



# DRDP (2015) 2017-2018 Differential Item Functioning (DIF) Analyses Report

## Introduction

---

The DRDP (2015), developed by the California Department of Education (CDE), is a judgment-based, authentic assessment instrument. Observation-based assessments such as the DRDP (2015) are completed by assessors (e.g., teachers, special education service providers) who interact regularly with the children being assessed. Assessors use observations and other documentation to inform their ratings of developmental and learning continua (measures) organized under eight domains:

1. Approaches to Learning—Self-Regulation (ATL-REG),
2. Social and Emotional Development (SED),
3. Language and Literacy Development (LLD),
4. English Language Development (ELD),
5. Cognition, including Math and Science (COG),
6. Physical Development—Health (PD-HLTH),
7. History-Social Science (HSS), and
8. Visual and Performing Arts (VPA).

The instrument is appropriate for use with children from birth to kindergarten entry (birth through 5 yrs of age) and is required for use with children participating in early childhood settings funded through two CDE divisions: the Early Education and Support Division (EESD) and the Special Education Division (SED).

The domain-specific content on the DRDP (2015) is based on developmental research and constructs specified in the California Infant/Toddler Learning and Development Foundations and Preschool Learning Foundations (California Department of Education, 2008, 2010, 2012) as well as the California Preschool Curriculum Framework Volumes 1-3 (California Department of Education, 2010, 2011, 2013). DRDP (2015) content is aligned to and used for reporting related to the OSEP Child Outcomes required by the U.S. Department of Education, Office of Special Education Programs (OSEP) (U.S. Department of Education, 2005) and the Head Start Early Learning Outcomes Framework (HSELOF) required by the U.S. Department of Health and Human Services, Office of Head Start (U.S. Department of Health and Human Services, 2015). The content of the DRDP (2015) reflects the knowledge, skills, or behaviors important for infants, toddlers, and preschool children to learn (California Department of Education, 2015).

For the present DIF analyses, all data were collected using the calibration version of the DRDP (2015). The calibration version of the DRDP (2015) had two views: an Infant/Toddler View and a Preschool View. The Infant/Toddler View was comprised of 27 measures and the Preschool View was comprised of 29 additional measures, for a total of 56 measures. Measures contained in the two views of the calibration version of the DRDP (2015) are virtually identical to the measures contained in the Infant/Toddler View and the Preschool Comprehensive View of the of the DRDP (2015) instrument currently in use in California<sup>1</sup>.

---

<sup>1</sup> Note: In the fall of 2016 an additional Preschool View of the DRDP (2015) was introduced, the Preschool Fundamental View. The Fundamental View is comprised of a subset of 43 measures from the Preschool Comprehensive View and focuses on domains of school readiness. The Preschool Fundamental View is currently used in some child development programs and nearly all special education programs as of fall 2016.

## Context for the 2017-2018 DIF Analyses

At the onset of the development of the DRDP (2015), agencies contracted by the California Department of Education outlined a series of assessment specifications to establish the objectives of the instrument. Adherence to the *2014 Standards for Educational and Psychological Testing* developed by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) guided these specifications (AERA/APA/NCME, 2014).

One assessment specification focused on universal design principles for the DRDP (2015). Among the evidence for universal design is the absence of bias. Detecting and reducing the presence of measurement bias is desired in educational and psychological measurement contexts, particularly for judgment-based, authentic assessments such as the DRDP (2015), which relies on observations to inform performance ratings. Specifically, the current analyses address Standard 3.6 of the Standards for Educational and Psychological Testing:

---

Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (AERA/APA/NCME, 2014).

---

**Figure 1:** Standard 3.6 of the Standards for Educational and Psychological Testing

One approach for examining measurement bias is to explore measurement invariance. As Millsap (2007, p. 462) noted, “at its root, the notion of measurement invariance is that some properties of a measure should be independent of the characteristics of the person being measured, apart from those characteristics that are the intended focus of the measure.” Differential item functioning (DIF) analyses are a statistical approach often used to explore measurement invariance.

DIF is used to investigate whether measures on an assessment instrument function differently for distinct groups of children. That is, the measures are not invariant. Examination of DIF is important when developing instruments, such as the DRDP (2015), to determine whether various subgroups of children who possess comparable ability levels have equal (or different) likelihood of receiving the same ratings on the measures. Subgroups of children could include those of different age, gender, or type of disability, among other attributes.

