January 2022

DRDP Technical Report for Early Infancy Through Kindergarten:

Structural Validity and Reliability Information for the Desired Results Developmental Profile

Lead author:
Karen Draney, Berkeley Evaluation and Assessment Research Center

Contributing Authors, BEAR:
Joshua Sussman
Perman Gochyyev

Contributing Authors, WestEd:
Kerry Kriener-Althen
Emily Newton
Peter Mangione

2

Table of Contents

Since 2001, the California Department of Education Early Learning and Care Division (CDE ELCD) has collaborated with early childhood assessment experts from WestEd and the Berkeley Evaluation and Assessment Research Center at the University of California, Berkeley (UC BEAR), to provide the Desired Results Developmental Profile (DRDP) formative child assessment system to publicly funded early care and education programs throughout California. This collaboration has resulted in the implementation of three generations of DRDP instruments for CDE ELCD programs.

The most recent generation of this instrument for infants and toddlers and preschool-age children, the DRDP (2015),[1] was developed through a collaboration between the CDE ELCD and the Special Education Division for use in the state's early childhood programs as well as for federal reporting to the Office of Special Education (CDE, 2015a; DRDP Collaborative Research Group, 2018). CDE ELCD further extended the DRDP developmental continua for use in kindergarten[2] (CDE, 2015b); the full continuum from birth through kindergarten will be the focus of this technical report. The goal of the current study is to further establish the measurement approach of the DRDP by assessing internal consistency and internal structure in the expanded version of the assessment. The remainder of this introduction describes key elements of the DRDP assessment system: different versions of the instrument, the DRDP items (called measures), and DRDP interpretation and use. These descriptions summarize the evidence for formative and summative assessment with the DRDP and situate the current study in the DRDP's ongoing program of validation research.

DRDP Views

The DRDP instruments were created through iterative processes, grounded in child development research literature (WestEd, 2018b; WestEd 2018d), developed through consultation with nationally recognized child development experts, and refined through numerous qualitative and quantitative research studies with early childhood and kindergarten teachers. Thus, the DRDP provides a continuum of measurement of children's developmental progress, from early infancy through the end of kindergarten, that can be used to support all children in early learning settings, including dual language learners and children with disabilities or other special needs.

---

[1] Hereinafter referred to as the DRDP, unless multiple generations of the instrument are being discussed.

[2] The Race to the Top–Early Learning Challenge (RTT-ELC) Initiative (American Recovery and Reinvestment Act of 2009), announced on May 25, 2011, by the U.S. Department of Education and U.S. Department of Health and Human Services, was a grant competition that provided awards to states for creating comprehensive plans to transform early learning systems for children, birth to age five, with better coordination and assessment mechanisms, clearer learning standards, and meaningful workforce development and family engagement initiatives. California received one of the first RTT-ELC grants in January 2012, with the goals of improving the quality of early learning programs and closing the readiness gap for subpopulations of young children. One use of California's RTT-ELC grant funds was to build on the DRDP assessment for infant/toddler and preschool settings to provide for developmentally appropriate formative assessment in kindergarten.

The DRDP continuum is presented in three different instrument views, for use in (1) infant/toddler group care settings, (2) preschool settings, and (3) kindergarten classrooms. The three views provide for developmentally appropriate assessment for children within each setting. The instrument is completed by a child's teacher,[3] who uses it to assess the child's knowledge and skills based on ongoing documentation of observations in the setting, reports from family members, and examples of the child's work.

The choice of teacher observation as the method for assessment across the three age ranges allows for consistency across the three views, on definitions of authentic assessment in early childhood education as well as on practical strategies for reliable and valid assessment of young children's competencies. Mathematica Policy Research has asserted that it is wiser "to invest in training teachers to be better observers and more reliable assessors than to spend those resources training and paying for outside assessors to administer on-demand tasks to young children in unfamiliar contexts that will provide data with the added measurement error inherent in assessing young children from diverse backgrounds" (Atkins-Burnett, 2007).

Some researchers have raised concerns about the reliability and validity of assessment data gathered through on-demand performance task assessments, which may not allow children to show all that they know. Specifically, young children may be unable to focus in testing situations, comprehend what is being asked of them, and produce responses that are consistent with their underlying proficiencies (Bagnato et al., 2010; Gelman & Gallistel, 1978; Greenspan & Wieder, 1998; Meisels, 1994, 1996; Meisels & Provence, 1989; National Research Council, 2008). Direct assessments are often mostly a measure of language ability (Shepard et al., 1998), and these concerns are heightened for children with developmental delays (Anastasi, 1988; Bagnato & Neisworth, 1994; Cronbach, 1990) and children who are dual language learners (Barrueco et al., 2012).

Online adaptive inventories are increasingly used in elementary school, including for kindergarten assessment purposes (Poggio & McJunkin, 2012). However, because direct assessments often must be administered individually to children, both time and staffing can be limitations to their use. Although online adaptive inventories can reduce both the time it takes to complete the assessment—because the adaptive nature can reduce the number of items needed—and teacher time—because the child may be able to complete the test independently—children's proficiency with the application and attention to the task may limit the accuracy of the assessments (Clemens et al., 2015).

---

[3] The term "teachers" in this document refers to teachers in early childhood programs, including infant care teachers in CDE-funded infant/toddler programs; teachers in CDE-funded preschool programs, including California State Preschool Programs and Head Start; teachers in transitional kindergarten and traditional kindergarten classrooms; staff in Family Child Care Home Education Network Programs who complete the DRDP (2015) on behalf of family child care home providers; and special education teachers and service providers working in CDE local education agencies.

DRDP Measures and Domains

The items of the DRDP, called "measures," are organized into a comprehensive set of research-based domains and sub-domains. The number of measures and domains vary across the three age-level views in the following ways: (1) the view used in infant/toddler settings has 29 measures within five domains; (2) a second view used in preschool settings has 56 measures across 10 domains or sub-domains; and (3) a third view for use with children in kindergarten has 55 measures across 11 domains or sub-domains (see Table 1 for the list of DRDP domains, number of measures per domain per age-level instrument view, and sub-domains available in the preschool and kindergarten views). In addition to the *comprehensive* set of DRDP domains and sub-domains for each view, two smaller sets of domains and measures, referred to as the *fundamental* set and the *essential* set, are also available to teachers who desire to use a shorter assessment (all sets are aligned with the five "essential domains of readiness" recommended by the National Education Goals Panel and Race to the Top–Early Learning Challenge Initiative [RTT-ELC]; American Recovery and Reinvestment Act of 2009).

Two domains, English Language Development (in the preschool and kindergarten views) and Language and Literacy Development in Spanish (in the kindergarten view), focus on dual language learning and complement the five essential domains. The infant/toddler view does not include a fundamental set; the preschool fundamental view consists of 37 measures within six domains, and the kindergarten fundamental set consists of 37 measures within seven domains. The essential set of measures is available for each view, with 23 measures within five domains in the infant/toddler essential view; 29 measures within six domains in the preschool essential view; and 33 measures within seven domains in the kindergarten essential view.

Table 1. DRDP domains, number of measures per domain per age-level instrument view, and sub-domains available in the preschool and kindergarten views.

Domains	Infant/Toddler (IT) View	Preschool (PS) View	Kindergarten (K) View	Sub-domains (PS and K only)
Approaches to Learning – Self-Regulation (ATL-REG)*	5	7	4	N/A
Social and Emotional Development (SED)*	5	5	5	N/A
Language and Literacy Development (LLD)*	5	10	10	Language (LANG) Literacy (LIT)
Cognition, including Math and Science (COG)*	6	11	10	Math (COG:MATH) Science (COG:SCI)
Physical Development – Health (PD-HLTH)*	8	10	9	Physical Development (PD) Health (HLTH)
History – Social Science (HSS)	N/A	5	5	N/A
Visual and Performing Arts (VPA)	N/A	4	4	N/A
English Language Development (ELD)*	N/A	4	4	N/A
Language and Literacy Development in Spanish (SPAN)*	N/A	N/A	4	N/A
Total number of measures per view (comprehensive set)	29	56	55	

Note. Asterisks (*) indicate domains included in the fundamental and essential sets. ELD and SPAN are "conditional" domains. ELD is completed only when a child in preschool or kindergarten has a home language other than English. SPAN is completed only when a child is enrolled in a kindergarten classroom in which the primary instructional language is Spanish.

Together, the measures within each domain of the DRDP cover the breadth and depth of content across California's early learning goals, which are embodied in the state's Infant/Toddler Learning and Development Foundations (ITLDF; CDE, 2009a), Preschool Learning Foundations (PLF; CDE, 2008, 2010, 2012), and Kindergarten Content Standards (KCS; CDE, 2000, 2001, 2006, 2009b, 2014b), including the Common Core State Standards (CDE, 2013, 2014a) and the Next Generation Science Standards (CDE, 2015c). In California, the term "foundations" is used to define goals for early learning and to emphasize

that the learning and development that happens in the early years of life is foundational to children's lifelong achievement, both in and out of school. As a group, DRDP measures for any given domain cover the full range of knowledge and skills presented in the foundations and standards for that domain. Measures also correspond to the most salient constructs, or the essential knowledge and skills depicted in the standards for each domain. For more information about how the DRDP was initially developed, see the earlier technical report (DRDP Collaborative Research Group, 2018).

Structure and Interpretation of DRDP Measures

Each DRDP measure consists of a measure name, definition, and developmental continuum.[4] Each measure's developmental continuum is expressed as five or more developmental levels, presented in a horizontal progression that is read from left to right. The number of developmental levels varies across infant/toddler, preschool, and kindergarten views of the DRDP, with overlapping levels across the three views.

