



## Comparing the Potential Utility of Kindergarten Entry Assessments to Provide Evidence of English Learners' Knowledge and Skills

*Debra J. Ackerman*



**POLICY INFORMATION REPORT**

This Policy Information Report was written by:

**Debra J. Ackerman**  
Educational Testing Service, Princeton, NJ

Policy Information Center  
Mail Stop 19-R  
Educational Testing Service  
Rosedale Road Princeton, NJ 08541-0001  
(609) 734-5212  
pic@ets.org

Copies can be downloaded from: [www.ets.org/research/pic](http://www.ets.org/research/pic)

The views expressed in this report are those of the authors and do not necessarily reflect the views of the officers and trustees of Educational Testing Service.

### **About ETS**

At ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English language learning, and elementary, secondary and postsecondary education, and by conducting education research, analysis and policy studies. Founded as a nonprofit in 1947, ETS develops, administers and scores more than 50 million tests annually—including the *TOEFL*<sup>®</sup> and *TOEIC*<sup>®</sup> tests, the *GRE*<sup>®</sup> tests and *The Praxis Series*<sup>®</sup> assessments—in more than 180 countries, at over 9,000 locations worldwide.

## RESEARCH REPORT

# Comparing the Potential Utility of Kindergarten Entry Assessments to Provide Evidence of English Learners' Knowledge and Skills

Debra J. Ackerman

Educational Testing Service, Princeton, NJ

In this report I share the results of a document-based, comparative case study aimed at increasing our understanding about the potential utility of state kindergarten entry assessments (KEAs) to provide evidence of English learner (EL) kindergartners' knowledge and skills and, in turn, inform kindergarten teachers' instruction. Using a sample of 9 purposely selected state KEAs and their respective policies, the focus of the study was to what extent these measures contain items that are specific to ELs, allow or mandate the use of linguistic accommodations, and have policies on KEA assessor or observer linguistic capacity. Also of interest was the degree to which research supports the use of these measures with EL students. The study's results suggest that the 9 sample KEAs represent 3 different profiles along a less-to-more continuum of EL-specific content, linguistic accommodations and assessor/observer linguistic capacity policies, and EL-relevant research. These results have implications for policymakers who are tasked with selecting or developing a KEA aimed at informing kindergarten teachers' instruction as well as the future research to be conducted on KEAs designed for this purpose.

**Keywords** Kindergarten entry assessments; English learners; linguistic accommodations; assessor linguistic capacity policies

doi:10.1002/ets2.12224

Two early education “hot topics” across the United States are kindergarten entry assessments (KEAs), which typically are administered in the first few months of school as a means for informing teachers' instruction (Ackerman, 2018), and the growing population of students entering the nation's schools who are considered to be dual language or English learners (ELs),<sup>1</sup> 77% of whom speak Spanish as their home language (McFarland et al., 2017). Recent interest in KEAs stems in part from federal Race to the Top-Early Learning Challenge (RTT-ELC) awards to 20 states as well as Enhanced Assessment Grants awarded to two consortia comprising 17 states (including some RTT-ELC states) and one additional state (Early Learning Challenge Technical Assistance, 2016). These funding competitions also highlighted the number and percentage of children ages 0–5 who were identified as ELs in applicant states (Ackerman & Tazi, 2015) as well as the need for valid and reliable assessment data to support the learning of all kindergartners (Maryland State Department of Education, 2013; North Carolina Department of Public Instruction, 2013; Texas Education Agency, 2013).

The focus on ELs and KEAs is salient. Although being bilingual, or having the ability to speak and understand two languages, can benefit young children's learning and development (Ackerman & Friedman-Krauss, 2017), analyses of Early Childhood Longitudinal Study data demonstrate gaps in EL kindergartners' early reading and math skills as compared to their non-EL peers (Han, Lee, & Waldfogel, 2012). Such gaps can persist as children progress through school, with Grades 4 and 8 National Assessment of Education Progress data revealing further differences in students' mathematics and reading scores (McFarland et al., 2017). KEAs have the potential to mitigate these emerging gaps by highlighting kindergartners' strengths and areas needing more support and, in turn, informing teachers' practice. However, various construct-irrelevant issues can challenge the validity and reliability of KEA data (Ackerman, 2018). Moreover, although a variety of socioeconomic factors contribute to ELs' unequal academic outcomes, one critical contributor is teachers' capacity to effectively respond to their learning needs (Takanishi & Le Menestrel, 2017).

Given this interrelated policy context, it is important from an educational equity perspective to examine to what extent KEAs are valid and reliable for all kindergartners (Goldstein & Flake, 2016). Such an examination can shed light on needed EL-related education policy and practice revisions as well (Abedi, Hofstetter, & Lord, 2004; Robinson-Cimpian, Thompson, & Umansky, 2016). However, researchers have not yet looked across specific state KEAs and hypothesized to

*Corresponding author:* D. Ackerman, E-mail: dackerman@ets.org

what extent these measures are likely to produce valid and reliable evidence of what EL students know and can do, much less tested out such hypotheses in a methodologically rigorous manner.

With the aim of increasing our understanding about the potential utility of KEAs to provide evidence of EL kindergartners' knowledge and skills, in this report I share the results of a comparative case study of nine state KEAs. To begin, I provide a brief overview of the current KEA policy context. I then highlight some potential validity and reliability challenges when KEA data are to be used to inform kindergarten teachers' instruction. After discussing the study's results, I conclude with some implications for policymakers and researchers.

### Current KEA Context

Teachers' capacity to differentiate their instruction is an important aspect of improving student learning outcomes (Tomlinson, 2008). For example, some kindergartners enter school with a firm grasp of letter-sound knowledge and are able to read a few simple words, whereas other students are still learning upper- and lower-case letter names (Bassok & Latham, 2017). Therefore, to plan for instruction, kindergarten teachers need accurate data on each student's current knowledge and skills (Al Otaiba et al., 2011; Moon, 2005).

One source of such data is KEAs, which are being developed, piloted, field tested, or implemented in at least 40 states and the District of Columbia (The Center on Standards & Assessment Implementation, 2017). These measures generally are administered in the first few months of kindergarten, and in some cases, they include a family survey or meeting to help teachers learn more about a student's background, such as their prior learning experiences in group settings. However, KEAs also differ. They include commercially available and state-developed instruments and can reflect a direct assessment or observational rubric approach. In some states, the KEA mandated for use reflects both of these assessment approaches. In addition, the role of the kindergarten teacher in the assessment process can vary. Some direct assessments are administered by a teacher on a one-on-one basis, while other assessments are computer administered. Observational measures typically require teachers to collect evidence of a student's knowledge or skill via notes, photos, and work samples. KEAs also differ in terms of the focus of their domains (e.g., language, literacy, mathematics, etc.), total number of items, prescribed administration periods, and purposes for which these data are collected (Ackerman, 2018).

Research on KEAs over the past decade has investigated the psychometric properties of individual KEAs (e.g., Ferrara & Lambert, 2015, 2016; Goldstein, McCoach, & Yu, 2017; Irvin, Tindal, & Slater, 2017; Tindal, Irvin, Nese, & Slater, 2015) and initial KEA implementation issues (Golan, Woodbridge, Davies-Mercier, & Pistorino, 2016). Additional studies have explored teachers' perspectives on administering these measures and to what extent they are useful for informing instruction (e.g., Ohle & Harvey, 2017; Schachter, Strang, & Piasta, 2017). Ackerman (2018) conducted case studies of seven KEAs that were developed, piloted, and field-tested with financial support from the RTT-ELC competition. This research highlighted some of the assessment- and teacher-related validity and reliability issues that prompted modifications to the content and administration policies of these KEAs. A key takeaway from all of these studies is that implementing a KEA policy may not be sufficient for ensuring that the measure's data can effectively inform teachers' practice and contribute to efforts to close school readiness gaps.

### Potential EL-Related KEA Validity and Reliability Issues

Two key factors that both policymakers tasked with developing or selecting an assessment and assessment data users need to keep in mind are validity and reliability. *Validity* refers to the extent to which the interpretation of a measure's scores provides sufficient evidence to inform a specific purpose and population of students (Bonner, 2013; Kane, 2013). The focus on purpose and population is important, as the inferences made from these scores can be more or less valid depending on the purpose for which the test is used (e.g., inform teachers' instruction vs. determine the extent to which all kindergartners meet certain learning benchmarks) and the population for which inferences are made (ELs vs. all kindergartners). *Reliability* refers to the extent to which a measure provides consistent results over different assessors, observers, testing occasions (assuming there is a minimal time lapse between testing occasions and there has been no change in students' underlying ability), and test forms (Livingston, 2018).

Although validity and reliability are complex issues that are deserving of their own report, three issues have particular relevancy to the potential utility of KEAs to help teachers meet their EL students' learning needs: the content of a measure, the linguistic accommodations that may be used when assessing or observing ELs and teachers' capacity to implement

those accommodations, and prior research demonstrating that a measure can provide valid and reliable evidence of what EL students know and can do.

## KEA Content

One key validity and reliability topic to consider when examining the potential utility of a measure is to what degree its items sufficiently focus on the constructs that are considered to be critical to informing the purpose for collecting the data in the first place. A common aim of collecting KEA data is to inform teachers' instruction in specific academic domains (e.g., English language arts or mathematics). Moreover, the specific topics to be covered within each of these domains often need to reflect a state's kindergarten learning standards. Therefore, assessment developers or those tasked with selecting a KEA typically investigate to what extent the measure contains age- and grade-appropriate items that are aligned with these standards (Roach, McGrath, Wixson, & Talapatra, 2010).

In states with English language proficiency or development standards, another aim for collecting KEA data may be to help teachers determine the language and literacy development of students identified as ELs via home language surveys or classroom placement tests (Gillanders & Castro, 2015). The standards aimed at upper elementary and secondary students can admittedly vary in terms of the extent to which they emphasize proficiency in individual academic content domains (Lee, 2018). However, at the kindergarten level, these standards tend to focus on a student's capacity to understand oral commands and read-alouds, participate in conversations with their peers and teachers, and use key English words when discussing specific academic topics (Council of Chief State School Officers, 2014).

Some states also are integrating what are known as learning progressions into their English language proficiency or development standards as a means for improving teachers' EL-focused instruction (Bailey & Heritage, 2014). Similarly, to provide teachers with actionable data, KEA items related to these standards should reflect the different developmental trajectories of second language acquisition. For example, some young ELs may learn their home language and English simultaneously, whereas other children may enter kindergarten as monolingual speakers of a non-English language. In addition, depending on the different linguistic contexts in which they interact with their families or in the community, young children will vary in their social and academic language proficiency as well as their levels of expressive and receptive language in either type of language. As a result, ELs may engage in *code switching* or *code mixing*, meaning a student may use words or grammar from more than one language to ask or respond to a question or while talking with peers, their teacher, or other adults (Guzman-Orth, Lopez, & Tolentino, 2017; Hoff, 2013).

No matter what the focus of a KEA, it is important to remember that entering kindergartners will typically exhibit a wide range of beginner to more advanced skills (Bernstein, West, Newsham, & Reid, 2014). Therefore, to support a KEA's validity and reliability for informing teachers' practice, both assessment developers and policymakers need to ensure that these items are broad enough to document the wide range of children's skills at kindergarten entry (Karelitz, Parrish, Yamada, & Wilson, 2010; Quirk, Nylund-Gibson, & Furlong, 2013). If a measure only reflects beginner or advanced skills, it likely will not be appropriate for assessing all of the students in a teacher's classroom (Johnson & Buchanan, 2011; Jonas & Kassner, 2014).

## Linguistic Support Accommodations

Another key validity and reliability issue to consider when selecting or developing a KEA to be used with ELs is the *accommodations* that may be used. This term refers to the changes that may be made to an assessment's administration or the ways in which a student may participate in a direct assessment or display their knowledge and skills as part of an observational measure's evidence-gathering process. In either case, the purpose of using the accommodation is to mitigate construct-irrelevant issues that might otherwise impact the validity and reliability of a student's scores for a particular purpose (Pitoniak et al., 2009).

Typically assessment accommodations are made for students with identified special needs. In this case, the changes could include extending the testing or response time or adjusting how a test is presented (e.g., enlarging the size of the print; Division for Early Childhood of the Council for Exceptional Children, 2007; Thurlow, 2014). When thinking about potential accommodations that are responsive to the needs of ELs, it is helpful to consider that a major source of variation in student scores can be proficiency in the language in which an assessment is given (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). For example,

an EL kindergartner may easily count 10 items when prompted to do so in Spanish but not respond to the request when issued in English. If the item aimed to generate evidence of the student's cardinality skills, two likely construct-irrelevant factors are his or her English receptive and expressive vocabulary levels.

To mitigate the potential effect of these “confounders,” policymakers may consider using what are known as direct linguistic supports (Wolf et al., 2008). The research base on the effectiveness of such supports is admittedly not yet unequivocal and mainly focuses on students in Grade 4 and above (Kieffer, Lesaux, Rivera, & Francis, 2009; Kieffer, Rivera, & Francis, 2012). Yet, additional research suggests the potential usefulness of three specific linguistic supports to support the validity of KEAs aimed at informing kindergarten teachers' instruction so that they can best meet their EL students' learning needs.

