14*(1), 1–58. http://dx.doi.org/10.14507/epaa.v14n1.2006.
Pearson, P. D. (2004). The reading wars. *Educational Policy*, *18*(1), 216–252. http://dx.doi.org/10.1177/0895904803260041.
Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209.
Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*(2), 161–182.
Pyle, A., & DeLuca, C. (2013). Assessment in the kindergarten classroom: An empirical study of teachers' assessment approaches. *Early Childhood Education Journal*, *41*(5), 373–380.
Ravitch, D. (2011). *The death and life of the great american school system: How testing and choice are undermining education*. NY: Basic Books.
Roberts, T. A., & Meiring, A. (2006). Teaching phonics in the context of children's literature or spelling: Influences on first-grade reading, spelling, and writing and fifth-grade comprehension. *Journal of Educational Psychology*, *98*(4), 690–713. http://dx.doi.org/10.1037/0022-0663.98.4.690.
Rosenshine, B. (2003). High-stakes testing: Another analysis. *Education Policy Analysis Archives*, *11*, 1–8. http://dx.doi.org/10.14507/epaa.v11n24.2003.
Rous, B., McCormick, K., Gooden, C., & Townley, K. F. (2007). Kentucky's early childhood continuous assessment and accountability system: Local decisions and state supports. *Topics in Early Childhood Special Education*, *27*(1), 19–33. http://dx.doi.org/10.1177/02711214070270010201.
Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482–496.
Schultz, T., Kagan, S. L., & Shore, R. (2007). *Taking stock: Assessing and improving early childhood learning and program quality*. Washington DC: Pew Charitable Trusts. Retrieved October, 06, 2016 from http://policyforchildren.org/wp-content/uploads/2013/07/Taking-Stock.pdf.
Shen, J., & Cooley, V. E. (2008). Critical issues in using data for decision-making. *International Journal of Leadership in Education*, *11*(3), 319–329.
Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., ... Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities*, *40*(4), 331–347. http://dx.doi.org/10.1177/00222194070400040401.
Solley, B. A. (2007). On standardized testing: An ACEI position paper. *Childhood Education*, *84*(1), 31–37. http://dx.doi.org/10.1080/00094056.2007.10522967.
Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. *American Educational Research Journal*, *47*(2), 358–389. http://dx.doi.org/10.3102/0002831209349215.
Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*(2), 113–141.
Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*(3), 692–700. http://dx.doi.org/10.3758/s13428-011-0076-x.
Tourangeau, K., Lê, T., Hagedorn, M. C., Daly, P., & Najarian, M. (2015). *Early childhood longitudinal study, kindergarten class of 2010–11 (ECLS-K: 2011). user's manual for the ECLS-K: 2011 kindergarten data file and electronic codebook, public version. NCES 2015-074*. National Center for Education Statistics.
Vanhoof, J., & Van Petegem, P. (2007). Matching internal and external evaluation in an era of accountability and school development: Lessons from a Flemish perspective. *Studies in Educational Evaluation*, *33*(2), 101–119. http://dx.doi.org/10.1016/j.stueduc.2007.04.001.
Wilson, D. M. (2009). Developmentally appropriate practice in the age of testing. *Pediatrics*, *123*(2), 431–436.
Xue, Y., & Meisels, S. J. (2004). Early literacy instruction and learning in kindergarten: Evidence from the early childhood longitudinal study- kindergarten class of 1998–1999. *American Educational Research Journal*, *41*(1), 191–229. http://dx.doi.org/10.3102/00028312041001191.
Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, *112*(4), 521–548.
Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models problems and solutions. *Organizational Research Methods*, *12*(4), 695–719 [10.1177/1094428108327450].

Haesung Im Ph.D. is currently working as an instructor in the Department of Early Childhood Education at the Ewha Womans University in South Korea. She received her doctoral degree in Curriculum and Instruction at Arizona State University in the United States. She is a dedicated researcher and teacher educator who has an expertise in early childhood assessment, teacher performance assessment and curriculum and instruction