When developmentally appropriate, the full continuum across all three views describes development from early infancy through early first grade. In this way, the assessment effectively minimizes floor and ceiling effects and provides for appropriate assessment for each age level and linkage between age levels. See Figures 1a–1c for an example of how the Social and Emotional Development (SED) measure "Identity of Self in Relation to Others" is presented in the infant/toddler (Figure 1), preschool (Figure 2), and kindergarten (Figure 3) views, which are similar, with the difference being that the first five levels are only available for rating in the infant/toddler and preschool views of the DRDP and the last two levels are only available for rating in the kindergarten view. Note that Figure 1 (infant/toddler) shows the earliest level (on the far left) as Responding Earlier and the latest ratable level (the eighth level) as Integrating Earlier; Figure 2 (preschool) shows the same; Figure 3 (kindergarten) shows Building Earlier as the earliest ratable level and Integrating Later as the latest level. Levels that appear across views are worded identically (such as the Building Earlier level, which appears in all three instrument views).

---

[4] The Definitions of Terms can be found in Infant/Toddler View, Preschool View, and the DRDP-K.

SED 1: Identity of Self in Relation to Others

Child shows increasing awareness of self as distinct from and also related to others

Mark the latest developmental level the child has mastered:

Responding		Exploring		Building			Integrating
Earlier ○	Later ○	Earlier ○	Later ○	Earlier ○	Middle N/A	Later N/A	Earlier N/A
Responds in basic ways to others	**Uses senses to explore self and others**	**Recognizes self and familiar people**	**Communicates own name and names of familiar people (e.g., "dada," "mama," "grandma," or sibling's name)**	**Expresses simple ideas about self and connection to others**	**Describes self or others based on physical characteristics**	**Describes own preferences or feelings;** *and* **Describes the feelings or desires of family members, friends, or other familiar people**	**Compares own preferences or feelings to those of others**

Possible Examples

• Attends to a familiar adult during feeding. • Quiets when hears a familiar adult. • Grasps an adult's finger when palm of child's hand is touched.	• Examines own hand or foot by looking at it or mouthing it. • Touches others' hair when it is within reach. • Plays with sound by repeating grunts and squeals.	• Orients toward a familiar adult when own name is spoken or signed. • Points to picture of self on the wall. • Smiles when a familiar adult enters the room.	• Communicates, "Me llamo Luis," ["My name is Luis," in Spanish]. • Communicates names of immediate family members in a photo. • Looks to new baby sister and communicates her name.	• Acts out roles from own family in pretend play. • Communicates, "I'm making cookies—just like Grandma!" while rolling play dough. • Draws picture of a house and communicates, "This is my house."	• Communicates, using communication board, "His hair is red!" • Identifies own height, as indicated on a growth chart posted on the wall. • Narrates details while drawing a picture of a friend. • Draws a picture of own family, representing traits such as heights and hair colors.	• Communicates to an adult, "I was mad when it rained because we couldn't go outside." • Communicates that a friend is happy because he is going to have a birthday party. • Says, "Ayokong hawakan ang susô. Na tatakot ako," ["I don't want to touch the snail. It scares me," in Tagalog].	• Selects a pink scarf for a friend whose favorite color is pink, then selects a blue scarf for self. • Communicates to a peer that they both like peanut butter and jelly sandwiches. • Communicates, "我喜歡游泳, 但是我姐姐不喜歡," ["I love to swim, but my sister doesn't," in Chinese].

○ Child is emerging to the next developmental level
○ Unable to rate this measure due to extended absence

SED 1 Identity of Self in Relation to Others **SED 1**

Figure 1. The Identity of Self in Relation to Others measure in the infant/toddler view.

SED 1: Identity of Self in Relation to Others
Child shows increasing awareness of self as distinct from and also related to others

Mark the latest developmental level the child has mastered:

Responding		Exploring		Building			Integrating
Earlier ⬭	Later ⬭	Earlier ⬭	Later ⬭	Earlier ⬭	Middle ⬭	Later ⬭	Earlier ⬭
Responds in basic ways to others	**Uses senses to explore self and others**	**Recognizes self and familiar people**	**Communicates own name and names of familiar people (e.g., "dada," "mama," "grandma," or sibling's name)**	**Expresses simple ideas about self and connection to others**	**Describes self or others based on physical characteristics**	**Describes own preferences or feelings; *and* Describes the feelings or desires of family members, friends, or other familiar people**	**Compares own preferences or feelings to those of others**
Possible Examples • Attends to a familiar adult during feeding. • Quiets when hears a familiar adult. • Grasps an adult's finger when palm of child's hand is touched.	• Examines own hand or foot by looking at it or mouthing it. • Touches others' hair when it is within reach. • Plays with sound by repeating grunts and squeals.	• Orients toward a familiar adult when own name is spoken or signed. • Points to picture of self on the wall. • Smiles when a familiar adult enters the room.	• Communicates, "Me llamo Luis," ["My name is Luis," in Spanish]. • Communicates names of immediate family members in a photo. • Looks to new baby sister and communicates her name.	• Acts out roles from own family in pretend play. • Communicates, "I'm making cookies—just like Grandma!" while rolling play dough. • Draws picture of a house and communicates, "This is my house."	• Communicates, using communication board, "His hair is red!" • Identifies own height, as indicated on a growth chart posted on the wall. • Narrates details while drawing a picture of a friend. • Draws a picture of own family, representing traits such as heights and hair colors.	• Communicates to an adult, "I was mad when it rained because we couldn't go outside." • Communicates that a friend is happy because he is going to have a birthday party. • Says, "Ayokong hawakan ang susô. Na tatakot ako," ["I don't want to touch the snail. It scares me," in Tagalog].	• Selects a pink scarf for a friend whose favorite color is pink, then selects a blue scarf for self. • Communicates to a peer that they both like peanut butter and jelly sandwiches. • Communicates, "我喜歡游泳, 但是我姐姐不喜歡," ["I love to swim, but my sister doesn't," in Chinese].

⬭ Child is emerging to the next developmental level
⬭ Unable to rate this measure due to extended absence

SED 1	Identity of Self in Relation to Others	SED 1

Figure 2. The Identity of Self in Relation to Others measure in the preschool view.

SED 1: Identity of Self in Relation to Others

Child shows increasing awareness of self as distinct from and also related to others

Mark the latest developmental level the child has mastered:

Building			Integrating		
Earlier ⬭	Middle ⬭	Later ⬭	Earlier ⬭	Middle ⬭	Later ⬭
Expresses simple ideas about self and connection to others	**Describes self or others based on physical characteristics**	**Describes own preferences or feelings;** *and* **Describes the feelings or desires of family members, friends, or other familiar people**	**Compares own preferences or feelings to those of others**	**Describes and compares self and others using personality characteristics**	**Identifies and evaluates strengths and weaknesses by comparing self with others**

⎯ Possible Examples ⎯

▸ Acts out roles from own family in pretend play. ▸ Communicates, "I'm making cookies— just like Grandma!" while rolling play dough. ▸ Draws picture of a house and communicates, "This is my house."	▸ Communicates, using communication board, "His hair is red!" ▸ Identifies own height, as indicated on a growth chart posted on the wall. ▸ Narrates details while drawing a picture of a friend. ▸ Draws a picture of own family, representing traits such as heights and hair colors.	▸ Communicates to an adult, "I was mad when it rained because we couldn't go outside." ▸ Communicates that a friend is happy because he is going to have a birthday party. ▸ Says, "I don't want to touch the caterpillar. It scares me."	▸ Selects a pink crayon for a friend whose favorite color is pink, then selects a blue crayon for self. ▸ Communicates to a peer that they both like peanut butter and jelly sandwiches. ▸ Communicates, "I love to swim, but my sister doesn't."	▸ Tells her grandma, "I'm a good friend in school because I share the markers," at the end of the day when she comes to pick her up. ▸ Communicates, "I like to be first to the door, but Michael doesn't care about being first." ▸ Communicates, "I'm shy." ▸ Communicates to a peer, "I like to talk to my friends when I get to school, and Brianna likes to read books when she gets to school."	▸ Communicates to a peer, "I'm great at math, but I'm not so good at reading," after an adult explains they will do a reading activity next. ▸ Moves to the back of a group of children when an adult asks, "Who knows how to use the microscope?" ▸ Communicates, "I think he likes her better than me."

⬭ **Child is emerging to the next developmental level**
⬭ **Unable to rate this measure due to extended absence**

SED 1 (of 5)	Identity of Self in Relation to Others	SED 1 (of 5)

Figure 3. The Identity of Self in Relation to Others measure in the kindergarten view.

9

Each measure's developmental continuum consists of level names, descriptors, and examples. Developmental *level names* are presented at the top of the developmental continuum. Below each level name, the measure's developmental progression is demonstrated through *descriptors* of observable knowledge and skills unique to that DRDP measure. Several *examples* are presented below the descriptors to demonstrate the broad range of behaviors covered by the descriptor and account for individual differences among children.

Teachers complete a rating for a DRDP measure by selecting the latest developmental level that a child has "mastered," which means that the child was observed to demonstrate the descriptor's knowledge and skills consistently in natural settings over time. Examples depicting the descriptors and definitions for each measure guide the teacher's selection of the latest developmental level mastered for each measure. Once a developmental level is selected, teachers may also choose to indicate whether the child may be emerging to the next developmental level.

The DRDP serves multiple purposes, including (1) providing teachers with valid and reliable measurement of individual children's development in key domains; (2) helping teachers observe and reflect on children's learning and development and plan curriculum for both individual children and small groups of children; (3) facilitating reflection with family members about their children's developmental progress and ways to support their children's learning and development in the classroom setting and at home; (4) supporting transition and alignment between infant/toddler programs and preschool, between preschool and kindergarten, and between kindergarten and first grade; (5) guiding professional development for teachers and ongoing quality improvement; and (6) providing state, district, and school administrators with results to inform the development of programs and policies at the state and local levels.