The first accommodation is providing students with the directions for participating in a direct assessment in his or her home language and English (Abedi et al., 2004). This accommodation makes intuitive sense when considering that the directions for computer-administered tests may include how to use the mouse, keyboard, or touchscreen. Research also demonstrates that young children are not equally proficient with, or experienced in using different digital devices (Barnes, 2015). Moreover, a student's level of digital familiarity and fluency can impact that student's capacity to demonstrate his or her knowledge and skills in a valid and reliable way (Dady, Lyons, & DePascale, 2018).

A second potential linguistic accommodation is translating the specific item prompts to which kindergartners must respond when being assessed with a direct KEA. Analyses of a nationally representative, large-scale database suggests that such translations may be particularly critical when assessing young ELs' mathematics knowledge and skills (Robinson, 2010). This finding also makes intuitive sense when considering young children's difficulty in mastering written mathematics symbols (Ginsburg, Lee, & Boyd, 2008). However, for all content areas, attention must be paid to providing accurate and culturally relevant translations (Educational Testing Service, 2016; Mihai, 2010; Turkan, Oliveri, & Cabrera, 2013).

Because young ELs may be more comfortable using a mix of English and their home language (García, 2011), a third potential accommodation to be considered—and one that can be used with both direct and observational KEAs—is allowing students to use the language or languages in which they are most proficient to demonstrate their knowledge and skills. Such an accommodation might also include allowing a student to use gestures, such as nodding or pointing (Guzman-Orth et al., 2017). The research base on this support also is in its early stages (Lopez, Turkan, & Guzman-Orth, 2017).

Of course, a key related issue to consider is teachers' capacity to implement these linguistic accommodations in a valid and reliable way (and a point to which I return in the next section). In fact, the National Association for the Education of Young Children (2009) urges young ELs to be assessed by staff who are not only fluent in a child's home language but also familiar with preferred cultural interaction styles. A similar cultural background may be particularly essential when assessing young children's social–emotional development (Espinosa & López, 2007). Adults who assess young ELs also may need a thorough understanding of bilingual language acquisition so that they can distinguish between inadequate content knowledge and a student's lack of English language or cultural proficiency (Alvarez, Ananda, Walqui, Sato, & Rabinowitz, 2014; Bedore & Peña, 2008).

## Research Supporting a KEA's Validity and Reliability for Teachers' Instruction of ELs

A final—and critical—issue to be considered prior to selecting and using any assessment is to identify the extent of adequate research evidence demonstrating its validity and reliability for a specific purpose and population (Snow & Van Hemel, 2008). This topic is also complex (Sanchez et al., 2013). In addition, this research will vary based on the different individuals and systems involved in the entire testing and data use process and whether the measure is a direct assessment or observational rubric. Also factoring into decisions about the research to be conducted are how often an assessment is administered and the array of purposes for which the measure is being used (Mislevy, Wilson, Ercikan, & Chudowsky, 2003).

However, two key topics have particular relevance to KEAs that are used to inform kindergarten teachers about their incoming EL students' knowledge and skills. First, young children's direct assessment responses or their knowledge and skills as defined by an observational measure can be influenced by their English language proficiency or the community- or ethnic-related culture in which they live (Rollins, McCabe, & Bliss, 2000; Snow & Van Hemel, 2008). Therefore, one key research topic is the extent to which an assessment accurately measures the same constructs across groups of children or, conversely, differs among groups of children with otherwise equal ability (Guhn, Gadermann, & Zumbo, 2007; Huang

& Konold, 2014). To investigate this topic, researchers typically analyze data from a representative sample of students from different demographic, linguistic, and cultural backgrounds (Quirk, Rebelez, & Furlong, 2014; Vogel et al., 2008). When there is concern about a particular population (e.g., ELs), the sample also should include students with similar characteristics (Acosta, Rivera, & Willner, 2008; Espinosa & García, 2012; Takahashi & Le Menestrel, 2017). If students' scores appear to be significantly different, additional analyses are needed to determine if the difference is due to item bias, or conversely, an attribute of the subgroup, such as their English language capacity or targeted ability (Guhn et al., 2007; Najarian, Snow, Lennon, Kinsey, & Mulligan, 2010).

To further confirm the adequacy of a measure's scores for providing evidence of EL students' knowledge and skills, researchers can investigate the relationship between this population's scores on the measure of interest and other measures that focus on similar constructs. For example, if an observational KEA measures children's early mathematics skills, researchers could compare teachers' ratings on that measure and the scores on a direct assessment of these same skills. If the scores for the two measures appear to be correlated, those tasked with selecting the target measure can have greater confidence that the KEA is valid for that particular purpose and population (Goldstein et al., 2017; Howard et al., 2017).

The second KEA-relevant topic is teacher reliability when administering a measure and producing scores (Wakabayashi & Beal, 2015; Waterman, McDermott, Fantuzzo, & Gadsden, 2012). Establishing appropriate reliability levels promotes consistency within and across classrooms and provides confidence that children's assessment scores reflect their performance on a particular occasion rather than the capacity of the assessor (R. J. Wright, 2010). In the case of KEAs specifically, rater reliability is particularly important if the data are also used to inform district- or state-level decisions. This topic is especially relevant when using observational measures, as score reliability can be impacted by teacher beliefs, perceptions regarding students' skills, or even concerns regarding how the scores will be used (Cash, Hamre, Pianta, & Myers, 2012; Harvey, Fischer, Weieneth, Hurwitz, & Sayer, 2013; Mashburn & Henry, 2005; Waterman et al., 2012; Williford, Downer, & Hamre, 2014). While such bias is rarely intentional, it may result in over- or underestimating children's skills. A closely related issue is whether an observational measure's authors provide sufficient guidance for what constitutes valid and sufficient evidence upon which to base a judgment (Dever & Barta, 2001; Goldstein & McCoach, 2011; Ohle & Harvey, 2017). Such guidance may be particularly critical for items focused on EL students' language development (Gillanders & Castro, 2015).

In summary, three EL-relevant validity and reliability topics are (a) the content of the measure, (b) the linguistic support accommodations that may be used and the capacity of teachers to implement those changes, and (c) the research supporting the use of these measures with this population of students. These topics may be particularly important if one aim of collecting KEA data is to inform teachers' instruction of students who are at greater risk for experiencing academic achievement issues due to their EL status.

## The Current Study

A growing research base focuses on KEAs, but researchers have not yet had the opportunity to look across these measures and consider their potential to accurately provide evidence of EL kindergartners' knowledge and skills. In light of the increased population of at-risk ELs entering kindergarten and the key role assessment data can play in helping teachers differentiate their instruction in response to students' learning needs, this lack of research is an important educational equity issue. With the aim of increasing our understanding about the potential utility of KEA data to provide evidence of what EL kindergartners know and can do, I conducted a comparative case study (Bartlett & Vavrus, 2017) of nine states' KEAs and their relevant policies and research. This approach is particularly well suited to the study's aim, because although each KEA is governed by its respective state's policies, the measures can be compared due to sharing similar components and goals. Three research questions guided my study:

- 1 Which KEA items are specific to ELs?
- 2 What are the linguistic accommodations that may or must be used? In turn, what are states' policies regarding the linguistic capacity needed by KEA assessors or observers when using these accommodations?
- 3 To what extent does test validity research or, if applicable, observer reliability research support the use of these KEAs with EL kindergartners?

## Methodology

### Sample

The sample for my study consisted of the KEAs and related policies used in nine states in the 2017–2018 school year. At least 40 states and the District of Columbia are at some stage of the KEA implementation process (The Center on Standards & Assessment Implementation, 2017), but given my use of a comparative case study design, I wanted to limit my sample size yet also examine KEAs that reflected the broader U.S. policy context. I therefore purposefully selected these nine KEAs because they represented the three assessment approaches being used across the United States for the purpose of informing kindergarten teachers' instruction in the first few weeks of school. In addition, these nine KEAs offered the opportunity to compare measures with similar content within each approach (see Table 1).

For example, the first five KEAs are teacher-administered observational measures. In addition, California's Desired Results Developmental Profile (DRDP-K; California Department of Education, 2015a) was developed in collaboration with the Illinois State Board of Education, where the nearly identical KEA is known as the Kindergarten Individual Development Survey (KIDS; California Department of Education Child Development Division, 2017; Illinois State Board of Education, 2015). Delaware and Washington each use a state-customized subset of items from the Teaching Strategies GOLD observational measure (Heroman, Burts, Berke, & Bickart, 2010), which has served as the basis for additional states' KEAs (Weisenfeld, 2016) and been widely used in state-funded PreK programs (Ackerman & Coley, 2012).

Pennsylvania's Kindergarten Entry Inventory (Pennsylvania Office of Child Development and Early Learning, 2017b), a state-developed observational measure, does not share a KEA source with any of the other measures in the sample but is similar in approach to the DRDP-K.

Florida's and Mississippi's respective KEAs are computer-administered, direct assessments and consist of items from the commercially available Star Early Literacy direct assessment (Renaissance Learning, 2017a). Both measures focus on language, early literacy, and mathematics. The final two measures rely on both a teacher-administered direct assessment and an observational rubric. Oregon's Kindergarten Assessment (Office of Teaching, Learning, & Assessment, Oregon Department of Education, 2017) is comprised of English letter name and letter sound items, math items written by Oregon educators or adapted from the easyCBM (Anderson et al., 2014), and the Child Behavior Rating Scale (CBRS)

**Table 1** Sample Kindergarten Entry Assessments (KEAs)

KEA	Fall 2017 Administration Domains								Age 0–8 ELs (%)
	Language	Literacy	Math	Science	Self-regulation/ approach to learning	Social– emotional	Physical	Other	
Teacher-administered observational rubric									
CA Desired Results Developmental Profile (DRDP-K)	X	X	X	X	X	X	X	X	60
DE Early Learner Survey	X	X	X		X	X	X	X	13 <sup>5</sup>
IL Kindergarten Individual Development Survey (KIDS)	X	X	X		X	X			34
PA Kindergarten Entry Inventory	X	X	X		X	X	X		19
WA Kindergarten Inventory of Developing Skills (WaKIDS)	X	X	X		X	X	X		32
Computer-administered direct assessment									
FL Kindergarten Readiness Screener (FLKRS)	X	X	X						40
MS Star Early Literacy	X	X	X						4 <sup>6</sup>
Teacher-administered direct assessment and observational rubric									
OR Kindergarten Assessment		X	X		X	X			28
UT Kindergarten Entry and Exit Profile (KEEP)		X	X		X	X			22

*Note.* EL = English learner; CA = California; DE = Delaware; IL = Illinois; PA = Pennsylvania; WA = Washington; FL = Florida; MS = Mississippi; OR = Oregon; UT = Utah. *N* = 9.

observational rubric (Bronson, Goodson, Layzer, & Love, 1990), which focuses on children's classroom self-regulation and interpersonal skills. Utah's new state-developed Kindergarten Entry and Exit Profile (Utah State Board of Education, 2017c) also relies on both direct items and an observational rubric, with the focus of the latter measure limited to students' behavior while participating in the KEA testing process.

I also selected these KEAs because they represent variations in two EL-relevant state policy contexts. First, these states rely on different kindergarten English language development or proficiency standards for ELs. Washington and Oregon belong to the English Language Proficiency Assessment for the 21st Century (ELPA21)<sup>2</sup> consortium, whereas Delaware, Florida, Illinois, Pennsylvania, and Utah are members of WIDA.<sup>3</sup> California has its own kindergarten English language development standards (California State Board of Education, 2012), and Mississippi uses the Teachers of English to Speakers of Other Languages (TESOL) PreK-12 English Language Proficiency standards (Mississippi Department of Education Office of Elementary Education and Reading, Student Intervention Services Pre-K-12, 2016). Second, the percentage of young children who are reported as being ELs in the KEAs' respective states varies, with the rates ranging from 4% in Mississippi to 60% in California (Delaware Department of Education, 2014; Park, O'Toole, & Katsiaficas, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g; personal communication, Mississippi Department of Education, May 4, 2018).

## Data Sources and Analysis

To address the study's research questions, I relied exclusively on document-based data. Documents can serve as useful sources of descriptive data (Yin, 2014), particularly when analyzing legislation-based education policies (Bowen, 2009; Gibton, 2016). The documents I analyzed to address Research Questions 1 and 2 included the nine sample KEAs used during the 2017–2018 school year, their administration manuals and score sheets, and if not indicated in the KEA itself, their respective states' accommodation guidelines. These documents were obtained from state departments of education and assessment developers' websites.

To address Research Question 3, I relied on all of the publicly available research I could identify that used a sample of preschool- or kindergarten-aged ELs and addressed the validity of these KEAs (or their previous iterations) for demonstrating what EL students know and can do or, in the case of the observational KEAs, was relevant to observer reliability. This research included psychometric studies and annual KEA score reports. Some of this research was available on the state department of education websites or available through a URL included in the KEAs or their related technical, administration, or accommodation documents. I identified other research by searching scholarly journal websites. However, I also included reports published by university-based or private research institutions. This latter source was particularly important given the involvement of these institutions in the development and ongoing refinement of some of my sampled KEAs.

Analysis of the documents used to inform Research Questions 1 and 2 followed the approach advocated by Bowen (2009). First, I read each KEA and its related administration and accommodations documents to identify any EL-specific items, allowable linguistic accommodations, and the policies on assessor/observer linguistic capacity when using these accommodations. I then recorded this information as well as the document in which it was found in an Excel spreadsheet; each row represented a single state/KEA, and the columns represented the research question categories. Due to my small sample size, my analysis of these data consisted of generating simple cross-case descriptive statistics in terms of which KEAs had EL-specific items, any of the three types of linguistic accommodations mentioned in my literature review, and an articulated assessor/observer linguistic policy.

To address Research Question 3, my analyses were aimed at determining the relative strength of the research base on a KEA's validity for demonstrating what EL students know and can do, as well as observer reliability. I therefore elected to characterize this research using the terms "strong," "moderate," or "not available." Although these terms are admittedly subjective, I took into consideration not only the quantity of studies, but also each study's methods, sample size, and reported results. For example, if I could identify several validity studies that used samples of at least 100 EL students and demonstrated positive results, I coded this KEA as having "strong evidence." Conversely, if I only could identify a single study, or multiple studies that used small samples or had mixed results, I coded this KEA as "moderate." I also noted when I could not identify any measure validity or observer reliability research. All of the research used to arrive at these determinations is discussed below.

## Results

### EL-Specific Items

My first research question focused on the extent to which any of the nine sample KEAs contained items that are specific to the development of EL kindergartners. My analyses of the study's data suggested that four of these KEAs include EL-specific items (see Table 2).

#### *California and Illinois*

The first two measures with EL-specific items are California's DRDP-K and Illinois's KIDS observational rubric. These measures were collaboratively developed, and their respective full versions are nearly identical. Not surprisingly, both rubrics include a domain entitled "English Language Development" that is conditional for students learning both their home language and English. The four items in this domain are comprehension of English; self-expression in English; understanding and response to English Literacy activities; and symbol, letter, and print knowledge in English (California Department of Education, 2015a; California Department of Education Child Development Division, 2017; WestEd Center for Child and Family Studies, 2015).

These four EL-conditional items were originally developed for the 2010 DRDP-Preschool measure aimed at children ages 36 months to kindergarten entry and in response to California's preschool learning standards (California Department of Education Child Development Division, 2013). They were subsequently added to the DRDP-School Readiness measure (California Department of Education Child Development Division, 2012)—on which the DRDP-K and KIDS are based—to align with Common Core State Standards for kindergarten language and literacy standards (WestEd Center for Child and Family Studies Evaluation Team, 2013). In addition, the inclusion of two items, self-expression in English and understanding and response to English literacy activities, in the DRDP-K is aligned with California's Kindergarten English Language Development standards (WestEd Center for Child and Family Studies, n.d.).

The scoring scale for these items is slightly different than what is used for the DRDP-K's and KIDS's other items in that it has five steps ranging from *discovering English* to a midpoint of *developing English* and culminating with *integrating English*. When collecting evidence for the measure's other domains, teachers are asked to place a student on a seven-step developmental scale ranging from a low of *earlier building* to a high of *later integrating* or even *emerging to the next developmental level*. However, no matter what domain is being assessed, all rubrics include item-specific examples for each step.

#### *Delaware and Washington*

The final two KEAs with EL-specific items are Delaware's Early Learner Survey and Washington's Kindergarten Inventory of Developing Skills (WaKIDS), both of which are customized versions of the Teaching Strategies GOLD observation

**Table 2** Kindergarten Entry Assessments (KEAs) With English Learner–Specific Items

KEA	Language items	Literacy items	No items
Teacher-administered observational rubric			
CA Desired Results Developmental Profile (DRDP-K)	X	X	
IL Kindergarten Individual Development Survey (KIDS; DRDP-K)	X	X	
DE Early Learner Survey	X		
WA Kindergarten Inventory of Developing Skills (WaKIDS)		X	
PA Kindergarten Entry Inventory			X
Computer-administered direct assessment			
FL Kindergarten Readiness Screener (FLKRS) (Star Early Literacy items)			X
MS Star Early Literacy			X
Teacher-administered direct assessment and observational rubric			
OR Kindergarten Assessment			X
UT Kindergarten Entry and Exit Profile (KEEP)			X

*Note.* CA = California; IL = Illinois; DE = Delaware; WA = Washington; PA = Pennsylvania; FL = Florida; MS = Mississippi; OR = Oregon; UT = Utah.

(Heroman et al., 2010). However, teachers in these two states are not required to use the same EL-specific items. Delaware's KEA has a two-item English language acquisition domain that focuses on English expressive and receptive skills (Delaware Department of Education, n.d.-a). Kindergarten teachers are required to observe students with these two items if they are identified as an EL through use of the state's Home-Language Survey Form (Delaware Department of Education, 2017). These items are scored on the same scale used for all of the Teaching Strategies GOLD items, which asks teachers to place a student on a developmentally progressive, 10-step *Not yet to Level 9* scale. Examples are provided for four of the 10 levels, as well.

Washington teachers using the WaKIDS measure with EL kindergartners are not required but have the option to gather evidence for the English language acquisition domain items. In addition, if teaching in a bilingual Spanish–English classroom and literacy instruction is only in Spanish, teachers are asked to replace the English literacy items with those focused on Spanish literacy (Washington State Department of Early Learning, 2015a, 2017a). As is the case with Delaware's Early Learner Survey, these items are scored on the same Teaching Strategies GOLD developmental progression scale.

### ***Remaining KEAs***

The remaining five KEAs in my sample did not include any EL-specific items at the time of the study. However, it should be noted that during the 2014–2015 to 2016–2017 school years, Oregon required Spanish-speaking EL kindergartners to be assessed with Spanish letter name and letter sound recognition items. Both of these subtests were suspended for the 2017–2018 school year so that the state's Department of Education could determine if they are the best choice for informing teachers' EL-focused literacy instruction (Oregon Department of Education, 2014, 2016b, 2017b).

### **Allowable Linguistic Support Accommodations and Related Assessor/Observer Policies**

My second research question focused on KEA linguistic support accommodations and, in turn, any policies regarding teachers' linguistic capacity when serving as a KEA assessor or observer. As is displayed in Table 3, I identified three different accommodations across all nine sample KEAs. However, the level of accommodation between states varies widely. In addition, I could not identify five states' policies regarding teachers' or translators' linguistic capacities when using these accommodations.

### ***Translated Direct Assessment Directions***

The first accommodation I identified was using a student's home language to provide him or her with directions for completing a direct assessment. As can be seen in Table 3, both states using a computer-administered KEA allow test directions to be translated into students' home language and then provided orally. These directions include an introduction regarding the number of items and average length of time to complete the test, as well as instructions for using the keyboard, mouse, or tablet to select a response (Florida Department of Education, 2017; Mississippi Department of Education & Renaissance Learning, 2017; Office of Assessment, Florida Department of Education, 2017; Office of Student Assessment, Mississippi Department of Education, 2017; Renaissance Learning, 2016, 2017b).

Although Florida allows EL students to elect to have an adult read the test's directions in his or her home language, I could not locate any specific guidance regarding the minimum linguistic capacity or other qualifications of the person translating for or reading the directions to the student. In Mississippi, the only guidance provided is that "consideration needs to be given as to whether the assessment should be explained to the student in his or her native language. .. unless it is clearly not feasible to do so" (Office of Student Assessment, Mississippi Department of Education, 2017, p. 7).

### ***Translated Direct Assessment Item Prompts***

Oregon's and Utah's respective KEAs do not include any standardized operational directions to be translated. However, these states provide examples of variations in state policies on translating direct assessment prompts. More specifically, all of the Oregon Kindergarten Assessment's direct item text may be translated and read to a student in his or her home

**Table 3** Allowable Linguistic Accommodations and Assessor/Observer/Translator Requirements

KEA	Test directions in home language	Item prompts in home language	Demonstrate skills and knowledge in home language or via gestures	Assessor/observer/translator linguistic requirements
Teacher-administered observational rubric				
CA Desired Results Developmental Profile (DRDP-K)	Not applicable	Not applicable	All items	Must speak language
IL Kindergarten Individual Development (KIDS; DRDP-K)	Not applicable	Not applicable	All items	Must speak language
DE Early Learner Survey	Not applicable	Not applicable	17/34 items	Could not determine
WA Kindergarten Inventory of Developing Skills (WaKIDS)	Not applicable	Not applicable	22/31 items	Must be trained and verified
PA Kindergarten Entry Inventory	Not applicable	Not applicable	27/30 items	Could not determine
Computer-administered direct assessment				
FL Kindergarten Readiness Screener (FLKRS) (Star Early Literacy items)	Yes	No	No	Could not determine
MS Star Early Literacy	Yes	No	Gesture or dictate to scribe	Could not determine
Teacher-administered direct assessment and observational rubric				
OR Kindergarten Assessment	Not applicable	Yes	Math items, CBRS	Spanish provided; Other: must be trained/endorsed
UT Kindergarten Entry and Exit Profile (KEEP)	Not applicable	Unscored, opening task only	No	Could not determine

*Note.* KEA = kindergarten entry assessment; CA = California; IL = Illinois; DE = Delaware; WA = Washington; PA = Pennsylvania; FL = Florida; MS = Mississippi; OR = Oregon; UT = Utah.

language. Oregon's state Department of Education provides the Spanish text. Any non-Spanish translations should be completed in advance by an individual who is trained and endorsed by the district. Individuals who administer the test in Spanish or any other language must be endorsed by the district as well (Office of Teaching, Learning, & Assessment, Oregon Department of Education, 2017; Oregon Department of Education, 2016a, 2017b, 2017c).

In contrast, Utah's state-developed Kindergarten Entry and Exit Profile (KEEP), which was being field-tested during the 2017–2018 school year (Utah State Board of Education, 2017b), allows the use of translated item prompts on a much more limited basis. The teacher begins the administration of the KEEP by asking a student to state his or her name and age. These items are not scored and are the only item prompts that may be translated into a student's home language. The prompts for all of the remaining items must be read verbatim in English (Utah State Board of Education, 2017a, 2017c). I could not find any information regarding the linguistic capacity of the teacher or person who may provide these translations or ask the students to state their name and age in any alternate language.

### **Displaying Skills/Knowledge**

The third linguistic accommodation I identified across the study's KEAs was allowing students to use their home language or gestures to display their knowledge and skills in response to direct assessment prompts or as part of an observational rubric. As can be seen in Table 3, all of the states using an observational scale-only KEA allow this accommodation to some degree, as do one of the direct assessment KEAs and one of the two KEAs that use a direct and observational approach.

For example, both California's and Illinois' observational KEAs allow students to demonstrate their knowledge and skills in their home language, in English, or in both languages within a conversation. In addition, the policies governing both of these measures require the teacher who is collecting evidence for these observations to speak the same language(s)

as the EL student. If the teacher does not have that capacity, he or she should obtain help from another adult with that capacity, such as a teaching assistant or parent (California Department of Education, 2015a; California Department of Education Child Development Division, 2017; Illinois State Board of Education, 2017).

A similar accommodation may be used when rating EL kindergartners with almost all of Pennsylvania's Kindergarten Entry Inventory rubric. Three of the measure's language and literacy development domain items (print concepts/letters and phonics, and conventions of English language) must be scored based on a student's English proficiency. For the measure's 27 remaining items—including nine in the language and literacy development domain and all seven mathematics items—children may use their dominant language or non-language based strategies to display their knowledge and skills. In addition, the measure's scoring rubric includes EL-specific examples. For example, for the print concepts/words item, an example of the evident rating is "Jorge labels his house on a drawing, copying 'casa' from the word wall" (Pennsylvania Office of Child Development and Early Learning, 2017b, p. 7). I could not find evidence that Pennsylvania kindergarten teachers must have a minimum level of proficiency in a child's home language to rate EL students. However, teachers may use input from families and other school staff (e.g., paraprofessionals) for all of the items. In addition, if a teacher does not feel comfortable scoring an EL student on an item, he or she may indicate that a "student is non-English speaking" (Pennsylvania Office of Child Development and Early Learning, 2017a, p. 5).