Resources available to support teachers with completing the DRDP and using assessment results include free online information and learning modules; fee-based in-person training; print and web-based tools, including the DRDP Portfolio App; an online rater certification system; and downloadable reports for teachers and parents. The tutorials and webinars are instructional in nature and help teachers learn the basic principles of how and when to complete the assessment; how to observe, document, and reflect on students' learning; and how to use the DRDPOnline™ data system.[5]

Method: Background and Current Desired Results Developmental Profile Studies

The study described in this report is one of a set of four studies designed to address the extent to which the DRDP can be appropriately used both formatively and summatively to assess children's learning and development from early infancy through kindergarten. The three additional studies explored the

---

[5] More information about DRDP training and resources is available at www.desiredresults.us and drdpk.org.

measurement properties of the DRDP when used formatively (DRDP Collaborative Research Group, 2018), including external validity evidence (WestEd & UC BEAR, 2021b) and inter-rater reliability (WestEd & UC BEAR, 2021a). In addition, we have studied a standard-setting process designed to enable state-level aggregate reporting of kindergarten readiness (Kriener-Althen et al., 2020).

CALIBRATION STUDY

The study described in this report was used to establish the measurement approach for the DRDP infancy through kindergarten continuum and to assess the internal consistency and internal structure of the instrument, including item fit and intercorrelation of domains. Pilot and field studies had occurred during instrument development stages in preceding years.

Sample and measures. The sample for the DRDP calibration consisted of 21,490 children, including 2,452 infants/toddlers, 18,629 preschool-age children, and 409 kindergarten children who were primarily enrolled in publicly supported child care programs or kindergarten classrooms throughout California in fall 2014 or spring 2015 (see Tables 2–5 for the demographic characteristics of children from infant/toddler, preschool, and kindergarten settings in the DRDP calibration study). The measures used for the study were the comprehensive set of the three DRDP instrument views (infant/toddler, preschool, kindergarten).

Table 2. Age of children from infant/toddler, preschool, and kindergarten settings in the DRDP calibration study.

	All Children (*N*=21,490)	Infant/ Toddler (*N*=2,452)	Preschool (*N*=18,629)	Kindergarten (*N*=409)
Mean Age (in months)	52.6	26.3	55.8	67.9
Standard Deviation	(12.0)	(9.0)	(7.1)	(2.9)

Table 3. Children's gender from infant/toddler, preschool, and kindergarten settings in the DRDP calibration study.

Gender	All Children (*N*=21,490)	Infant/ Toddler (*N*=2,452)	Preschool (*N*=18,629)	Kindergarten (*N*=409)
Male	50.0 %	50.4 %	49.9 %	52.3 %
Female	49.4 %	48.9 %	49.5 %	46.9 %

Table 4. Demographic characteristics of children from infant/toddler, preschool, and kindergarten settings in the DRDP calibration study.

	All Children (*N*=21,490)	Infant/Toddler (*N*=2,452)	Preschool (*N*=18,629)	Kindergarten (*N*=409)
Dual Language Learners (preschool and kindergarten)	54.0 %	-	54.0 %	54.3 %
IEP or IFSP	2.7 %	1.7 %	2.8 %	8.1 %

Note. IEP = Individualized Education Plan; IFSP = Individualized Family Service Plan

Table 5. Race/ethnicity of children from infant/toddler, preschool, and kindergarten settings in the DRDP calibration study.

Race/Ethnicity	All Children (*N*=21,490)	Infant/Toddler (*N*=2,452)	Preschool (*N*=18,629)	Kindergarten (*N*=409)
Hispanic or Latino	49.8 %	27.7 %	53.8 %	3.2 %
White	9.9 %	14.0 %	9.5 %	NA
Black/African American	8.9 %	8.2 %	8.6 %	25.9 %
Asian or Pacific Islander	1.26 %	3.3 %	1.0 %	1.0 %
Multiple Ethnicities	2.6 %	18.8 %	0.6 %	NA
No Response	27.5 %	27.9 %	26.5 %	69.9 %

Study procedures. Teachers were recruited for the study from the pool of teachers choosing to be "early adopters" of the DRDP (2015) during the 2014–15 academic year, which was a transition year when early childhood programs and kindergarten teachers could choose to use either the DRDP (2015) instrument or a previous generation of the DRDP instrument. The study was announced through communication to programs and school districts via CDE's email distribution list. In addition, study researchers contacted agencies, schools, and teachers from a variety of programs throughout the state to ensure sufficient representation of various sub-populations of children in the study (e.g., young infants, kindergarten-age children, dual language learners, and children with disabilities).

In total, 409 teachers of infants/toddlers, 1,088 teachers of preschool children, and 24 teachers of kindergarten children participated in the study. Teachers participated in online training, distributed information packets and consent forms to parents, and completed the DRDP for children in their classrooms during both fall 2014 and spring 2015. Purposeful sampling was used to include either the fall 2014 or spring 2015 ratings for each child in the calibration sample and to ensure that sufficient

numbers of children from six-month age categories (e.g., 0 to 6 months, 6 to 12 months, 12 to 18 months, 18 to 24 months, … 66 to 72 months), dual language learners, and children with disabilities and other delays were included in the calibration sample, without duplicating individual children.

To complete the DRDP, teachers observed and documented children's developmental competencies over (a) a six-week period after fall enrollment and (b) a six-week period in the spring that was approximately six months after the fall observation period. Following each six-week period of observation and documentation, teachers made rating determinations for each measure and entered the ratings into the online software system. After the close of the data entry period in the spring, data were extracted from the data system and cleaned for analyses. Missing data were removed using listwise deletion, which resulted in child records with complete data for every measure that was appropriate to each child's age and setting.

Analysis approach. Multidimensional Item Response Theory (IRT), specifically the multidimensional partial credit model, was the analytical approach used to calibrate the DRDP (introductions to IRT can be found in Embretson & Reise, 2000, and in Hambleton et al., 1991). IRT was used to produce a set of step difficulties for each measure, and domain scale scores for each child, derived from DRDP measure ratings. IRT is a method of analysis that considers the overall difficulty of each item and the difficulty of steps within each item, allowing direct comparisons of each item and step to the overall performance of an individual child's domain score and to a distribution of such scores.

IRT has multiple advantages. First, IRT allows for different item levels on a measure to vary in their amount of challenge or complexity and thus to reflect different probabilities of a child being assigned a rating at each level of a measure. As applied to the DRDP, this attribute of IRT means that knowledge and skills that typically develop earlier or later than other knowledge and skills in the same domain are accounted for in the quantitative models that produce the domain scales. Second, IRT allows for the analysis of ordinal ratings for assessment items. Thus, a child who has reached a given developmental level of the DRDP has reached or surpassed all earlier levels. Third, IRT scores are accompanied by standard errors of measurement that account for the uncertainty of measurement.

A number of factors can create variability in a child's rating, including who completed it (e.g., different teachers may base their ratings on differing amounts of experience with a child) and when the assessment was completed. Because all assessments contain some such variability, a child's true level of knowledge and skills can never be measured exactly with any assessment tool. Thus, IRT provides a standard error, which represents the area in which one can be confident that a child's true rating lies. The result of the calibration analysis was a continuous scale for each domain of the DRDP, spanning early infancy through early first grade.[6] Domain-level thresholds, the median scale value at which

---

[6] The DRDP continuum extends to early first grade to avoid a ceiling effect when used with children in kindergarten.

children most likely would move from one developmental level to the next across all measures in a domain, were also established using IRT analyses and used for score interpretation.

The analysis approach for this background calibration study was similar to the approach reported for the calibration of the DRDP with children in infant/toddler and preschool settings in the *DRDP (2015) Technical Report* (DRDP Collaborative Research Group, 2018). The current study differed in three key ways: (1) the inclusion of kindergarten children in the calibration model to place infant/toddler, preschool, and kindergarten DRDP domain scores on the same scale; (2) the use of an estimation procedure (Warm's likelihood estimates [WLE], Warm, 1989) that produced profiles of person estimates that were easier for teachers to understand; and (3) the use of methodology to enable direct comparisons of children's domain scale scores across the "essential domains of school readiness" (Shepard et al., 1998).

Items showing significant model misfit were removed or modified as needed. Reasons for unexpected variability in child-level responses at the class group level were investigated as well. Multidimensional IRT models allow the estimation of the latent correlation between the dimensions, by controlling for measurement error and providing a disattenuated[7] estimate of the intercorrelations. The ConQuest software (Wu et al., 2007), which was designed to fit the Multidimensional Random Coefficients Multinomial Logit Model (Adams et al., 1997), was used to perform all IRT calibration analyses.

One drawback inherent in multidimensional models is that they center each dimension separately on zero, prohibiting comparisons between item difficulty or person proficiency estimates across dimensions. It has recently been suggested that an effective strategy for supporting learning in young children is to make conceptually meaningful connections between developmental domains while maintaining domain-specific distinctions (Institute of Medicine and National Research Council, 2015). It is helpful for teachers and parents to have assessment information that helps them understand children's developmental progress in each domain while being able to compare the domain-specific information with progress in other domains. Thus, we used a technique known as delta dimensional alignment (DDA; Feuerstahler & Wilson, 2019, 2021; Schwartz & Ayers, 2011) to make direct comparisons between domains possible.

---

[7] "Disattenuated" means that the correlations have been corrected for measurement error. The typical range for disattenuated domain correlations for technical information reported for early childhood test materials range from 0.3 to 0.8. Disattenuated correlations for other highly respected international assessments, such as PISA, are even higher, ranging from 0.6 to 0.9, with no indication that these tests lack discriminant validity (Organisation for Economic Cooperation and Development [OECD], 2017).

Results

INTERNAL CONSISTENCY RELIABILITY

WLE reliabilities were calculated as measures of internal consistency reliability for all DRDP domains and sub-domains. See Appendix A for DRDP WLE reliabilities by view for the five-dimensional DDA model for the five readiness domains, associated sub-domains, and three unidimensional domains; reliabilities are shown in the white diagonal cells of the tables (which also contain intercorrelations that are discussed in the validity section regarding multidimensionality). The WLE reliabilities ranged from 0.93 to 0.96 for the infant/toddler view, from 0.87 to 0.96 for the preschool view, and from 0.81 to 0.96 for the kindergarten view. The WLE reliabilities across all domains, and all but one sub-domain (HLTH for preschool and for kindergarten), were greater than 0.9, indicating that the DRDP domains and sub-domains have excellent overall precision of measurement. In addition, information about inter-rater reliability can be found in WestEd and UC BEAR (2021a).