Delaware and Washington's respective KEAs are state-customized versions of the GOLD measure (Heroman et al., 2010) but have very different policies for this specific accommodation. For example, both states consider all items in the social emotional (e.g., interacts with peers), physical (e.g., uses drawing and writing tools) and cognitive (e.g., solves problems) domains to be "language neutral." As a result, EL kindergartners may demonstrate their knowledge and skills for any of these items in English or their home language. However, they differ in the specific "language dependent" mathematics, language, and literacy domain items for which students must be rated according to the extent to which they display their skills and knowledge in English. EL kindergartners in Delaware may use their home language to demonstrate the literacy skill, writes name, and the mathematics skills, quantifies and compares and measures. Students' remaining literacy and mathematics skills, including interacting with read-alouds and book conversations, counting, and connecting numerals with their quantities, must be rated from an English-only perspective (Delaware Department of Education, 2017, n.d.-b). In contrast, Washington's EL kindergartners may use their home language to demonstrate their capacity for all four mathematics items (counts, quantifies, connects numerals with their quantities, and understand shapes). As the Washington State Department of Early Learning (2017a, p. 1), "If the child can count to only 10 in English but can count to 30 in Spanish, then the child can count to 30." However, all seven of Washington's literacy items, including writes name, must be assessed from an English perspective.

Delaware and Washington also differ in the extent to which their policies articulate the linguistic capacity needed by kindergarten teachers when collecting evidence of their EL students' knowledge and skills as part of the KEA process. At the time of the study, I was unable to locate any Delaware policies regarding kindergarten teachers' linguistic capacity when collecting evidence for the three literacy and mathematics items for which students may use their home language. However, in Washington, if a teacher is not proficient in a child's home language, kindergarten teachers must seek assistance from a trained individual whom their district has verified to be fully proficient in that language (Washington State Department of Early Learning, 2015a, 2017a).

Oregon's KEA, which relies on both direct items and an observational rubric, also provides an example of this specific accommodation. The measure's letter name and sound items are rated based on a student's response in English. If a student provides a letter name in a non-English language, teachers or others administering the test must ask the child to say the letter name in English. However, to respond to the math items, students may respond in their home language or English, a mix of either language, or even point to the answer. In addition, kindergartners are not required to use any specific language while being rated by their teachers using the CBRS rubric (Díaz, McClelland, Thompson, & The Oregon School Readiness Research Consortium, 2017). As mentioned earlier, when translations are used, the test administrator must be a bilingual individual who is trained and endorsed by a school district (Office of Assessment, & Accountability, Oregon Department of Education, 2016; Office of Teaching, Learning, & Assessment, Oregon Department of Education, 2017; Oregon Department of Education, 2016a, 2017a, 2017b, 2017c).

Finally, Mississippi's KEA—which is computer-administered—is the only direct measure that allows ELs to use a response accommodation. Because there is no opportunity for a student to be scored based on a non-English verbal

response, EL kindergartners may dictate or gesture answers to a scribe, who then will select that answer as part of the testing process (Office of Student Assessment, Mississippi Department of Education, 2017).

## EL-Relevant Research

My final research question focused on the extent to which the research that has been conducted supports the validity of each these KEAs for demonstrating EL students' skills and knowledge, and for the observational measures specifically, how it supports observer reliability when gathering evidence of what EL students know and can do. As can be seen in Table 4, my analysis of this research base suggests that the DRDP and KIDS measures have strong evidence for both of these topics, but more research is needed for each of the remaining KEAs. All of the studies that I analyzed to make these determinations are discussed next.

### *Desired Results Developmental Profile*

The various iterations of the DRDP, the most recent of which is the KEA used in both California and Illinois, have undergone a variety of EL-relevant studies. For example, in 2005 the DRDP-Revised measure underwent an initial rating validation study using a sample of 610 preschoolers, with 28% of the sample speaking Spanish at home and 17% considered to be bilingual (California Department of Education Child Development Division, 2011). Although this version of the DRDP did not yet include the English language development items, the language and literacy domain items, in particular, were found to have a high level of internal consistency reliability when using a sample of 4-year-olds, 44% of whom spoke Spanish as their only, primary, or home language (Vogel et al., 2008). Next, the 2010 DRDP-Preschool measure, a precursor to the DRDP-K and KIDS, was validated on a sample of 450 children ages 31 – 66 months. Of this sample, 40% spoke Spanish at home and 17% were bilingual (i.e., English and Spanish or another language). This study also demonstrated modest intercorrelations between students' ratings on the language and literacy development and English language development items, as well as reliability of the English language development items alone for EL students (California Department of Education Child Development Division, 2013).

Researchers also examined the relationship between preschoolers' scores on versions of the DRDP and other measures of preschoolers' skills. One study of the implementation of Los Angeles's universal preschool program during the 2006–2007 school year used a sample of 418 4-year-olds, 44% of whom spoke Spanish as their only, primary, or home language. In this study, researchers found a moderate, but positive relationship between children's DRDP-Revised language and literacy ratings and their scores on direct assessments of their expressive and receptive vocabulary (Vogel et al., 2008).

**Table 4** Publicly Available Research Supporting Kindergarten Entry Assessment (KEA) Validity and Observer Reliability

KEA	Measure validity for ELs	Observer reliability with ELs
Teacher-administered observational rubric		
CA DRDP-K and IL KIDS	Strong	Strong
Teaching Strategies GOLD	Strong	Moderate
DE Early Learner Survey (customized subset of GOLD items)	No research identified	Moderate
WA Kindergarten Inventory of Developing Skills (customized subset of GOLD items)	Moderate	Moderate
PA Kindergarten Entry Inventory	Moderate	Moderate
Computer-administered direct assessment		
Star Early Literacy	Moderate	(Not relevant)
FL Kindergarten Readiness Screener (FLKRS; Star Early Literacy items)	No research identified	(Not relevant)
MS Star Early Literacy	No research identified	(Not relevant)
Teacher-administered direct assessment and observational rubric		
OR Kindergarten Assessment	Moderate	Strong
UT Kindergarten Entry and Exit Profile (KEEP)	No research identified	No research identified

*Note.* EL = English learner; CA DRDP-K = California Desired Results Developmental Profile - Kindergarten; IL KIDS = Illinois Kindergarten Individual Development Survey; DE = Delaware; WA = Washington; PA = Pennsylvania; FL = Florida; MS = Mississippi; OR = Oregon; UT = Utah.

In a second study, researchers found a modest, but positive relationship between children's scores on these same direct assessments of children's expressive and receptive vocabulary and teachers' DRDP-School Readiness ratings on the English language development items (WestEd Center for Child and Family Studies, 2012; WestEd Center for Child and Family Studies & University of California — Berkeley Evaluation and Assessment Research Center, 2012). A third study (California Department of Education Child Development Division, 2013) used a sample of 310 ethnically diverse preschoolers to demonstrate a strong correlation between preschoolers' DRDP-Preschool Language and Literacy Development scores and ratings for similar items contained in the Creative Curriculum (Teaching Strategies, 2001) observational measure.

In addition, researchers examined interrater reliability rates for the DRDP and, in particular, ratings for the English language development domain items. In the Los Angeles universal preschool implementation study, interrater reliability on the DRDP-Revised was considered to be relatively weak (Vogel et al., 2008). Researchers found a similar result for novice users of the 2010 DRDP-Preschool measure. However, in this latter study, the domain with the highest overall level of reliability was the English language development domain (California Department of Education Child Development Division, 2013; WestEd Center for Child and Family Studies & University of California — Berkeley Evaluation and Assessment Research Center, 2012). A third study was conducted during the 2014–2015 school year using the current DRDP measure (California Department of Education, 2015a) and based on the scores of special education teachers. As was the case with the other two studies, the majority of the student sample was of preschool age (36–59 months). On average, raters scored all of the items exactly the same 64% of the time, but this rate rose to 71% for the English language development items (California Department of Education, 2015b).

### ***Teaching Strategies GOLD***

Delaware and Washington each use a customized subset of GOLD items as their respective KEAs. Accordingly, two sets of studies are relevant to the current EL-focused investigation: those focused on the original GOLD observational rubric, and studies specific to each of these KEAs. In terms of the former, an array of studies examine the extent to which GOLD is valid for determining the knowledge and skills of EL kindergartners. For example, research aimed at establishing normative scores for the measure's scoring rubric used a sample of over 54,000 3- and 4-year-olds, 22.5% of whom spoke Spanish or other languages (Lambert, 2012). A subsequent study examined GOLD's validity for 32,736 3- to 5-year-old ELs vs. 60,841 monolingual English students. With the expected exception of the uses conventional grammar item, this study found little evidence that GOLD items functioned differently across the two groups (Kim, Lambert, & Burts, 2013).

Three recent studies examined the extent to which GOLD items functioned differently across three different groups of EL preschoolers: children who mostly used their home language in and out of school, children who mostly spoke their home language but also used some English at home or at school, and children who mostly spoke English in the classroom and their home language outside of the classroom. These studies also used large national samples of 4-year-olds with teacher-generated GOLD ratings. Analyses of these data determined that students in the three subgroups differed in terms of their initial ratings and monthly growth rates, as well as in comparison to their non-EL peers (Kim, Lambert, & Burts, 2018; Kim, Lambert, Durham, & Burts, 2018; Lambert, Kim, Durham, & Burts, 2017).

Other studies using smaller samples of EL preschoolers and kindergartners have found there is a moderate relationship between children's GOLD ratings and measures of their language, literacy, mathematics and self-regulation skills (Lambert, Kim, & Burts, 2015; Miller-Bains, Russo, Williford, DeCoster, & Cottone, 2017). However, both of these studies also suggest potential GOLD rater reliability issues. The authors of both studies hypothesized that this variance might have been due to teachers' lack of experience with the measure or the amount of training received.

### ***Delaware and Washington's State-Customized Versions of GOLD***

Publicly available research focused specifically on Delaware's Early Learner Survey is limited to a single study of child outcomes for the 2015–2016 and 2016–2017 school years, with 19% and 16% of Delaware's kindergartners, respectively, identified as ELs. These analyses suggest that ELs are less likely than their non-EL peers to have ratings in the accomplished category in every domain, and the domain with the largest difference in scores is language and literacy. However, because the gap in mathematics items ratings grew considerably over the two school years, the study's authors hypothesized that the state's kindergarten teachers may not be scoring these items in a reliable manner (Hanover Research, 2017).

Much of the research on Washington's WaKIDS measure is focused on the number of districts, schools, students, and teachers that have participated in the scaling up of this measure as well as some brief overall results (Washington State Department of Early Learning, 2013, 2014, 2015b, 2016, 2017b, 2018). However, the first part of an initial validity and reliability study took place in the fall of 2012 and relied on a convenience sample of 54 volunteer teachers, 28 of whom reported that they had completed the interrater reliability certification offered by GOLD's developers. In addition, teachers had an average of 1 year of experience with WaKIDS. These findings were not surprising given that achieving interrater reliability was optional during the Fall 2012 WaKIDS pilot and only 174 out of 1,003 teachers completed the process in place at that time (Office of the Governor, State of Washington, 2013). The teachers were asked to watch videos of four students and rate them using the WaKIDS rubric. Three of the kindergartners spoke English, and the remaining student was categorized as a Latino dual language learner. One of the English-speaking students was considered to be developmentally low functioning. Analysis of the teachers' WaKIDS ratings suggested interrater reliability was moderate overall, but teachers were less likely to be in exact agreement with a master coder when rating the dual language learner or low-functioning student. In addition, 10 of the 13 items with the largest standard deviations from the master rater score for the dual language learner student were in the language and literacy domains (Soderberg et al., 2013).

The second part of this WaKIDS study used a sample of 333 kindergartners enrolled in the classrooms of the 54 teachers who participated in the video rating study to investigate the extent to which students' WaKIDS domain level scores (e.g., language, literacy, mathematics, etc.) were correlated with scores from similar measures of kindergartners' skills. Of these kindergartners, 65% percent spoke English as their primary language and 35% had a non-English primary language. Researchers found that WaKIDS items moderately measured their intended constructs, although the results were stronger for language, literacy, and mathematics than for children's social-emotional and physical development. One important caveat to these findings is the extent to which students' WaKIDS scores were reliable in the first place given the variations noted in teachers' ratings of the video portfolios (Soderberg et al., 2013).

A final WaKIDS study (Gotch et al., 2017) used a sample of 2,210 kindergartners, 44% of whom spoke Spanish as their home language, to examine the relationship between children's WaKIDS language and literacy item scores and their scores on a direct assessment of early literacy skills. This study found a moderate correlation between the scores, suggesting the measures may be targeting different underlying constructs. However, because the authors did not break down the results by student home language, it is difficult to discern to what extent this variable contributed to the results.

### ***Pennsylvania Kindergarten Entry Inventory***

The majority of research on this KEA has focused on implementation of the measure as it was piloted and field tested from 2011 to 2016 (Pennsylvania Department of Education, 2015, 2016, 2017; Pennsylvania Office of Child Development and Early Learning, 2013, 2014a, 2014b), as well. However, preliminary research also was conducted to determine to what extent mean ratings for individual pilot items differed across demographic subgroups, including ELs in comparison to the full sample. Items with statistically significant differences between these two groups included writing, conventions of English language, numbers, number systems & relationships, and geometry. In response, the state's department of education revised the training provided to teachers, as well as the evidence teachers could use to rate EL students on these items (Pennsylvania Office of Child Development and Early Learning, 2013, 2014a, 2014b).

Howard et al. (2017) conducted additional research to examine the extent to which the current version of Pennsylvania's KEA operates consistently across different subgroups of kindergartners. This study used 2014, 2015, and 2016 data from 59,147 kindergartners, 7% of whom were considered to be dual language learners. Their analyses suggested that the items that are potentially biased for this subgroup of students included conventions of English language, expressive language, and receptive language. As part of this latter study, researchers also investigated the relationship between KEA ratings and scores on direct assessments of children's early language, literacy, mathematics, and self-regulation skills. This study had mixed results. However, the data also suggest that there may be a relationship between teachers' scoring reliability and their education levels and teaching experience. For the purposes of the current study, it also is important to note that the sample for this follow-on study purposefully excluded children who could not pass an English language screener.

### **Star Early Literacy**

Both Florida and Mississippi use Star Early Literacy as their KEA. Despite its name, this computer-administered direct assessment contains both early literacy and numeracy items. Research specifically related to Mississippi's measure is limited to reports of average scale scores for individual school districts and schools since the measure was first used in 2014 (Mississippi Department of Education, 2014, 2015; C. M. Wright, 2016, 2017). Florida only began using the measure as its KEA in the 2017–2018 school year (Lyons, 2017), and therefore no publicly available data were available at the time of the current study.

It is not clear to what extent research supports the validity and reliability of the Star Early Literacy measure as a KEA aimed at informing kindergarten teachers' EL-specific instruction. The 2015 technical manual (Renaissance Learning, 2015) indicates that 2.7% of the 2014 norming sample was identified as ELs. However, I could not identify research specifically focused on the measure's construct validity for ELs. Additionally, when I used the assessment developer's online research library advanced search function<sup>4</sup> to request reports related to the drop-down choices of the Star Early Literacy measure, Kindergarten through Grade 3, and focused on English Language learners, no results were generated. It also should be noted that the assessment developer itself not only stresses that scores on the measure need to be interpreted in light of an EL student's English language proficiency (Renaissance Learning, 2010), but also offers a Spanish version of Star Early Literacy (Renaissance Learning, 2017c).

The measure's assessment developers suggest that two small studies demonstrate the positive relationship between students' Star Early Literacy scores and scores on other similar measures. The first study used a sample of 98 kindergartners, 18% of whom were considered to be ELs, to examine to what extent their Star scores predicted students' end of year reading skills. The results of this study, which is limited by its small sample size, were mixed (Clemens et al., 2015). The second study used a larger sample of 633 kindergarten, Grade 1, and Grade 2 students to examine the relationship between Star literacy scores and scores on other literacy assessments. However, only four of the students in this latter sample were considered to be ELs. The results of this study suggested a weak correlation between kindergartners' scores (McBride, Ysseldyke, Milone, & Stickney, 2010).

### **Oregon**

Oregon's KEA is distinct from the other measures in the study's sample in that it is a composite consisting of a letter name and letter sound fluency subtest, direct math items either written by Oregon educators or adapted from the easyCBM, and observational items from the CBRS. Numerous studies demonstrate the concurrent and predictive validity of the letter name, letter sound, and mathematics subtests that formed the basis of the original direct items included in the state's KEA (Lai, Alonzo, & Tindal, 2013; Lai, Nese, Jamgochian, Alonzo, & Tindal, 2010; Wray, Alonzo, & Tindal, 2013), but these studies do not report any data for EL students. However, additional research demonstrates the validity and reliability of the CBRS across different populations of students (Matthews, Cameron Ponitz, & Morrison, 2009; McClelland et al., 2014; Schmitt, Pratt, & McClelland, 2014; von Suchodoletz et al., 2013; Wanless et al., 2011; Wanless et al., 2013; Wanless, McClelland, Acock, Chen, & Chen, 2011; Wanless, McClelland, Tominey, & Acock, 2011).

Two published studies used KEA data to examine the validity of the entire measure. The first study, which used data from schools participating in the state's 2012 KEA pilot, demonstrated a moderate relationship between kindergartners' easyCBM math and literacy scores (which formed the basis for the measure's direct items at that time) and teachers' CBRS ratings of students. Of the 1,189 students sampled, 217, or 18%, were reported to be limited English proficient. However, these results were not broken out by subgroup (Tindal et al., 2015). The second study used data from Oregon's 2013 KEA field test to examine the extent to which students' scores varied on the math, literacy, and CBRS subtests by population subgroup. Data from 41,170 kindergartners from 189 Oregon districts were included in the sample, with 18% of students classified as limited English proficient. This study found that while the same constructs appear to be measured across different population subgroups, significant disparities exist between the literacy and mathematics scores attained by the students classified as limited English proficient, as well as some other subgroups, and their peers (Irvin et al., 2017).

### **Utah**

Utah's KEEP was operationally field tested in the state's school districts at the time of the current study. Accordingly, I identified just five preliminary investigations (Carter, 2017; The National Center for the Improvement of Educational

Assessment, 2017a, 2017b; Throndsen, 2018; Utah State Board of Education, 2017b), none of which provided definitive information about the measure's validity for providing evidence of what EL kindergartners know and can do. Also to be addressed through future research is teacher reliability when using the observational portion of KEEP, which focuses on students' affect when responding to the test's items (e.g., reluctant vs. confident), as well as their attention, focus, listening, and following directions skills (Utah State Board of Education, 2017c).

## Discussion and Implications

In this study, I used a purposefully selected sample of nine state KEAs to examine to what extent these measures contain items that are specific to ELs and allow or mandate the use of linguistic accommodations when assessing or observing EL kindergartners with these measures and, in turn, have policies on KEA assessor or observer linguistic capacity. Also of interest was the degree to which research supports the use of these KEAs with EL students. My aim in conducting the study was to expand our understanding of the potential utility of KEA data to provide evidence of EL kindergartners' knowledge and skills. Such an inquiry is timely given the increasing use of KEA data across the United States, the number of EL students entering kindergarten, and research documenting achievement gaps between EL students and their non-EL peers.

If an aim of collecting KEA data is to produce valid and reliable evidence of what EL kindergartners know and can do, the results of this study suggest that the nine sampled measures may not offer comparable levels of utility. Such a finding is not entirely surprising given that states have had significant leeway to select or develop a KEA that best meets their respective needs. Some of these variations may stem from intentional efforts to support the validity of a KEA for its original purpose, as well. For example, Oregon's literacy items—which are part of a larger KEA—are designed to generate data on children's skills in English. Policymakers in any of the sampled states also may have determined that other mandatory assessments can supplement the data generated by KEAs, and thus a more extensive KEA focus is not necessary.

Yet, these results have implications for KEA decision makers and researchers. These implications include considering the net effect of these variations on the potential utility of a KEA for informing kindergarten teachers' EL-focused instruction and the EL-relevant topics to be addressed through future research.

### “Ballparking” the Potential EL-Relevant Utility of a KEA

As part of the assessment selection or development process, assessment decision makers may look to other states for guidance regarding possible measures to be used (Wakabayashi & Beal, 2015; Weisenfeld, 2017; A. Wright, Farmer, & Baker, 2017). Broadly speaking, the KEAs that formed the sample for the current study represent the different options available. However, researchers caution that the mere availability of measure does not necessarily mean it will accurately assess young ELs' knowledge and skills (Barrueco, López, Ong, & Lozano, 2012).

When viewed through a validity and reliability lens, the first implication of this study for policymakers and others tasked with making assessment decisions is the importance of carefully examining the overall potential of each KEA option to provide accurate evidence of what ELs know and can do. The nine KEAs highlighted here have substantive differences in their EL-relevant content, linguistic accommodations and stated assessor/observer policies, and measure validity and observer reliability research. Furthermore, when combining these differences, the result is three different profiles along a less-to-more content, accommodations, and research continuum (see Figure 1).

For example, Profile A is located on the *less* side of the continuum and consists of the KEAs used in Florida, Mississippi, and Utah. As highlighted above, while students may hear the directions for completing Florida and Mississippi's respective computer-administered KEAs in their home language, none of these three KEAs include EL-focused items, allow any scored item prompts to be translated, or provide students with the option to use their home language to respond to any KEA questions. Furthermore, I could not identify any research that specifically supports the use of these KEAs with EL kindergartners.

Profile C is on the *more* end of the continuum and consists of California's DRDP-K and Illinois' KIDS. Both observational measures contain an English language development domain, with its four items focusing on expressive and receptive vocabulary and early literacy skills. The scoring rubric for this domain reflects the developmental span of young children's second language acquisition and use as well. In addition, with the exception of the EL-conditional items, both measures

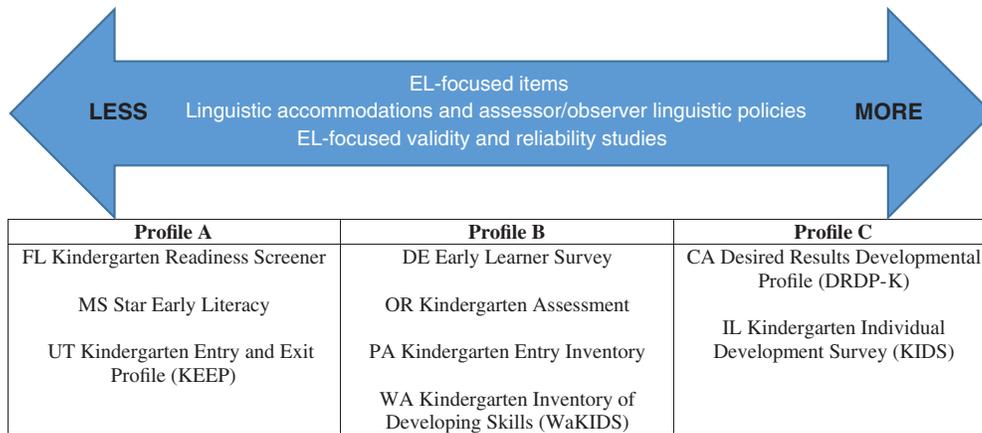


Figure 1 KEA potential EL-related utility profiles

allow children to use whichever language(s) they speak to display their skills and knowledge, and there are related policies regarding observers’ linguistic capacity. Finally, the various iterations of the DRDP have undergone a variety of EL-relevant measure validity and observer reliability studies.

Profile B represents the KEAs with a mix of both *less* and *more* attributes. For example, Oregon’s KEA does not currently have EL-specific items. And, additional research is needed on the validity of the current direct items for demonstrating what EL students know and can do. However, all of the measure’s direct item prompts may be translated and EL students may demonstrate their math skills and knowledge in their home language. The state provides Spanish translations and has specific policies regarding the assessors’ linguistic capacity as well. Pennsylvania’s observational KEA does not have EL-specific items, either, and also could benefit from further measure validity and observer reliability research. But EL students may use any language to display what they know and can do for nearly all of the items.

This “less and more” Profile B status also applies to Delaware’s and Washington’s respective observational KEAs. The Early Learner Survey contains a two-item English language acquisition domain, whereas Washington’s teachers have the option to use these items. In addition, Washington’s Spanish bilingual kindergarten classroom teachers have the option to replace the English literacy items with those focused on Spanish literacy. However, both measures limit the items for which EL students may use their home language to display their knowledge and skills. In addition, although the original GOLD measure has a wealth of EL-relevant research, research on these state customized versions is limited.

Interestingly, these profiles suggest that a state’s English language proficiency or development standards does not necessarily signal similar KEA potential utility for assessing ELs. For example, Florida and Illinois are both WIDA members but are located at opposite ends of the continuum. There also does not appear to be a clear relationship between the percentage of ELs ages 0–8 in a state (as displayed in Table 1) and a KEA’s profile placement. For example, California and Illinois, both of which fall under Profile C, had the first and third highest percentage of ELs ages 0–8, respectively. Mississippi had the lowest percentage and falls under Profile A. Yet, Florida has the second highest percentage of children in this population and also fits Profile A. Although not the focus of this study, further inquiry should examine to what extent the KEAs used in the sample states are aligned with their respective standards. It also may be that other state policies (e.g., “English only”) override efforts to align KEAs with such standards.

Of course, my three profiles do not represent a definitive judgment regarding the in-practice utility of these measures for informing kindergarten teachers’ instruction of their EL students. This observation is especially the case given that the criteria I used may not be equally important to the potential validity or reliability of a KEA for this purpose. At the same time, because research data can potentially help policymakers evaluate whether education policies and practices are adequate for supporting EL students’ academic outcomes (Robinson-Cimpian et al., 2016), the study’s results also suggest three areas for future research.

## Implications for Future EL-Focused KEA Research

The first research implication of the current study's results is the need for additional studies focusing on the validity of KEAs for determining what EL kindergartners know and can do. Furthermore, this research is not limited to recently developed KEAs such as Utah's KEEP. Instead, researchers should examine to what extent prior findings hold for the KEAs that are customized versions of already developed measures. A similar issue to be investigated is KEA rater reliability when observing EL students.

A second key research topic is the extent to which the use of the direct linguistic support accommodations described here contribute to the validity and reliability of a KEA for informing teachers' EL instruction. Given that ELs will vary in terms of their home languages and their degree of academic and social language proficiency (Park, Zong, & Batalova, 2018), these outcomes may differ based on specific student and teacher characteristics as well as assessment approach used. To inform these findings, also of interest are studies examining how these accommodations are being used, by which teachers, and with which students, as well as the larger policy and practice contexts that appear to contribute to any variations in research outcomes.