VALIDITY

In this section, we will focus on one of the types of validity evidence recommended in *The Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014): evidence based on internal structure. Evidence based on relations to external variables can be found in WestEd and UC BEAR (2020b); evidence based on content can be found in the extensive alignment of the DRDP with the Infant/Toddler Learning and Development Foundations, the Preschool Learning Foundations, and the Kindergarten Content Standards (see WestEd, 2018a, 2018c); and evidence based on response processes can be found in the descriptions of cognitive interviews in DRDP Collaborative Research Group (2018). Validity evidence based on internal structure is provided through item fit statistics, Wright maps, and the dimensionality of the various domains.

Fit statistics. Mean-square fit statistics indicate the degree to which the model predictions and the empirical data are consistent at the individual item level. Because the multidimensional partial credit model is a probabilistic model, some degree of randomness in response is expected (i.e., sometimes a child will be assigned a rating that is slightly lower or slightly higher on a particular measure than what would be predicted). Too little randomness in responses to an item will result in a low fit statistic (usually not regarded as problematic). Too much randomness, however, is regarded as problematic because it indicates that the model is not sufficiently good at predicting which children will receive high ratings and which will receive low ratings; most often, this problem is a result of poor item writing, vague scoring rubrics, or an item that is not well enough aligned with the dimension it is intended to represent.

Table 6 provides the ranges of minimum and maximum fit statistics for the measures within each DRDP domain. Values more than 1.0 indicate that the observed variance is more than the expected variance for the measure (and the data for a measure may be overly random), and the common convention of

4/3 (1.33) is used as an acceptable upper bound. Although values less than 0.75 can be used as a lower bound, indicating that the data for the measure may be overly consistent, the item fit statistics suggest that all the measures fit reasonably well.

Table 6. Ranges of minimum and maximum fit statistics for the measures within each DRDP domain.

Domain	Minimum Fit Statistic	Maximum Fit Statistic
Approaches to Learning – Self-Regulation (ATL-REG)	0.59	1.04
Social and Emotional Development (SED)	0.47	0.52
Language and Literacy Development (LLD)	0.55	1.07
Cognition, including Math and Science (COG)	0.51	1.20
Physical Development – Health (PD-HLTH)	0.60	1.13
History – Social Science (HSS)	0.58	0.83
Visual and Performing Arts (VPA)	0.69	1.01
English Language Development (ELD)	0.50	0.98

Wright map. The theory of child development, on which the measures and developmental levels were based, implies that the difficulty of attaining a particular developmental level should line up reasonably well across all the items within a particular domain. For most domains, this alignment does in fact occur, implying a consistency in how the developmental levels work across items (refer to Figure 4 for the Wright map for the DRDP SED domain that covers infant/toddler, preschool, and kindergarten views of the instrument). Figure 4 shows a modification of an IRT-based graph that is commonly known as a Wright map (Wilson, 2005). The left-hand column shows a distribution of all child IRT-based scores as a histogram; higher levels on the page indicate later levels of development. The Thurstonian thresholds between the developmental levels are shown on the right-hand side of the map; each DRDP measure is represented in one column, with the name of the measure at the bottom; the levels higher on the page indicate later levels of development (e.g., with *RSP_L* indicating the threshold between the Responding Earlier and Responding Later developmental levels and *INTG_L* indicating the threshold between the Integrating Middle and Integrating Later developmental levels).

The interpretation of the level threshold is as follows: the children represented in the histogram on the left, who are at the same vertical position on the map as a measure threshold that is presented to the right of the histogram, have a 50 percent likelihood of being rated at or later than that level on that measure, and a 50 percent likelihood of being rated earlier than that level (given the spacing of the developmental levels, the likelihood of being more than one level later or earlier than one's location is usually very low). For example, in Figure 4, the children depicted in the histogram bar that is the second one from the bottom of the page have a 50 percent likelihood of being rated at or later than the RSP_L

level on the SED1 measure, and a 50 percent likelihood of being rated earlier than that level (given the spacing of the developmental levels). Children above a given level on a measure are more likely to be rated at or later than that level, and children below a level are less likely to be rated at or later than that level; these likelihoods can be computed exactly for any specific child scoring at any specific level using IRT, if desired. Children whose development is below the Responding Later thresholds are more than 50 percent likely to be rated in the Responding Earlier category; Responding Earlier does not appear on the map because the likelihood of being rated at this level is accounted for in relation to the Responding Later threshold.

Note that in Figure 4, the Wright map for SED, the thresholds line up very well with one another within each level. There is a clear space in between each level across all measures, indicating that the measures and the developmental levels within them are performing as expected given our theory of development; again, this type of structure is reasonably consistent across domains and provides evidence supporting the internal structure of the instrument.
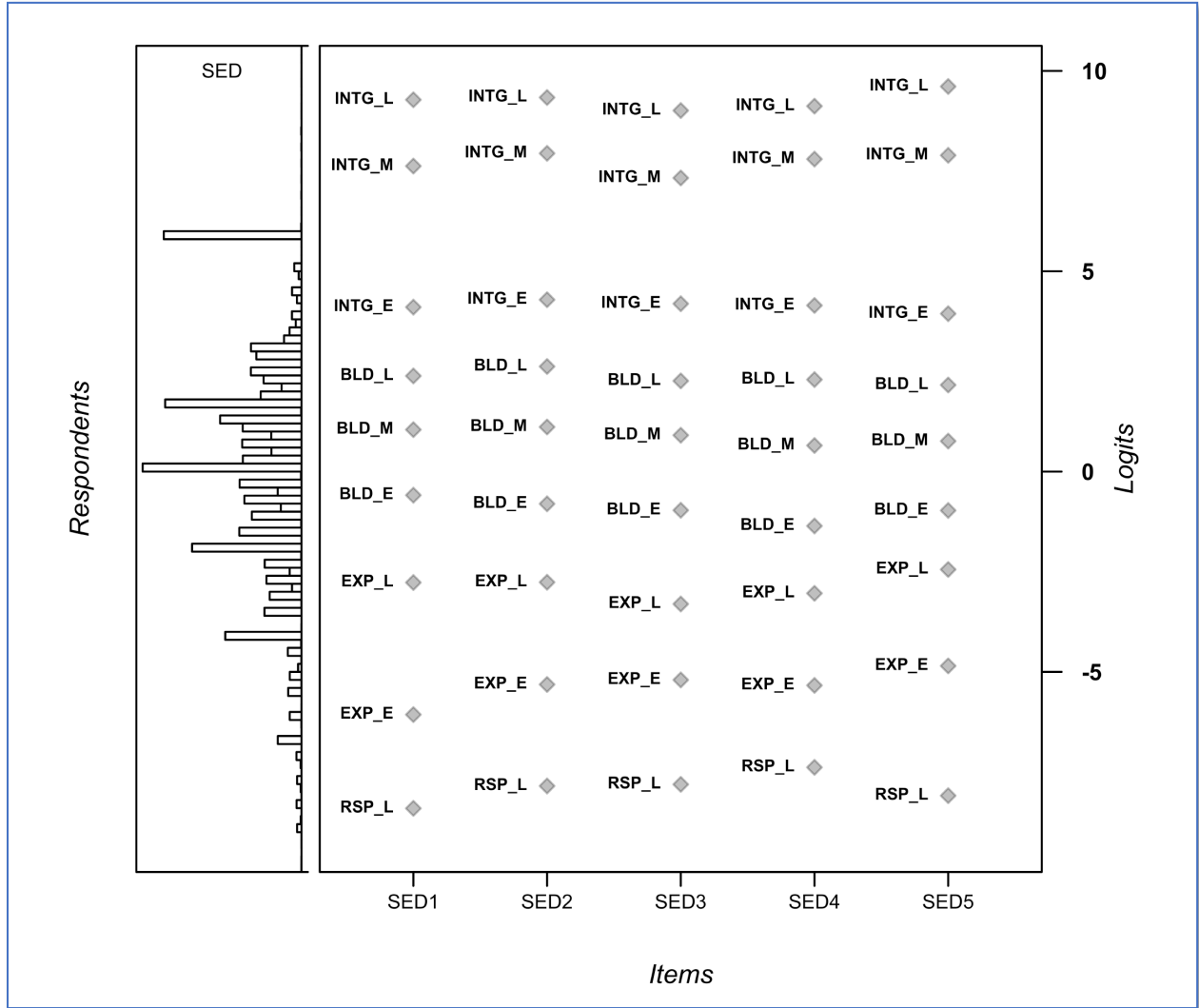
Figure 4. Wright Map for the Social and Emotional Development domain.

Multidimensionality. Another aspect of internal structure that is relevant for a multidimensional assessment is evidence that it is in fact functioning multidimensionally. Three pieces of evidence are presented here, including attenuated and disattenuated correlations between domains and sub-domains, model fit, and the percentage of children who showed a significant discrepancy on at least one of the five essential readiness domains.

First, the fit of a multidimensional partial credit model was compared to the fit of a unidimensional partial credit model for the five readiness domains (to which a multidimensional model was fit), using a likelihood ratio test, which follows a X^2 distribution with degrees of freedom equal to the number of additional parameters in the more complex model. The multidimensional model shows better fit ($X^2 =$ 47,149, *df* = 18, *p* <.001) in the sense of statistical significance.

Second, the correlations between the dimensions were examined. To reduce the effect of children's ages on the correlations, we compute them separately for the infant/toddler, preschool, and kindergarten views. One potential criticism that has been levied against the DRDP (Stipek et al., 2019) is that the correlations between dimensions are "too high," and that this may indicate a lack of discriminant validity. However, we point out three reasons this criticism is incorrect.

First, we examined relative model fit, comparing a multidimensional partial credit model to a unidimensional partial credit model for these data, using a likelihood ratio test, which follows a chi-square distribution with degrees of freedom equal to the number of additional parameters in the more complex model. The multidimensional model shows better fit (X2 = 47,149, df == 18, p < 0.001) in the sense of statistical significance.