Third, to shed light on the on-the-ground, practical utility of these KEAs for informing teachers' EL-focused instruction and policies aimed at supporting EL students' academic outcomes, it would be useful for researchers to collect feedback from kindergarten teachers. One important topic is to what extent teachers find the focus of these measures' items to be useful for helping their EL students advance to the next developmental level of a state's learning standards, including those related to English language development. For example, after administering the KEEP's 16 scored literacy and numeracy items in English only, Utah teachers must complete an observational rubric regarding the "behaviors exhibited by the student during administration of the assessment" (Utah State Board of Education, 2017c, p. 31). These behaviors include students' affect when responding to the test's items (e.g., reluctant vs. confident) as well as their attention, focus, listening, and following directions skills. It will be interesting to learn how these observational data are used to inform teachers' instruction of EL kindergartners.

A related topic is teachers' perceptions of the accommodations or supports that must or may be used—or conversely, are not allowed to be used—and to what extent these policies and practices impact teachers' understanding of what their EL students know and can do. All of these topics may have further policy implications for the linguistic capacity needed by kindergarten teachers or other personnel in schools serving large populations of ELs as well as the additional training or technical assistance needed to assess or observe EL kindergartners in a reliable manner.

## Study Limitations

In addition to the potential inadequacies of my profile continuum, my study has two main limitations that hinder its generalizability. First, my sample was composed of nine purposefully selected KEAs. Because at least 40 states and the District of Columbia have incorporated policies on these measures, my sample may not be representative, particularly in terms of other KEAs' EL-specific content, accommodation policies, or research supporting the use of these measures with ELs.

A second limitation is my reliance on publicly available documents and research to seek out data relevant to the study's research questions. Although I often was able to triangulate these data by looking for multiple sources for any topic, it is not clear to what extent the information reported in any of these documents was complete. It is possible that there is additional information and research that I could not include here because they were not publicly available. Future research should therefore build on these initial findings as a means for increasing the field's understanding of the factors to keep in mind when relying on document-based data to investigate KEAs and their related policies.

## Conclusion

This study expands our understanding of the potential utility of KEA data to provide evidence of EL kindergartners' knowledge and skills. Such an understanding is useful given the widespread adoption of KEAs across the United States over the past 6 years as well as the growing number of ELs entering the nation's schools. Although KEAs have the potential to inform teachers' EL-focused instruction and policymakers' efforts to close school readiness gaps, these measures need to be psychometrically strong if they are to provide sufficient evidence for these purposes. The results of the current study

demonstrate that merely selecting a KEA may not necessarily accomplish either goal. Instead, policymakers, assessment developers, researchers, and kindergarten teachers need to work collaboratively to examine the validity and reliability of these measures for demonstrating what EL kindergartners know and can do.

### Acknowledgments

I gratefully acknowledge the financial support received from the W. T. Grant Foundation under Grant 188326 to conduct this study. I also wish to thank Sandra Barrueco, Alexandra Figueras-Daniel, Richard Lambert, and Tomoko Wakabayashi, who served as external reviewers for this report, as well as Danielle Guzman-Orth and Carly Slutzky, who served as ETS technical reviewers. The views expressed in this report are the author's alone.

### Notes

- 1 In this report, the terms *English learner* and *dual language learner* both refer to preschool- and kindergarten-aged children who come from a home where a language other than English is spoken, are continuing to develop this non-English home language, and are learning English as their second language (Administration for Children and Families, U.S. Department of Health and Human Services, & U.S. Department of Education, 2017).
- 2 <http://www.elpa21.org/standards-initiatives/ells-elpa21>
- 3 <https://www.wida.us/membership/states/>
- 4 [http://research.renaissance.com/advancedsearch.asp?\\_ga=2.184826155.369518452.1521562966-1178945380.1521562966](http://research.renaissance.com/advancedsearch.asp?_ga=2.184826155.369518452.1521562966-1178945380.1521562966)
- 5 Percentage is based on the number of ELs enrolled in PreK—Grade 3 in 2013–2104.
- 6 Percentage is based on the number of ELs enrolled in PreK—Grade 2 in 2017–2018.

### References

- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*, 1–28. <https://doi.org/10.3102/00346543074001001>
- Ackerman, D. J. (2018). *Real world compromises: Policy and practice impacts of kindergarten entry assessment-related validity and reliability challenges* (Research Report Series No. RR-18-13). Princeton, NJ: Educational Testing Service.
- Ackerman, D. J., & Coley, R. (2012). *State preK assessment policies: Issues and status*. Princeton, NJ: Educational Testing Service Policy Information Center.
- Ackerman, D. J., & Friedman-Krauss, A. H. (2017). *Preschoolers' executive function: Importance, contributors, research needs and assessment options*. Princeton, NJ: Educational Testing Service.
- Ackerman, D. J., & Tazi, Z. (2015). *Enhancing young Hispanic dual language learners' achievement: Exploring strategies and addressing challenges*. Princeton, NJ: Educational Testing Service Policy Information Center.
- Acosta, B. D., Rivera, C., & Willner, L. S. (2008). *Best practices in state assessment policies for accommodating English language learners: A Delphi study*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.
- Administration for Children and Families, U.S. Department of Health and Human Services, & U.S. Department of Education. (2017). *Policy statement on supporting the development of children who are dual language learners in early childhood programs*. Retrieved from [https://www.acf.hhs.gov/sites/default/files/ece/dll\\_guidance\\_document\\_final.pdf](https://www.acf.hhs.gov/sites/default/files/ece/dll_guidance_document_final.pdf)
- 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. *Elementary School Journal*, *111*, 535–560. <https://doi.org/10.1086/659031>
- Alvarez, L., Ananda, S., Walqui, A., Sato, E., & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners*. San Francisco, CA: WestEd.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Anderson, D., Alonzo, J., Tindal, G., Farley, D., Irvin, P. S., Lai, C., ... & Wray, K. A. (2014). *Technical manual: easyCBM*. Eugene: Behavioral Research and Teaching, University of Oregon.
- Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly*, *48*, 480–506. <https://doi.org/10.1002/tesq.176>
- Barnes, S. K. (2015). Computer-based testing and young children. In O. N. Saracho (Ed.), *Contemporary perspectives on research in assessment and evaluation in early childhood education* (pp. 373–396). Charlotte, NC: Information Age.

- Barrueco, S., López, M., Ong, C., & Lozano, P. (2012). *Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures*. Baltimore, MD: Brookes.
- Bartlett, L., & Vavrus, F. (2017). *Rethinking case study research: A comparative approach*. New York, NY: Routledge.
- Bassok, D., & Latham, S. (2017). Kids today: The rise in children's academic skills at kindergarten entry. *Educational Researcher*, 46, 7–20. <https://doi.org/10.3102/0013189X17694161>
- Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, 11, 1–29. <https://doi.org/10.2167/beb392.0>
- Bernstein, S., West, J., Newsham, R., & Reid, M. (2014). *Kindergartners' skills at school entry: An analysis of the ECLS-K*. Princeton, NJ: Mathematica Policy Research.
- Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 87–106). Thousand Oaks, CA: SAGE. <https://doi.org/10.4135/9781452218649.n6>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child behavior rating scale*. Cambridge, MA: Abt.
- California Department of Education. (2015a). *Desired Results Developmental Profile: Kindergarten (DRDP-K)*. Sacramento, CA: Author.
- California Department of Education. (2015b). *DRDP (2015) 2014–2015 interrater agreement study report*. Sacramento, CA: Author.
- California Department of Education Child Development Division. (2011). *DRDP-R technical report*. Sacramento, CA: Author.
- California Department of Education Child Development Division. (2012). *Desired Results Developmental Profile: School Readiness (DRDP-SR)*. Sacramento, CA: Author.
- California Department of Education Child Development Division. (2013). *California Desired Results Developmental Profile (2010) technical report*. Sacramento, CA: Author.
- California Department of Education Child Development Division. (2017). *Kindergarten Individual Development Survey (KIDS) user's guide & instrument*. Sacramento, CA: Author.
- California State Board of Education. (2012). *California English language development standards (Electronic edition): Kindergarten through grade 12*. Sacramento, CA: Author.
- Carter, C. (2017). *Utah KEEP standard setting process* (Memorandum). Salt Lake City: Utah State Board of Education.
- Cash, A. H., Hamre, B. K., Pianta, R. C., & Myers, S. S. (2012). Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly*, 27, 529–54. <https://doi.org/10.1016/j.ecresq.2011.12.006>
- The Center on Standards & Assessment Implementation. (2017). *State of the states: Pre-K/K assessment*. Retrieved from <http://www.csa-online.org/sos>
- Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., ... & Jones, M. (2015). The predictive validity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review*, 44, 76–97. <https://doi.org/10.17105/SPR44-1.76-97>
- Council of Chief State School Officers. (2014). *English language proficiency (ELP) standards with correspondences to K-12 English language arts (ELA), mathematics, and science practices, K-12 ELA standards, and 6–12 literacy standards*. Washington, DC: Author.
- Dady, N., Lyons, S., & DePascale, C. (2018). The comparability of scores from different digital devices: A literature review and synthesis with recommendations for practice. *Applied Measurement in Education*, 31, 30–50. <https://doi.org/10.1080/08957347.2017.1391262>
- Delaware Department of Education. (n.d.-a). *Delaware Early Learner Survey objectives for Development and Learning*. Dover, DE: Author. Retrieved from <https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/534/Summary%20-%20DE-ELS%20ODL.pdf>
- Delaware Department of Education. (n.d.-b). *Language-free and Language-dependent objectives by domain*. Retrieved from <https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/534/Language-free%20and%20Language-dependent%20Objectives%20By%20Domain.pdf>
- Delaware Department of Education. (2014). *Annual report of Delaware's English language learners, staff, and programs*. Dover, DE: Author.
- Delaware Department of Education. (2017). *Delaware Department of Education home language survey*. Dover, DE: Author.
- Dever, M. T., & Barta, J. J. (2001). Standardized entrance assessment in kindergarten: A qualitative analysis of the experiences of teachers, administrators, and parents. *Journal of Research in Childhood Education*, 15, 220–233. <https://doi.org/10.1080/02568540109594962>
- Díaz, G., McClelland, M., Thompson, K., & The Oregon School Readiness Research Consortium. (2017). *English Language learner students entering kindergarten in Oregon*. Eugene: Oregon State University.
- Division for Early Childhood of the Council for Exceptional Children. (2007). *Promoting positive outcomes for children with disabilities: Recommendations for curriculum, assessment, and program evaluation*. Missoula, MT: Author.
- Early Learning Challenge Technical Assistance. (2016). *Kindergarten entry assessments in RTT-ELC grantee states*. Retrieved from <https://elc.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=15013>