Second, it is important to note that, following the practices endorsed in *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), we have reported both the disattenuated correlations, which have been corrected for measurement error, and the raw (attenuated) correlations, which are most often used by researchers who are less experienced with using disattenuated correlations. Disattenuated correlations will be higher than raw correlations (although the latter is what has been most often reported in test manuals and studies in the early childhood literature). These correlations are shown separately for each DRDP instrument view, in the tables in Appendix A; attenuated correlations are below the diagonal and disattenuated above the diagonal.

For the infant/toddler view, raw correlations ranged from 0.93 to 0.98, and disattenuated correlations ranged from 0.94 to 0.99. These correlations are high, as expected, likely due to rapid development and less differentiation between domains at the younger ages. For the preschool view, raw correlations among domains ranged from 0.57 to 0.90 (this range does not include correlations of domains with sub-domains, some of which are very high, because sub-domains share measures with the domains with which they are associated). Disattenuated correlations for the preschool view ranged from 0.61 to 0.94. Similarly, raw correlations between domains for the kindergarten view ranged from 0.46 to 0.90, and disattenuated correlations ranged from 0.51 to 0.94.

As a reference, other assessments (e.g., Programme for International Student Achievement (PISA); Organisation for Economic Cooperation and Development [OECD], 2016) have shown similarly high disattenuated correlations between different dimensions (including math, reading, and science, but also subscales such as collaborative problem solving and financial literacy), but this has not been interpreted as evidence for lack of discriminant validity of the test. Rather, it is interpreted as consistency in student development. For comparative purposes, we have presented both attenuated and disattenuated correlations among subtests for several commonly used early childhood tests in Table 7, as well as for the PISA. It can be seen that these correlations cover a wide range, from low (e.g., 0.20) to very high (e.g., 0.89); in the PISA tests, they tend to be higher.

Table 7. Disattenuated correlations among subtests for several commonly used early childhood tests and the PISA test.

Assessment	Age Range	Areas Assessed	Disattenuated	Range of Correlations
Bayley Scales of Infant and Toddler Development	Infant/toddler	Receptive Communication, Expressive Communication, Gross Motor, Fine Motor, Cognitive, Social–Emotional	Yes	0.20–0.60
Wechsler Preschool and Primary Scale of Intelligence	2 years and 6 months to 3 years and 11 months	Information, Receptive Vocabulary, Picture Naming, Block Design, Object Assembly, Picture Memory, Zoo Locations	Yes	0.33–0.74
Early Development Instrument	Kindergarten	Physical Health and Well-Being, Social Competence, Emotional Maturity, Language and Cognitive Development, Communication Skills and General Knowledge	Yes	0.52–0.89
Kindergarten Observation Form	Kindergarten entry	Self-Care and Motor Skills, Self-Regulation, Social Expression, Kindergarten Academics	?	Unknown/unpublished
PISA 2015	15 years	Math, Reading, Science, Collaborative Problem Solving, Financial Literacy	Yes	0.64–0.88 (International) 0.71–0.90 (USA)

Third, because correlations are a limited way of looking at dimensionality, it is useful to examine what can be gained from using a multidimensional approach to assessing child development, which aligns with the design of the DRDP (DRDP Collaborative Research Group, 2018), rather than simply a unidimensional approach, as some (e.g., Stipek et al., 2019) have suggested. To that end, we undertook an analysis to determine to what extent children showed notable discrepancies in one or more of the five DRDP readiness domains (Sussman & Gochyyev, 2019). Using data for 37,672 children rated on the preschool view in fall 2017, we computed both a unidimensional overall developmental scaled score across the five readiness domains and separate domain scaled scores for each of the five readiness domains in the multidimensional model. We then compared each of the multidimensional scores with the unidimensional score to determine whether they showed a statistically significant discrepancy.

The results (refer to Table 8) show that there were statistically significant discrepancies between unidimensional and multidimensional development on at least one of the domain scales for between 11 percent (for LLD) and 27 percent (for PD-HLTH) of children. Given the size of the sample, these percentages represent roughly 4,000 to 10,000 children, depending on the domain. We also divided the discrepancies by the mean estimated development per month to determine the approximate number of months associated with each statistically significant discrepancy. The median discrepancy between unidimensional and multidimensional estimates of development ranged from 6.7 months (for LLD) to 10.5 months (for ATL-REG). This suggests that the unidimensional model may imprecisely estimate children's learning and development, possibly overestimating or underestimating learning and development by six months or more, which is a substantial amount of time for children whose ages are often described in months rather years. The precision of measurement provided with the multidimensional DDA DRDP model supports teachers with identifying variations in children's development and appropriately differentiating instruction (Mangione et al., 2019).

Table 8. Comparison of child development using a unidimensional versus a multidimensional model.

Domain	% discrepant	Median discrepancy between unidimensional and multidimensional model in months
ATL-REG	14.6	10.5
SED	11.9	9.7
LLD	11.3	6.7
COG	13.7	7.8
PD-HLTH	27.0	7.6

Note. ATL-REG = Approaches to Learning–Self-Regulation, COG = Cognition, Including Math and Science, LLD = Language and Literacy Development, PD-HEALTH = Physical Development – Health, SED = Social and Emotional Development.

Evidence related to invariance in the DRDP measures has been gathered through several differential item functioning (DIF) analyses that examined uniform, measure-level DIF for focal groups defined by biological sex, DLL status, various racial/ethnic group membership, and disability status (whether a child has an Individualized Education Plan [IEP], or an Individualized Family Service Plan [IFSP]). DIF parameters and standard errors (estimated using the MRCML via the ConQuest software) are shown in Appendix A. DIF values (absolute values) below 0.426 are considered negligible, values between 0.426 and 0.638 are considered moderate, and values above 0.638 are considered large (Paek & Wilson, 2011).

This set of analyses—DIF based on sex/gender (female versus male), IEP/IFSP status (yes versus no), DLL status (DLL versus monolingual English speaker), and a set of comparisons of various racial and ethnic groups to non-Latinx White students—was performed using data from the DRDPtech system that was gathered between 2015 and 2018. For each of these analyses, a random sample of the data were drawn from the overall datasets, with over-sampling of smaller but important groups (e.g., monoracial Native American). All DIF comparisons are based on data from at least 25,000 children. These analyses resulted in 405 total DIF parameters, all of which were in the negligible range; values ranged from -0.409, in favor of the focal group, to 0.389, in favor of the reference group (non-Latinx, White). Taken as a whole, these results provide good evidence that the DRDP measures are invariant across gender, DLL status, disability status, and race/ethnicity.

Discussion

California, with the development and use of the DRDP, has elected to offer a whole-child approach to assess children's developmental progress from early infancy through kindergarten.

The extensive background studies described in this paper, as well as others (WestEd, 2018 a–d; WestEd & BEAR, 2021a, 2021b) establish that the DRDP was developed using best practices in the field. It was aligned to the California Infant/Toddler Learning and Development Foundations, Preschool Learning Foundations, and Kindergarten Readiness Standards (WestEd, 2018a, 2018c), where were based on the expertise of nationally recognized scholars in early childhood development. Internal consistency reliability is uniformly high for each domain, and inter-rater reliability is moderately high to high. It would be useful to conduct studies of rater consistency over time, with video recordings or documentation of the behaviors of specific children used to control for growth.

Extensive validity evidence around instrument content, response processes, and internal structure has been collected and is publicly available (DRDP Collaborative Research Group, 2018), or presented in the infant/toddler, preschool, and kindergarten calibration described in this document. The DRDP assessment, designed using multidimensional item response modeling with dimensional alignment, can be used to make comparisons between domains for individual children and facilitate the design of opportunities to support individualized learning and development.

Additional validity evidence to investigate associations of the DRDP domains with external assessments should be collected, with larger and more socioeconomically diverse samples, to more thoroughly examine the relationships between the DRDP and other assessments of learning and development that would be expected to show higher and lower correlations with the DRDP. In addition, evidence regarding the consequences of using the DRDP on teacher practices, teacher understanding of child development, and child outcomes should be collected. This could include studies of the effects of child care quality on children's learning and development outcomes.

Studies of the predictive value of the DRDP over time would also be very useful. For example, child DRDP scores at the preschool or kindergarten level could be matched with state-level testing outcomes in third grade to study potential relationships between DRDP domains and third-grade reading and mathematics test scores. In addition, the DRDP could be used to bridge the gap between observational assessment in infant/toddler, preschool, and kindergarten, and on-demand testing in third grade, to gain a more nuanced understanding of the relationships between child development and learning during the critical early education years and later educational outcomes. The DRDP could be expanded to include innovative assessment strategies: structured prompts or scenarios could be designed, such as activities in which individual students or small groups of students could participate. Teachers could then observe and rate students' behavior on an ordered developmental scale that extends the current DRDP levels. Results from such an assessment in the early grades could help first- and second-grade teachers identify the strengths and limitations of individual children and tailor instruction accordingly. Finally, additional studies of the fairness of the DRDP (e.g., DIF regarding other important sub-populations) are needed.

SIGNIFICANCE OF THE DESIRED RESULTS DEVELOPMENTAL PROFILE

The DRDP fills an important gap in the field of early childhood assessment because it was developed for both (a) formative curriculum planning to support individuals and groups of children and (b) summative state-level reporting of children's progress toward readiness. Five DRDP domains—(1) ATL-REG, (2) SED, (3) LLD, (4) COG, and (5) PD-HLTH—align with the "essential domains of school readiness" recommended by the National Education Goals Panel and by the federal government through RTT-ELC. The DRDP allows for documentation of learning and development for individual children along a continuum, which has relevance for other states undergoing similar efforts to develop standards-aligned early childhood and school readiness assessments. The assessment continua correspond to early learning foundations and kindergarten standards—spanning infancy, toddlerhood, preschool, and kindergarten—to support the teachers' ability to monitor children's progress throughout the year and the states' ability to document children's development as they approach school entry. The studies presented here provide evidence for how the DRDP serves as a valid and reliable assessment of children's learning and development from early infancy through kindergarten.