- Educational Testing Service. (2016). *ETS International principles for the fairness of assessments*. Princeton, NJ: Author.
- Espinosa, L. M., & García, E. (2012). *Developmental assessment of young dual language learners with a focus on kindergarten entry assessments: Implications for state policies*. (Working paper #1). Chapel Hill: The University of North Carolina, FPG Child Development Institute, CECER-EL.
- Espinosa, L. M., & López, M. L. (2007). *Assessment considerations for young English language learners across different levels of accountability*. Washington, DC: National Early Childhood Accountability Task Force and First 5 LA.
- Ferrara, A. M., & Lambert, R. G. (2015). *Findings from the 2014 North Carolina kindergarten entry formative assessment pilot*. Charlotte: Center for Educational Measurement and Evaluation, The University of North Carolina at Charlotte.
- Ferrara, A. M., & Lambert, R. G. (2016). *Findings from the 2015 statewide implementation of the North Carolina K-3 formative assessment process: Kindergarten entry assessment*. Charlotte: Center for Educational Measurement and Evaluation, The University of North Carolina at Charlotte.
- Florida Department of Education. (2017). *2017–18 Florida Kindergarten Readiness Screener (FLKRS): Star Early Literacy* (Preview webinar, May 3, 2017). Retrieved from <http://www.fldoe.org/core/fileparse.php/18494/urlt/FLKRS050317Webinar.pdf>
- García, O. (2011). The translanguaging of Latino kindergartners. In K. Potowski & J. Rothman (Eds.), *Bilingual youth: Spanish in English-speaking societies* (pp. 33–55). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/sibil.42.05gar>
- Gibton, D. (2016). *Researching education policy, public policy, and policymakers*. New York, NY: Routledge.
- Gillanders, C., & Castro, D. C. (2015). Assessment of English language acquisition in young dual language learners. In O. N. Saracho (Ed.), *Contemporary perspectives on research in assessment and evaluation in early childhood education* (pp. 291–311). Charlotte, NC: Information Age.
- Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report*, XXII(1), 1–24. <https://doi.org/10.1002/j.2379-3988.2008.tb00054.x>
- Golan, S., Woodbridge, M., Davies-Mercier, B., & Pistorino, C. (2016). *Case studies of the early implementation of kindergarten entry assessments*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability*, 28, 273–293. <https://doi.org/10.1007/s11092-015-9231-8>
- Goldstein, J., & McCoach, D. B. (2011). The starting line: Developing a structure for teacher ratings of students' skills at kindergarten entry. *Early Childhood Research & Practice*, 13(2). Retrieved from <http://ecrp.uiuc.edu/v13n2/goldstein.html>
- Goldstein, J., McCoach, D. B., & Yu, H. (2017). The predictive validity of kindergarten readiness judgments: Lessons from one state. *The Journal of Educational Research*, 110(1), 50–60. <https://doi.org/10.1080/00220671.2015.1039111>
- Gotch, C. M., Beecher, C. C., Lenihan, K., French, B. F., Juarez, C., & Strand, P. S. (2017). WaKIDS GOLD as a measure of literacy and language in relation to other standardized measures. *The WERA Educational Journal* 10(1), 44–51.
- Guhn, M., Gadermann, A., & Zumbo, B. D. (2007). Does the EDI measure school readiness in the same way across different groups of children? *Early Education and Development*, 18, 453–472. <https://doi.org/10.1080/10409280701610838>
- Guzman-Orth, D., Lopez, A. A., & Tolentino, F. (2017). *A framework for the dual language assessment of young dual language learners in the United States* (Research Report No. RR-17-37). Princeton, NJ: Educational Testing Service.
- Han, W.-J., Lee, R., & Waldfogel, J. (2012). School readiness among children of immigrants in the US: Evidence from a large national birth cohort study. *Children and Youth Services Review*, 34, 771–782. <https://doi.org/10.1016/j.childyouth.2012.01.001>
- Hanover Research. (2017). *Delaware early learner survey key findings*. Arlington, VA: Author.
- Harvey, E. A., Fischer, C., Weieneth, J. L., Hurwitz, S. D., & Sayer, A. G. (2013). Predictors of discrepancies between informants' ratings of preschool-aged children's behavior: An examination of ethnicity, child characteristics, and family functioning. *Early Childhood Research Quarterly*, 28, 668–682. <https://doi.org/10.1016/j.ecresq.2013.05.002>
- Heroman, C., Burts, D. C., Berke, K., & Bickart, T. (2010). *Teaching strategies GOLD objectives for development and learning: Birth through kindergarten*. Washington, DC: Teaching Strategies.
- Hoff, E. (2013). Interpreting the early language trajectories of children from low SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49, 4–14. <https://doi.org/10.1037/a0027238>
- Howard, E. C., Dahlke, K., Tucker, N., Liu, F., Weinberg, E., Williams R., ... & Brumley, B. (2017). *Evidence-based Kindergarten Entry Inventory for the commonwealth: A journey of ongoing improvement*. Washington, DC: American Institutes for Research.
- Huang, F. L., & Konold, T. R. (2014). A latent variable investigation of the Phonological Awareness Literacy Screening-Kindergarten assessment: Construct identification and multigroup comparisons between Spanish-speaking English-language learners (ELLs) and non-ELL students. *Language Testing*, 31, 205–221. <https://doi.org/10.1177/0265532213496773>
- Illinois State Board of Education. (2015). *Kindergarten individual development survey*. Springfield, IL: Author.
- Illinois State Board of Education. (2017). *KIDS Guidance for dual language learners*. Springfield, IL: Author.
- Irvin, P. S., Tindal, G., & Slater, S. (2017, April). *Examining the factor structure and measurement invariance of a large-scale kindergarten entry assessment*. Paper presented at the annual conference of the American Educational Research Association, San Antonio, TX.

- Johnson, T. G., & Buchanan, M. (2011). Constructing and resisting the development of a school readiness survey: The power of participatory research. *Early Childhood Research & Practice*, 13(1). Retrieved from <http://ecrp.uiuc.edu/v13n1/johnson.html>
- Jonas, D. L., & Kassner, L. (2014). *Virginia's Smart Beginnings kindergarten readiness assessment pilot: Report from the Smart Beginnings 2013/14 school year pilot of Teaching Strategies GOLD in 14 Virginia school divisions*. Richmond: Virginia Early Childhood Foundation.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1–73. <https://doi.org/10.1111/jedm.12000>
- Karelitz, T. M., Parrish, D. M., Yamada, H., & Wilson, M. (2010). Articulating assessments across childhood: The cross-age validity of the Desired Results Developmental Profile-Revised. *Educational Assessment*, 15, 1–26. <https://doi.org/10.1080/10627191003673208>
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: Meta-analysis on effectiveness and validity. *Review of Educational Research*, 79, 1168–1201. <https://doi.org/10.3102/0034654309332490>
- Kieffer, M. J., Rivera, M., & Francis, D. J. (2012). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments. 2012 update*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Kim, D., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of Teaching Strategies GOLD assessment tool for English language learners and children with disabilities. *Early Education and Development*, 24, 574–595. <https://doi.org/10.1080/10409289.2012.701500>
- Kim, D., Lambert, R. G., & Burts, D. C. (2018). Are young dual language learners homogeneous? Identifying subgroups using latent class analysis. *The Journal of Educational Research*, 111, 43–57. <https://doi.org/10.1080/00220671.2016.1190912>
- Kim, D., Lambert, R. G., Durham, S., & Burts, D. C. (2018). Examining the validity of GOLD with 4-year-old dual language learners. *Early Education and Development*, 29, 477–493. <https://doi.org/10.1080/10409289.2018.1460125>
- Lai, C., Alonzo, J., & Tindal, G. (2013). *EasyCBM reading criterion related validity evidence: Grades K–1*. Eugene: Behavioral Research and Teaching, University of Oregon.
- Lai, C., Nese, J. F. T., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM primary-level reading measures (Grades K-1) 2009–2010 version*. Eugene: Behavioral Research and Teaching, University of Oregon.
- Lambert, R. (2012). *Growth norms for the Teaching Strategies GOLD assessment system*. Charlotte: Center for Educational Measurement and Evaluation, University of North Carolina.
- Lambert, R., Kim, D., & Burts, D. C. (2015). The measurement properties of the Teaching Strategies GOLD assessment system. *Early Childhood Research Quarterly*, 33, 49–63. <https://doi.org/10.1016/j.ecresq.2015.05.004>
- Lambert, R. G., Kim, D., Durham, S., & Burts, D. C. (2017). Differentiated rates of growth across preschool dual language learners. *Bilingual Research Journal*, 40, 81–101. <https://doi.org/10.1080/15235882.2016.1275884>
- Lee, O. (2018). English language proficiency standards aligned with content standards. *Educational Researcher*, 47, 317–327. <https://doi.org/10.3102/0013189X18763775>
- Livingston, S. A. (2018). *Test reliability—Basic concepts* (Research Memorandum No. RM-18-01). Princeton, NJ: Educational Testing Service.
- Lopez, A. A., Turkan, S., & Guzman-Orth, D. (2017). *Conceptualizing the use of translanguaging in initial content assessments for newly arrived emergent bilingual students* (Research Report No. RR-17-07). Princeton, NJ: Educational Testing Service.
- Lyons, H. (2017). *Memorandum: Selection of new Florida Kindergarten Readiness Screener (FLKRS) for 2017–18; Implementation of program and training for test administrators*. Tallahassee, FL: Florida Department of Education.
- Maryland State Department of Education. (2013). *Early childhood comprehensive assessment system – A multi-state collaboration*. Retrieved from <https://www2.ed.gov/programs/eag/mdeag2013.pdf>
- Mashburn, A. J., & Henry, G. T. (2005). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice*, 23, 16–30. <https://doi.org/10.1111/j.1745-3992.2004.tb00165.x>
- Mathews, J. S., Cameron Ponitz, C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, 101, 689–704. <https://doi.org/10.1037/a0014240>
- McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E. (2010). Technical adequacy and cost benefit of four measures of early literacy. *Canadian Journal of School Psychology*, 25, 189–204. <https://doi.org/10.1177/0829573510363796>
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The Head–Toes–Knees–Shoulders task. *Frontiers in Psychology*, 5, Article 599. <https://doi.org/10.3389/fpsyg.2014.00599>
- McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., ... & Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Mihai, F. (2010). *Assessing English language learners in the content areas: A research-into-practice guide for educators*. Ann Arbor: University of Michigan Press. <https://doi.org/10.3998/mpub.2825611>

- Miller-Bains, K. L., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Examining the validity of a multidimensional performance-based assessment at kindergarten entry. *AERA Open*, 3(2), 1–16. <https://doi.org/10.1177/2332858417706969>
- Mislevy, R. J., Wilson, M. R., Ercikan, K., & Chudowsky, N. (2003). Psychometric principles in student assessment. In D. Stufflebeam & T. Kellaghan (Eds.), *International handbook of educational evaluation* (pp. 489–532). Dordrecht, The Netherlands: Kluwer Academic Press. [https://doi.org/10.1007/978-94-010-0309-4\\_31](https://doi.org/10.1007/978-94-010-0309-4_31)
- Mississippi Department of Education. (2014). *Kindergarten readiness assessment results October 2014*. Jackson, MS: Author.
- Mississippi Department of Education. (2015). *Kindergarten readiness assessment results May 2015*. Jackson, MS: Author.
- Mississippi Department of Education & Renaissance Learning. (2017). *MKAS 2 K-Readiness assessment teacher manual*. Wisconsin Rapids, WI: Author.
- Mississippi Department of Education Office of Elementary Education and Reading, Student Intervention Services Pre-K-12. (2016). *English Learner (EL) administrator and teacher guide*. Jackson, MS: Author.
- Moon, T. R. (2005). The role of assessment in differentiation. *Theory Into Practice*, 44, 226–233. [https://doi.org/10.1207/s15430421tip4403\\_7](https://doi.org/10.1207/s15430421tip4403_7)
- Najarian, M., Snow, K., Lennon, J., Kinsey, S., & Mulligan, G. (2010). *Early Childhood Longitudinal Birth Cohort (ECLS-B) preschool-kindergarten 2007 psychometric report*. Washington, DC: U.S. Department of Education.
- National Association for the Education of Young Children. (2009). *Where we stand on assessing young English language learners*. Washington, DC: Author.
- The National Center for the Improvement of Educational Assessment. (2017a). *Report on KEEP dimensionality analyses*. Dover, NH: Author.
- The National Center for the Improvement of Educational Assessment. (2017b). *Suggestions for revising the scoring rules for Utah's Kindergarten Entry and Exit Portfolio (KEEP) entry assessment*. Dover, NH: Author.
- North Carolina Department of Public Instruction. (2013). *Office of Elementary and Secondary Education (OESE): Enhanced assessment instruments grants program — Enhanced assessment instruments: Kindergarten entry assessment competition CFDA Number 84.368A*. Retrieved from <https://www2.ed.gov/programs/eag/nceag2013.pdf>
- Office of Assessment & Accountability, Oregon Department of Education. (2016). *Oregon kindergarten assessment specifications 2016–2017*. Salem, OR: Author.
- Office of Assessment, Florida Department of Education. (2017). *FLKRS & STAR early literacy FAQ*. Tallahassee, FL: Author.
- Office of Student Assessment, Mississippi Department of Education. (2017). *Mississippi testing accommodations manual revised*. Jackson, MS: Author.
- Office of Teaching, Learning, & Assessment, Oregon Department of Education. (2017). *Oregon kindergarten assessment specifications 2017–2018*. Salem, OR: Author.
- Office of the Governor, State of Washington. (2013). *Race to the Top Early Learning Challenge 2012 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.
- Ohle, K. A., & Harvey, H. A. (2017). Educators' perceptions of school readiness within the context of a kindergarten entry assessment in Alaska. *Early Child Development and Care*. <https://doi.org/10.1080/03004430.2017.1417855>.
- Oregon Department of Education. (2014). *Guidance for locating a Spanish bilingual assessor for the Oregon Kindergarten Assessment*. Retrieved from <http://www.ode.state.or.us/news/announcements/announcement.aspx?9909>
- Oregon Department of Education. (2016a). *2016–17 Kindergarten Assessment administration frequently asked questions*. Salem, OR: Author.
- Oregon Department of Education. (2016b). *Preliminary test administration manual: 2016–17 school year*. Salem, OR: Author.
- Oregon Department of Education. (2017a). *2017–18 Kindergarten Assessment early literacy & early math training required for DTCs, STCs, and TAs (Module 2)* (PowerPoint presentation). Retrieved from <http://www.oregon.gov/ode/educator-resources/assessment/Pages/Assessment-Training-Materials.aspx>
- Oregon Department of Education. (2017b). *Oregon accessibility manual for the Kindergarten Assessment: 2017–2018 school year*. Salem, OR: Author.
- Oregon Department of Education. (2017c). *Test administration manual: 2017–18 school year*. Salem, OR: Author.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017a). *Dual language learners: A demographic and policy profile for California*. Washington, DC: Migration Policy Institute.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017b). *Dual language learners: A demographic and policy profile for Florida*. Washington, DC: Migration Policy Institute.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017c). *Dual language learners: A demographic and policy profile for Illinois*. Washington, DC: Migration Policy Institute.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017d). *Dual language learners: A demographic and policy profile for Oregon*. Washington, DC: Migration Policy Institute.