Limitations

A primary limitation of these studies includes the low numbers of participating kindergarten classrooms and children. Because the use of the DRDP in California kindergartens is voluntary, and various kindergarten assessments are used across local school districts, DRDP data for this age group has always been much less available than that for the infant/toddler or preschool groups. Additional research regarding the DRDP's performance in kindergarten is needed.

The second major limitation of these studies is that the samples in these studies are not representative of the population of California as a whole. The studies' research samples primarily consist of children enrolled in publicly funded early childhood programs. Hence, the children in the sample are more likely a representative sample of California children who live in families with low socioeconomic status (SES), identify as a race/ethnicity other than White, and live in households where a language other than English is primarily used. Additional research involving children enrolled in child care settings that are not publicly funded would provide highly useful information about the DRDP.

Acknowledgments

Appendix A

The tables in this appendix show the WLE reliabilities on the diagonal, with the attenuated correlations below the diagonal, and the disattenuated correlations above the diagonal. Tables 9, 10, and 11 present this information for the Infant/Toddler view data, Preschool view, and Kindergarten view, respectively.

Table 9. Reliability and correlations between DRDP domains for the Infant/Toddler view.

	ATL-REG	SED	LLD	COG	PD-HLTH
ATL-REG	0.93	0.99	0.96	0.97	0.94
SED	0.93	0.94	0.98	0.98	0.96
LLD	0.91	0.93	0.96	0.97	0.95
COG	0.91	0.93	0.93	0.95	0.97
PD-HLTH	0.89	0.91	0.91	0.92	0.96

Note. WLE reliabilities are on the diagonal and represented by data cells that are filled in white. Attenuated correlations are below the diagonal and represented by data cells that are filled in orange. Disattenuated correlations are above the diagonal and represented by data cells that are filled in green.

25

Table 10. Reliability and correlations between DRDP domains and subdomains for the Preschool view.

	ATL-REG	SED	LLD	COD	PD-HLTH	MATH*	SCI*	LANG*	LIT*	PD*	HLTH*	VPA	HSS	ELD
ATL-REG	0.93	0.93	0.90	0.88	0.86	0.88	0.88	0.89	0.89	0.84	0.84	0.81	0.87	0.62
SED	0.87	0.94	0.94	0.89	0.87	0.88	0.90	0.94	0.92	0.85	0.86	0.84	0.88	0.65
LLD	0.85	0.90	0.96	0.94	0.87	0.94	0.92	1.00	1.00	0.85	0.85	0.84	0.88	0.71
COD	0.83	0.84	0.90	0.95	0.87	1.00	1.00	0.89	0.96	0.86	0.84	0.85	0.90	0.68
PD-HLTH	0.81	0.83	0.83	0.83	0.96	0.86	0.87	0.86	0.86	1.00	1.00	0.86	0.88	0.61
MATH*	0.81	0.82	0.89	0.99	0.81	0.92	0.97	0.89	0.97	0.85	0.84	0.84	0.89	0.68
SCI*	0.81	0.83	0.86	0.96	0.81	0.89	0.92	0.89	0.93	0.86	0.87	0.85	0.90	0.67
LANG*	0.83	0.88	0.96	0.84	0.81	0.83	0.82	0.93	0.93	0.85	0.84	0.81	0.85	0.71
LIT*	0.82	0.84	0.96	0.88	0.80	0.88	0.85	0.86	0.90	0.84	0.85	0.87	0.92	0.71
PD*	0.78	0.80	0.81	0.81	0.96	0.79	0.80	0.79	0.77	0.93	0.94	0.84	0.84	0.61
HLTH*	0.76	0.78	0.78	0.77	0.94	0.75	0.77	0.76	0.75	0.84	0.87	0.88	0.92	0.60
VPA	0.75	0.78	0.79	0.79	0.80	0.77	0.78	0.75	0.79	0.77	0.79	0.91	0.95	0.64
HSS	0.80	0.81	0.83	0.83	0.82	0.81	0.82	0.78	0.83	0.77	0.82	0.86	0.91	0.65
ELD	0.58	0.61	0.67	0.63	0.57	0.63	0.62	0.66	0.64	0.56	0.54	0.58	0.59	0.92

Note. Items with asterisks (*) are subdomains (e.g., lang and lit are subdomains of lld); domains will clearly show very high correlations with associated subdomains. WLE reliabilities are on the diagonal and represented by data cells that are filled in white. Attenuated correlations are below the diagonal and represented by data cells that are filled in orange. Disattenuated correlations are above the diagonal and represented by data cells that are filled in green.

Table 11. Reliability and correlations between DRDP domains and subdomains for the Kindergarten view.

	ATL-REG	SED	LLD	COD	PD-HLTH	MATH*	SCI*	LANG*	LIT*	PD*	HLTH*	VPA	HSS	ELD
ATL-REG	0.93	0.94	0.88	0.88	0.88	0.88	0.82	0.87	0.82	0.88	0.98	0.89	0.92	0.51
SED	0.88	0.94	0.85	0.86	0.90	0.85	0.78	0.86	0.77	0.88	1.00	0.89	0.93	0.53
LLD	0.83	0.81	0.96	0.94	0.79	0.93	0.89	1.00	0.99	0.77	1.00	0.92	0.84	0.61
COD	0.83	0.81	0.90	0.95	0.86	1.00	0.96	0.90	0.92	0.84	1.00	0.94	0.87	0.57
PD-HLTH	0.84	0.85	0.76	0.83	0.96	0.86	0.78	0.77	0.77	1.00	1.00	0.96	0.91	0.49
MATH*	0.82	0.79	0.87	0.98	0.81	0.92	0.88	0.88	0.91	0.84	0.99	0.91	0.87	0.57
SCI*	0.76	0.72	0.84	0.90	0.73	0.81	0.92	0.87	0.84	0.76	1.00	0.96	0.80	0.54
LANG*	0.81	0.81	0.95	0.85	0.72	0.82	0.81	0.93	0.86	0.75	0.99	0.89	0.83	0.60
LIT*	0.75	0.70	0.92	0.85	0.71	0.83	0.76	0.79	0.90	0.75	1.00	0.95	0.79	0.59
PD*	0.81	0.82	0.73	0.79	0.98	0.78	0.70	0.69	0.69	0.93	1.00	0.94	0.89	0.46
HLTH*	0.85	0.88	0.88	0.89	0.98	0.85	0.90	0.86	0.85	0.90	0.81	1.00	1.00	0.56
VPA	0.82	0.82	0.86	0.87	0.89	0.83	0.88	0.82	0.85	0.87	0.87	0.91	0.98	0.50
HSS	0.85	0.85	0.79	0.81	0.84	0.79	0.73	0.77	0.71	0.81	0.92	0.89	0.91	0.56
ELD	0.47	0.50	0.57	0.53	0.46	0.52	0.49	0.56	0.53	0.42	0.48	0.46	0.51	0.92

Note. Items with asterisks (*) are subdomains of related domains (e.g., lang and lit are subdomains of lld); subdomains will show lower reliability and very high correlations with associated domains. WLE reliabilities are on the diagonal and represented by data cells that are filled in white. Attenuated correlations are below the diagonal and represented by data cells that are filled in orange. Disattenuated correlations are above the diagonal and represented by data cells that are filled in green.

Appendix B

The tables in this appendix present the DIF values for each domain, first by gender, disability status, and DLL status, followed by race/ethnicity categories. These DIF values provide good evidence that the DRDP measures are invariant across gender, DLL status, disability status, and race/ethnicity.

Table 12. DIF values for gender, IEP status, and DLL status for measures in the Approaches to Learning—Self-Regulation domain.

	Female	IEP	DLL
Attention Maintenance	-0.107 [0.005]	0.005 [0.005]	0.012 [0.006]
Self-Comforting	0.060 [0.005]	0.025 [0.006]	-0.098 [0.006]
Imitation	-0.042 [0.006]	-0.045 [0.006]	0.104 [0.006]
Curiosity and Initiative in Learning	0.195 [0.004]	-0.089 [0.005]	0.097 [0.005]
Self-Control of Feelings and Behavior	-0.053 [0.004]	0.140 [0.004]	-0.070 [0.005]
Engagement and Persistence	0.005 [0.005]	-0.053 [0.005]	-0.021 [0.005]
Shared Use of Space and Materials	-0.058 [0.012]	0.017 [0.012]	-0.023 [0.013]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 13. DIF values for gender, IEP status, and DLL status for measures in the Social Emotional domain.

	Female	IEP	DLL
Identity of Self in Relation to Other	0.019 [0.005]	0.046 [0.005]	0.088 [0.005]
Social and Emotional Understanding	-0.006 [0.005]	0.088 [0.005]	0.047 [0.005]
Relationships and Social Interactions with Familiar Adults	0.026 [0.005]	-0.102 [0.005]	0.012 [0.005]
Relationships and Social Interactions with Peers	0.059 [0.005]	-0.024 [0.005]	-0.085 [0.005]
Symbolic and Sociodramatic Play	-0.098 [0.009]	-0.008 [0.009]	-0.062 [0.005]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 14. DIF values for gender, IEP status, and DLL status for measures in the Language and Literacy domain.

	Female	IEP	DLL
Understanding of Language (Receptive)	0.025 [0.006]	-0.224 [0.006]	-0.042 [0.006]
Responsiveness to Language	-0.078 [0.006]	0.213 [0.006]	-0.031 [0.006]
Communication and Use of Language (Expressive)	-0.010 [0.005]	0.109 [0.005]	0.033 [0.005]
Reciprocal Communication and Conversation	-0.026 [0.005]	-0.086 [0.005]	-0.144 [0.005]
Interest in Literacy	-0.016 [0.005]	0.021 [0.005]	0.061 [0.005]
Comprehension of Age-Appropriate Text	0.011 [0.005]	0.037 [0.005]	0.062 [0.005]
Concepts About Print	0.043 [0.005]	0.015 [0.005]	-0.001 [0.005]
Phonological Awareness	-0.056 [0.005]	-0.012 [0.005]	-0.054 [0.005]
Letter and Word Knowledge	0.037 [0.005]	-0.092 [0.005]	0.121 [0.005]
Emergent Writing	0.049 [0.005]	-0.029 [0.005]	0.008 [0.005]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 15. DIF values for gender, IEP status, and DLL status for measures in the Cognition domain.

	Female	IEP	DLL
Spatial Relationships	0.023 [0.017]	0.049 [0.017]	-0.012 [0.017]
Classification	0.033 [0.004]	-0.009 [0.004]	0.061 [0.005]
Number Sense of Quantity	-0.145 [0.005]	-0.170 [0.005]	-0.262 [0.005]
Number Sense of Math Operations	0.006 [0.005]	-0.004 [0.005]	0.002 [0.005]
Measurement	-0.015 [0.004]	0.332 [0.004]	0.147 [0.005]
Patterning	-0.035 [0.004]	0.357 [0.005]	0.097 [0.004]
Shapes	-0.105 [0.005]	-0.031 [0.005]	-0.052 [0.005]
Cause and Effect	0.016 [0.005]	0.087 [0.005]	0.032 [0.005]
Inquiry through Observation and Investigation	0.015 [0.005]	-0.131 [0.005]	-0.074 [0.005]
Documentation and Communication of Inquiry	0.055 [0.005]	-0.087 [0.005]	-0.024 [0.005]
Knowledge of the Natural World	0.177 [0.014]	-0.344 [0.005]	0.073 [0.014]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 16. DIF values for gender, IEP status, and DLL status for measures in the Physical Development – Health domain.

PD-HLTH Skills Assessed	Female	IEP	DLL
Perceptual-Motor Skills and Movement Concepts	0.146 [0.005]	-0.018 [0.005]	0.005 [0.005]
Gross Locomotor Movement Skills	0.389 [0.005]	-0.134 [0.005]	-0.025 [0.005]
Gross Motor Manipulative Skills	-0.098 [0.005]	0.193 [0.005]	0.106 [0.005]
Fine Motor Manipulative Skills	-0.025 [0.008]	-0.009 [0.008]	0.365 [0.008]
Safety	0.266 [0.005]	-0.104 [0.005]	0.001 [0.005]
Personal Care Routines:	0.306 [0.005]	-0.091 [0.005]	-0.042 [0.005]
Personal Care Routines: Hygiene	-0.155 [0.005]	0.049 [0.005]	-0.096 [0.005]
Personal Care Routines: Feeding	-0.252 [0.005]	0.137 [0.005]	-0.066 [0.005]
Personal Care Routines: Dressing	-0.212 [0.008]	-0.130 [0.008]	0.014 [0.008]
Active Physical Play	-0.166 [0.005]	0.079 [0.005]	-0.105 [0.005]
Nutrition	-0.051 [0.0056]	-0.061 [0.006]	-0.050 [0.006]
Knowledge of Wellness	-0.148 [0.019]	0.089 [0.019]	-0.108 [0.019]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 17. DIF values for race/ethnicity for measures in the Approaches to Learning—Self-Regulation domain.

	Monoracial Black	Multiracial Black	Monoracial Latinx	Latinx + White	Monoracial Native American	Multiracial Native American
Attention Maintenance	-0.071 [0.008	-0.045 [0.009]	-0.062 [0.008]	0.042 [0.008]	-0.013 [0.015]	0.036 [0.009]
Self-Comforting	-0.139 [0.008]	-0.114 [0.009]	-0.214 [0.008]	0.201 [0.008]	-0.066 [0.015]	-0.116 [0.009]
Imitation	-0.185 [0.008]	-0.154 [0.010]	0.028 [0.008]	0.027 [0.008]	-0.064 [0.015]	0.193 [0.009]
Curiosity and Initiative in Learning	0.086 [0.006]	0.067 [0.008]	0.195 [0.006]	0.159 [0.006]	0.095 [0.012]	0.148 [0.007]
Self-Control of Feelings and Behavior	-0.008 [0.006]	-0.014 [0.008]	-0.092 [0.006]	0.063 [0.006]	0.000 [0.012]	-0.119 [0.007]
Engagement and Persistence	0.168 [0.007]	0.124 [0.008]	0.083 [0.007]	0.069 [0.007]	0.007 [0.013]	-0.043 [0.008]
Shared Use of Space and Materials	0.149 [0.018]	0.136 [0.021]	0.062 [0.018]	0.051 [0.018]	0.041 [0.034]	-0.099 [0.020]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 18. DIF values for race/ethnicity for measures in the Social Emotional domain.

	Monoracial Black	Multiracial Black	Monoracial Latinx	Latinx + White	Monoracial Native American	Multiracial Native American
Identity of Self in Relation to Others	0.018 [0.006]	0.022 [0.008]	0.088 [0.005]	0.006 [0.006]	0.056 [0.012]	0.044 [0.007]
Social and Emotional Understanding	0.022 [0.006]	-0.013 [0.008]	0.047 [0.005]	0.006 [0.006]	0.017 [0.012]	-0.021 [0.007]
Relationships and Social Interactions with Familiar Adults	-0.044 [0.006]	-0.020 [0.008]	0.012 [0.005]	0.006 [0.006]	-0.035 [0.012]	0.101 [0.007]
Relationships and Social Interactions with Peers	-0.005 [0.007]	-0.017 [0.008]	-0.085 [0.005]	0.006 [0.007]	-0.045 [0.012]	-0.108 [0.007]
Symbolic and Sociodramatic Play	0.009 [0.013]	0.028 [0.015]	-0.062 [0.009]	0.013 [0.013]	0.007 [0.025]	-0.015 [0.014]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 19. DIF values for race/ethnicity for measures in the Language and Literacy domain.

	Monoracial Black	Multiracial Black	Monoracial Latinx	Latinx + White	Monoracial Native American	Multiracial Native American
Understanding of Language (Receptive)	-0.028 [0.006]	-0.125 [0.011]	-0.157 [0.009]	0.143 [0.009]	-0.069 [0.017]	0.096 [0.010]
Responsiveness to Language	0.079 [0.007]	-0.025 [0.010]	-0.051 [0.008]	-0.085 [0.008]	-0.061 [0.016]	-0.101 [0.009]
Communication and Use of Language (Expressive)	0.028 [0.007]	-0.019 [0.009]	0.016 [0.008]	0.002 [0.008]	-0.167 [0.015]	0.018 [0.009]
Reciprocal Communication and Conversation	-0.021 [0.006]	-0.016 [0.008]	-0.112 [0.007]	-0.131 [0.007]	0.037 [0.013]	-0.106 [0.008]
Interest in Literacy	-0.077 [0.006]	-0.035 [0.008]	0.033 [0.007]	0.097 [0.007]	0.180 [0.013]	0.066 [0.008]
Comprehension of Age-Appropriate Text	-0.040 [0.007]	0.028 [0.009]	0.103 [0.007]	0.101 [0.007]	0.126 [0.014]	0.075 [0.008]
Concepts About Print	0.036 [0.007]	0.089 [0.009]	0.077 [0.007]	0.068 [0.007]	0.064 [0.014]	0.034 [0.008]
Phonological Awareness	0.085 [0.007]	0.061 [0.009]	-0.019 [0.007]	-0.036 [0.007]	-0.019 [0.014]	-0.118 [0.008]
Letter and Word Knowledge	-0.025 [0.007]	0.036 [0.009]	0.113 [0.007]	0.115 [0.007]	0.067 [0.014]	0.014 [0.008]
Emergent Writing	-0.038 [0.020]	0.007 s[0.009]	0.022 [0.008]	0.042 [0.008]	0.083 [0.015]	0.052 [0.009]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 20. DIF values for race/ethnicity for measures in the Cognition domain.

	Monoracial Black	Multiracial Black	Monoracial Latinx	Latinx + White	Monoracial Native American	Multiracial Native American
Spatial Relationships	-0.114 [0.009]	-0.001 [0.029]	-0.023 [0.024]	-0.030 [0.024]	-0.076 [0.046]	-0.031 [0.027]
Classification	-0.056 [0.008]	-0.074 [0.008]	0.013 [0.006]	0.010 [0.006]	0.002 [0.012]	0.139 [0.007]
Number Sense of Quantity	-0.008 [0.008]	0.074 [0.008]	-0.192 [0.007]	-0.151 [0.007]	-0.001 [0.013]	-0.409 [0.008]
Number Sense of Math Operations	-0.005 [0.007]	-0.006 [0.008]	0.005 [0.007]	-0.011 [0.007]	-0.005 [0.013]	0.094 [0.007]
Measurement	-0.071 [0.007]	-0.056 [0.008]	0.119 [0.006]	0.101 [0.006]	0.013 [0.012]	0.138 [0.007]
Patterning	0.048 [0.007]	-0.100 [0.100]	0.005 [0.006]	0.011 [0.006]	-0.004 [0.012]	0.161 [0.007]
Shapes	0.110 [0.007]	-0.031 [0.008]	-0.088 [0.007]	-0.093 [0.007]	0.012 [0.013]	0.018 [0.007]
Cause and Effect	0.065 [0.007]	0.044 [0.008]	0.029 [0.007]	0.012 [0.007]	0.029 [0.013]	0.023 [0.007]
Inquiry through Observation and Investigation	0.012 [0.007]	0.066 [0.008]	0.006 [0.007]	0.002 [0.007]	-0.022 [0.013]	-0.095 [0.008]
Documentation and Communication of Inquiry	0.047 [0.008]	0.024 [0.008]	-0.008 [0.007]	-0.015 [0.007]	-0.062 [0.013]	-0.074 [0.008]
Knowledge of the Natural World	-0.027 [0.024]	0.058 [0.024]	0.112 [0.020]	0.135 [0.020]	0.037 [0.038]	0.005 [0.022]

Note. Standard errors (SE) are included in brackets following the estimate.