- Park, M., O'Toole, A., & Katsiaficas, C. (2017e). *Dual language learners: A demographic and policy profile for Pennsylvania*. Washington, DC: Migration Policy Institute.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017f). *Dual language learners: A demographic and policy profile for Utah*. Washington, DC: Migration Policy Institute.
- Park, M., O'Toole, A., & Katsiaficas, C. (2017g). *Dual language learners: A demographic and policy profile for Washington*. Washington, DC: Migration Policy Institute.
- Park, M., Zong, J., & Batalova, J. (2018). *Growing superdiversity among young U.S. dual language learners and its implications*. Washington, DC: Migration Policy Institute.
- Pennsylvania Department of Education. (2015). *Pennsylvania Kindergarten Entry Inventory cohort 1 (2014) summary report*. Harrisburg, PA: Author.
- Pennsylvania Department of Education. (2016). *Pennsylvania Kindergarten Entry Inventory cohort 2 (2015) summary report*. Harrisburg, PA: Author.
- Pennsylvania Department of Education. (2017). *Pennsylvania Kindergarten Entry Inventory cohort 3 (2016) summary report*. Harrisburg, PA: Author.
- Pennsylvania Office of Child Development and Early Learning. (2013). *The 2011 SELMA pilot*. Harrisburg, PA: Author.
- Pennsylvania Office of Child Development and Early Learning. (2014a). *The 2012 SELMA pilot*. Harrisburg, PA: Author.
- Pennsylvania Office of Child Development and Early Learning. (2014b). *The 2013 SELMA pilot*. Harrisburg, PA: Author.
- Pennsylvania Office of Child Development and Early Learning. (2017a). *A guide to using the Pennsylvania Kindergarten Entry Inventory*. Harrisburg, PA: Author.
- Pennsylvania Office of Child Development and Early Learning. (2017b). *Kindergarten Entry Inventory*. Harrisburg, PA: Author.
- Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.
- Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with Grade 2 achievement. *Early Childhood Research Quarterly, 28*, 437–449. <https://doi.org/10.1016/j.ecresq.2012.11.002>
- Quirk, M., Rebelez, J., & Furlong, M. (2014). Exploring the dimensionality of a brief school readiness screener for use with Latino/a children. *Journal of Psychoeducational Assessment, 32*, 259–264. <https://doi.org/10.1177/0734282913505994>
- Renaissance Learning. (2010). *Getting the most out of Star Early Literacy: Using data to inform instruction and intervention*. Wisconsin Rapids, WI: Author.
- Renaissance Learning. (2015). *Star Early Literacy technical manual*. Wisconsin Rapids, WI: Author.
- Renaissance Learning. (2016). *Star Early Literacy pretest instructions*. Wisconsin Rapids, WI: Author.
- Renaissance Learning. (2017a). *Star assessments for early literacy: Abridged technical manual*. Wisconsin Rapids, WI: Author.
- Renaissance Learning. (2017b). *Star Early Literacy test administration manual*. Wisconsin Rapids, WI: Author.
- Renaissance Learning. (2017c). *Star Spanish*. Wisconsin Rapids, WI: Author.
- Roach, A. T., McGrath, D., Wixson, C., & Talapatra, D. (2010). Aligning an early childhood assessment to state kindergarten content standards: Application of a nationally recognized alignment framework. *Educational Measurement: Issues and Practice, 29*, 25–37. <https://doi.org/10.1111/j.1745-3992.2009.00167.x>
- Robinson, J. P. (2010). The effect of test translation on young English learners' mathematics performance. *Educational Researcher, 39*, 582–590. <https://doi.org/10.3102/0013189X10389811>
- Robinson-Cimpian, J. P., Thompson, K. D., & Umansky, I. M. (2016). Research and policy considerations for English learner equity. *Policy Insights From the Behavioral and Brain Sciences, 3*, 129–137. <https://doi.org/10.1177/2372732215623553>
- Rollins, P. R., McCabe, A., & Bliss, L. (2000). Culturally sensitive assessment of narrative skills in children. *Seminars in Speech and Language, 21*, 223–234. <https://doi.org/10.1055/s-2000-13196>
- Sanchez, S. V., Rodriguez, B. J., Soto-Huerta, M. E., Villarreal, F. C., Guerra, N. S., & Flores, B. B. (2013). A case for multidimensional bilingual assessment. *Language Assessment Quarterly, 10*, 160–177. <https://doi.org/10.1080/15434303.2013.769544>
- Schachter, R. E., Strang, T. M., & Piasta, S. B. (2017). Teachers' experiences with a state-mandated kindergarten readiness assessment. *Early Years, 37*. <https://doi.org/10.1080/09575146.2017.1297777>
- Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development, 25*, 641–660. <https://doi.org/10.1080/10409289.2014.850397>
- Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.
- Soderberg, J. S., Stull, S., Cummings, K., Nolen, E., McCutchen, D., & Joseph, G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS)*. Seattle: University of Washington.

- Takanishi, R., & Le Menestrel, S. (Eds.). (2017). *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24677>
- Teaching Strategies. (2001). *The creative curriculum developmental profile for ages 3–5*. Washington, DC: Author.
- Texas Education Agency. (2013). Office of Elementary and Secondary Education (OESE): Enhanced assessment instruments grants program – Enhanced assessment instruments: Kindergarten entry assessment competition CFDA Number 84.368A. Austin, TX: Author. Retrieved from <https://www2.ed.gov/programs/eag/texaseag2013.pdf>
- Thronsdon, J. E. (2018). *Relationships among preschool attendance, type, and quality and early mathematical literacy* (Unpublished doctoral dissertation). Logan: Utah State University.
- Thurlow, M. L. (2014). Accommodation for challenge, diversity, and variance in human characteristics. *Journal of Negro Education*, 83, 442–464. <https://doi.org/10.7709/jnegroeducation.83.4.0442>
- Tindal, G., Irvin, P. S., Nese, J. F. T., & Slater, S. (2015). Skills for children entering kindergarten. *Educational Assessment*, 20, 197–319. <https://doi.org/10.1080/10627197.2015.1093929>
- Tomlinson, C. A. (2008). The goals of differentiation. *Educational Leadership*, 66(3), 26–30.
- Turkan, S., Oliveri, M. E., & Cabrera, J. (2013). Using translation as a test accommodation with culturally and linguistically diverse learners. In D. Tsagari & G. Floros (Eds.), *Translation in language teaching and assessment* (pp. 215–234). Newcastle upon Tyne, England: Cambridge Scholars.
- Utah State Board of Education. (2017a). *Materials prior to training* (PowerPoint presentation). Retrieved from <https://www.schools.utah.gov/file/86da295d-5b5b-43bc-83bf-9479d42ab815>
- Utah State Board of Education. (2017b). *Utah's KEEP report*. Salt Lake City, UT: Author.
- Utah State Board of Education. (2017c). *Utah's Kindergarten Entry and Exit Profile (KEEP) test administration manual*. Salt Lake City, UT: Author.
- Vogel, C., Aikens, N., Atkins-Burnett, S., Martin, E. S., Caspe, M., Sprachman, S., & Love, J. M. (2008). *Reliability and validity of child outcome measures with culturally and linguistically diverse preschoolers: The first 5 LA universal preschool child outcomes study spring 2008 pilot study*. Princeton, NJ: Mathematica.
- von Suchodoletz, A., Gestsdottir, S., Wanless, S. B., McClelland, M. M., Birgisdottir, F., Gunzenhauser, C., & Ragnarsdottir, H. (2013). Behavioral self-regulation and relations to emergent academic skills among children in Germany and Iceland. *Early Childhood Research Quarterly*, 28, 62–73. <https://doi.org/10.1016/j.ecresq.2012.05.003>
- Wakabayashi, T., & Beal, J. A. (2015). Assessing school readiness in children: Historical analysis, current trends. In O. N. Saracho (Ed.), *Contemporary perspectives on research in assessment and evaluation in early childhood education* (pp. 69–91). Charlotte, NC: Information Age.
- Wanless, S. B., McClelland, M. M., Acock, A. C., Chen, F., & Chen, J. (2011). Behavioral regulation and early academic achievement in Taiwan. *Early Education and Development*, 22, 1–28. <https://doi.org/10.1080/10409280903493306>
- Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S., Lan, X., ... & Li, S. (2011). Measuring behavioral regulation in four societies. *Psychological Assessment*, 23, 364–378. <https://doi.org/10.1037/a0021768>
- Wanless, S. B., McClelland, M. M., Lan, X., Son, S., Cameron, C. E., Morrison, F. J., ... & Sung, M. (2013). Gender differences in behavioral regulation in four societies: The United States, Taiwan, South Korea, and China. *Early Childhood Research Quarterly*, 28, 621–633. <https://doi.org/10.1016/j.ecresq.2013.04.002>
- Wanless, S. B., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2011). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Education and Development*, 22, 461–488. <https://doi.org/10.1080/10409289.2011.536132>
- Washington State Department of Early Learning. (2013). *Fall 2012 WaKIDS baseline data release*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2014). *Fall 2013 WaKIDS data summary*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2015a). *Guidance for WaKIDS teachers working with English language learners (ELL)*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2015b). *WaKIDS Fall 2014 data summary*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2016). *WaKIDS Fall 2015 data summary*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2017a). *Using Teaching Strategies GOLD to assess children who are English-language learners*. Retrieved from <http://www.k12.wa.us/WaKIDS/WaKIDSIntro101Training/AssessingChildrenwhoareEnglishLanguageLearners.pdf>
- Washington State Department of Early Learning. (2017b). *WaKIDS Fall 2016 data summary*. Olympia, WA: Author.
- Washington State Department of Early Learning. (2018). *WaKIDS Fall 2017 data summary*. Olympia, WA: Author.
- Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, 27, 46–54. <https://doi.org/10.1016/j.ecresq.2011.06.003>

- Weisenfeld, G. G. (2016). *Using Teaching Strategies GOLD within a kindergarten entry assessment system* (CEELO FastFact). New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.
- Weisenfeld, G. G. (2017). *Assessment tools used in kindergarten entry assessments (KEAs): State scan*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.
- WestEd Center for Child and Family Studies. (n.d.). *Desired Results Developmental Profile-Kindergarten (DRDP-K) correspondence to California learning standards: English language development (ELD) and California English language development standards-kindergarten (ELD-K)*. Sausalito, CA: Author.
- WestEd Center for Child and Family Studies. (2012). *Development of the Desired Results Developmental Profile – School Readiness (DRDP-SR)*. Camarillo, CA: Author.
- WestEd Center for Child and Family Studies. (2015). *Research summary: English language development (ELD) domain in the DRDP (2015) assessment instrument*. Sausalito, CA: Author.
- WestEd Center for Child and Family Studies Evaluation Team. (2013). *The DRDP-SR (2012) instrument and its alignment with the Common Core State Standards for kindergarten*. Sausalito, CA: Author.
- WestEd Center for Child and Family Studies & University of California – Berkeley Evaluation and Assessment Research Center. (2012). *DRDP-SR and Illinois KIDS: Psychometric approach and research studies* (PowerPoint presentation). Retrieved from <http://206.166.105.35/KIDS/pdf/KIDS-psychometrics1212.pdf>
- Williford, A. P., Downer, J. T., & Hamre, B. K. (2014). *The Virginia kindergarten readiness project: Concurrent validity of Teaching Strategies GOLD, Fall 2013, Phase 1 report*. Charlottesville: Center for Advanced Study of Teaching and Learning, Curry School of Education, University of Virginia.
- Wolf, M. K., Kao, J. C., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses – Practice review* (Part 2 of 3; CRESST Report 732). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Wray, K. A., Alonzo, J., & Tindal, G. (2013). *Internal consistency of the easyCBM CCSS math measures: Grades K-8*. Eugene: Behavioral Research and Teaching, University of Oregon.
- Wright, A., Farmer, D., & Baker, S. (2017). *Review of early childhood assessments: Methods and recommendations*. Dallas, TX: Annette Caldwell Simmons School of Education & Human Development Center on Research & Evaluation, Southern Methodist University.
- Wright, C. M. (2016). *Kindergarten readiness assessment results*. Jackson: Mississippi Department of Education.
- Wright, C. M. (2017). *2016–2017 Kindergarten readiness assessment results*. Jackson: Mississippi Department of Education.
- Wright, R. J. (2010). *Multifaceted assessment for early childhood education*. Thousand Oaks, CA: SAGE.
- Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: SAGE.

### Suggested citation:

Ackerman, D. J. (2018). *Comparing the potential utility of kindergarten entry assessments to provide evidence of English learners' knowledge and skills* (Research Report No. RR-18-36). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12224>

**Action Editor:** Heather Buzick

**Reviewers:** Danielle Guzman-Orth and Carly Slutzky

ETS, the ETS logo, and MEASURING THE POWER OF LEARNING. are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS RESEARCHER database at <http://search.ets.org/researcher/>