Table 21. DIF values for race/ethnicity for measures in the Physical Development – Health domain.

	Monoracial Black	Multiracial Black	Monoracial Latinx	Latinx + White	Monoracial Native American	Multiracial Native American
Perceptual-Motor Skills and Movement Concepts	0.034 [0.007]	0.010 [0.008]	0.020 [0.007]	0.023 [0.007]	0.066 [0.013]	0.004 [0.008]
Gross Locomotor Movement Skills	-0.126 [0.007]	-0.106 [0.009]	-0.049 [0.007]	-0.043 [0.007]	0.024 [0.014]	-0.101 [0.008]
Gross Motor Manipulative Skills	-0.051 [0.008]	-0.028 [0.009]	0.108 [0.008]	0.039 [0.008]	0.036 [0.015]	0.045 [0.009]
Fine Motor Manipulative Skills	-0.026 [0.011]	0.160 [0.013]	0.223 [0.011]	0.197 [0.011]	0.313 [0.022]	0.289 [0.013]
Safety	0.001 [0.007]	-0.005 [0.008]	0.015 [0.007]	0.055 [0.007]	-0.019 [0.013]	0.027 [0.008]
Personal Care Routines:	-0.071 [0.007]	-0.064 [0.008]	-0.066 [0.007]	-0.030 [0.007]	-0.012 [0.013]	-0.051 [0.008]
Personal Care Routines: Hygiene	0.086 [0.007]	0.028 [0.008]	-0.038 [0.007]	-0.023 [0.007]	-0.006 [0.013]	-0.059 [0.008]
Personal Care Routines: Feeding	0.081 [0.007]	0.026 [0.009]	-0.056 [0.007]	-0.037 [0.007]	-0.007 [0.014]	-0.151 [0.009]
Personal Care Routines: Dressing	0.061 [0.011]	0.039 [0.013]	-0.019 [0.011]	-0.051 [0.011]	-0.073 [0.022]	0.125 [0.013]
Active Physical Play	-0.023 [0.007]	-0.047 [0.009]	-0.120 [0.007]	-0.087 [0.007]	-0.008 [0.014]	-0.141 [0.008]
Nutrition	0.075 [0.008]	0.038 [0.010]	0.048 [0.008]	0.005 [0.008]	-0.183 [0.016]	0.147 [0.009]
Knowledge of Wellness	-0.041 [0.027]	-0.052 [0.032]	-0.067 [0.026]	-0.047 [0.027]	-0.132 [0.052]	-0.135 [0.030]

Note. Standard errors (SE) are included in brackets following the estimate.

References

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing.* AERA.

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5 (H.R. 1), February 17, 2009; 123 Stat. 115 as of May 21, 2012.

Anastasi, A. (1988). *Psychological testing* (6th ed.). Macmillan.

Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade.* Mathematica Policy Research. https://www.pewtrusts.org/en/research-and-analysis/reports/2007/07/03/measuring-childrens-progress-from-preschool-through-third-grade

Bagnato, S. J., & Neisworth, J. T. (1994). A national study of the social and treatment "invalidity" of intelligence testing for early intervention. *School Psychology Quarterly, 9*(2), 81–102.

Bagnato, S. J., Neisworth, J. T., & Pretti-Frontczak, K. (2010). *LINKing authentic assessment and early childhood intervention: Best measures for best practice* (2nd ed.). Paul H Brookes Publishing.

Barrueco, S., Lopez, M., Ong, C., & Lozano, P. (2012). Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures. Brookes Publishing Company.

California Department of Education. (2000). History—social science content standards for California public schools: Kindergarten through grade twelve. CDE Press.

California Department of Education. (2001). Visual and performing arts content standards for California public schools: Prekindergarten through grade twelve—Dance, music, theatre, visual Arts. CDE Press.

California Department of Education. (2006). Physical education model content standards for California public schools: Kindergarten through grade twelve. CDE Press.

California Department of Education. (2008). *California preschool learning foundations: Vol. 1.* CDE Press.

California Department of Education. (2009a). California infant/toddler learning & development foundations. CDE Press.

California Department of Education. (2009b). Health education content standards for California public schools: Kindergarten through grade twelve. CDE Press.

California Department of Education. (2010). *California preschool learning foundations: Vol. 2.* CDE Press.

California Department of Education. (2012). *California preschool learning foundations: Vol. 3.* CDE Press.

California Department of Education. (2013). California Common Core State Standards: English language arts & literacy in history/social studies, science, and technical subjects. CDE Press.

California Department of Education. (2014a). *California Common Core State Standards: Mathematics.* CDE Press. https://www.cde.ca.gov/be/st/ss/documents/ccssmathstandardaug2013.pdf

California Department of Education. (2014b). *California English language development standards: Kindergarten through grade 12.* CDE Press. https://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf

California Department of Education. (2015a). Desired Results Developmental Profile (2015): An early childhood developmental continuum. CDE Press.

California Department of Education. (2015b). Desired Results Developmental Profile—Kindergarten (2015): For use with transitional kindergarten and kindergarten-aged children. CDE Press.

California Department of Education. (2015c). *Next Generation Science Standards for California public schools, kindergarten through grade twelve.* https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp

Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., Gamez-Patience, B., & Jones, M. (2015). The predictive validity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review, 44*(1), 76–97.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). Harper & Row.

DRDP Collaborative Research Group. (2018). *Technical report for the Desired Results Developmental Profile (2015)*. Report prepared for the California Department of Education.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates, Inc.

Feuerstahler, L. M., & Wilson, M. (2019). Scale alignment in between-item multidimensional Rasch models. *Journal of Educational Measurement, 56*, 280–301.

Feuerstahler, L. M., & Wilson, M. (2021). Scale alignment in the between-item multidimensional partial credit model. *Applied Psychological Measurement*, *45*(4), 268–282. https://doi.org/10.1177/01466216211013103

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Harvard University Press.

Greenspan, S. I., & Wieder, S. (1998). The child with special needs: Encouraging intellectual and emotional growth. Perseus Books.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory: Vol. 2.* Sage Publications.

Institute of Medicine (IOM) & National Research Council (NRC). (2015). *Transforming the workforce for children birth through age 8: A unifying foundation.* The National Academies Press.

Kriener-Althen, K., Newton, E. K., Draney, K., & Mangione, P. (2020). Measuring readiness for kindergarten using the Desired Results Developmental Profile. *Early Education and Development 31*(5), 739–763. https://doi.org/10.1080/10409289.2020.1743160

Mangione, P. L., Osborne, T., & Mendenhall, H. (2019). What's next? How learning progressions help teachers support children's development and learning. *Young Children, 74*(3), 20–25.

Meisels, S. J. (1994). Designing meaningful measurements for early childhood. In B. L. Mallory & R. S. New (Eds.), *Diversity in early childhood education: A call for more inclusive theory, practice, and policy* (pp. 202–222). Teachers College Press.

Meisels, S. J. (1996). Charting the continuum of assessment and intervention. In S. J. Meisels & E. Fenichel (Eds.), *New visions for the developmental assessment of infants and young children* (pp. 27–52). Zero to Three/National Center for Infants, Toddlers, and Families.

Meisels, S. J., & Provence, S. (1989). Screening and assessment: Guidelines for identifying young disabled and developmentally vulnerable children and their families. National Center for Clinical Infant Programs.

National Research Council. (2008). *Early childhood assessment: Why, what, and how.* The National Academies Press.

Organisation for Economic Cooperation and Development (OECD). (2017). *PISA 2015 technical report.* https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf

Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel–Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement, 71*, 1023–1046.

Poggio, J., & McJunkin, L. (2012). History, current practice, perspectives and what the future holds for computer based assessment in K-12 education. Information Age.

Schwartz, R., & Ayers, E. (2011). Delta dimensional alignment: Comparing performances across dimensions of the learning progression for Assessing Data Modeling and Statistical Reasoning [Unpublished manuscript]. University of California, Berkeley.

Shepard, L. A., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments.* The National Education Goals Panel.

Stipek, D., & Colleagues. (2019). *Getting down to facts II – Technical report: Early childhood education in California.* Policy Analysis for California Education, Stanford University. https://gettingdowntofacts.com/technical-reports

Sussman, J., & Gochyyev, P. (2019, April). Multidimensional IRT analysis to support formative, longitudinal assessment of early childhood development. In K. Draney (Chair), *Supporting children's early learning and kindergarten goals with a standards-aligned assessment: Effective practices in statewide aggregate reporting* [Paper presentation]. Annual meeting of the American Educational Research Association, Toronto, Ontario, Canada. https://bearcenter.berkeley.edu/bibliography/multidimensional-irt-analysis-support-formative-longitudinal-assessment-early-childhood

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, *54*, 427–450.

WestEd. (2018a). *DRDP (2015) aligned to the California infant/toddler and preschool foundations.* Report prepared for the California Department of Education. https://www.desiredresults.us/drdp-2015-aligned-california-foundations

WestEd. (2018b). *DRDP (2015) research summaries.* Report prepared for the California Department of Education. https://www.desiredresults.us/research-summaries-drdp-2015-domain

WestEd. (2018c). *DRDP-K aligned to the California learning standards.* Report prepared for the California Department of Education. https://drdpk.org/cor_sum.html

WestEd. (2018d). *DRDP-K (2015) research summaries.* Report prepared for the California Department of Education and the Illinois State Board of Education. https://drdpk.org/res_sum.html

WestEd & University of California – Berkeley Evaluation and Assessment Research Center (UC BEAR). (2021a). *Desired results developmental profile (2015) early infancy through kindergarten continuum: Interrater reliability.* Report prepared for the California Department of Education.

WestEd & University of California – Berkeley Evaluation and Assessment Research Center (UC BEAR). (2021b). Desired results developmental profile (2015) early infancy through kindergarten

continuum: Validity evidence in relation to external measures of child development. Report prepared for the California Department of Education.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Psychology Press.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). ConQuest Version 2: Generalised Item Response Modelling Software. ACER Press.