Measures that exhibit DIF, and that are not invariant across subgroups, work “one way for one group of respondents and in a different way for another group” (de Ayala, 2009, p. 323). The amount of DIF associated with a measure impacts the assessor’s ability to make accurate and meaningful comparisons of performance between children across different subgroups. It is important to demonstrate that measures on instruments have minimal DIF or measurement bias, particularly for instruments such as the DRDP (2015) that may be used to make interpretations concerning a child’s performance and comparisons across subgroups.

The focus of the DIF analyses described in the present report was to provide evidence that the DRDP (2015) generally functions as intended for all children of the same ability level (i.e., no bias in a specific developmental area or skill). These analyses examined the extent to which children ages birth to five with similar ability levels but representing distinct subgroups received the same measure ratings on the DRDP (2015). The two subgroups examined in these analyses are children with disabilities (here defined as infants and toddlers with Individualized Family Service Plans (IFSPs) and preschool children with Individualized Education Programs (IEPs)) and children without disabilities. The following research question guided the analyses:

To what degree does DIF exist on any measure of the DRDP (2015) for children with disabilities in SED-funded programs versus children in EESD-funded programs, who do not have disabilities?

# Methodology

## Sample

All data for this investigation were drawn from a calibration study of the DRDP (2015) conducted in the Spring of 2015. All participants in the study were early interventionists, infant care and preschool teachers, or early childhood special education service providers selected by program administrators from EESD and SED-funded programs who responded to a request for study participants. Selected assessors participated in a DRDP (2015) online training session prior to conducting the assessment. Participants were currently working with the children in an early intervention setting (e.g., the child's home), preschool classroom, or other early childhood setting. Each study participant assessed one or more children with the DRDP (2015).

Children in the EESD group included infants and toddlers and preschool age-aged children enrolled in early care and education programs administered by the EESD. Children in the SED group included infants and toddlers and preschool-aged children receiving special education services and programs. For purposes of the calibration study, no children included in the EESD group had an IFSP or IEP. For the current DIF analyses, the EESD group is referred to as *children without disabilities* and the SED group is referred to as *children with disabilities*.

Tables 1-4 provide a summary of the demographic information of the child participants in the 2015 calibration study from which data were used for the DIF analyses. Only assessment records with complete assessment ratings across all measures within a domain were included for the present analyses (totaling 19,128 records across both samples).

Table 1: Demographic Information – Gender

Gender	Children without Disabilities [EESD Sample]		Children with Disabilities [SED Sample]	
	n	%	n	%
Male	8,678	49.1%	851	65.7%
Female	8,678	49.1%	851	65.7%
Total	17,675*	100%	1,295	100%

\*158 missing gender information

Table 2: Demographic Information – Age

Age in Yrs	Children without Disabilities [EESD Sample]		Children with Disabilities [SED Sample]	
	n	%	n	%
< 1	214	1.2%	31	2.4%
1	594	3.4%	117	9.0%
2	1,043	5.9%	217	16.8%
3	2,812	16.0%	265	20.5%
4	7,604	43.2%	447	34.5%
5	5,331	30.3%	218	16.8%
Total	17,598*	100%	1,295	100%

\*235 missing age information

**Table 3: Demographic Information – Children in Infant/Toddler and Preschool Settings**

Age Range	Children without Disabilities [EESD Sample]		Children with Disabilities [SED Sample]	
	n	%	n	%
Infant/Toddler	2,273	12.7%	378	29.2%
Preschool	15,560	87.3%	917	70.8%
Total	17,833	100%	1,295	100%

**Table 4: Disability Categories of Children with IFSPs and IEPs**

Disability Category	Children with Disabilities [SED Sample]	
	n	%
Autism	392	30.3%
Speech or Language Impairment	314	24.2%
Intellectual Disability	112	8.6%
Hard of Hearing	101	7.8%
Other Health Impairment	96	7.4%
Orthopedic Impairment	77	5.9%
Deafness	77	5.9%
Multiple Disability	45	3.5%
Visual Impairment	29	2.2%
Specific Learning Disability	22	1.7%
Established Medical Disability	20	1.5%
Total	1,295	100.0%*

\* Total may not add to 100% due to rounding.

Note: For the purposes of the calibration study no children included in the EESD sample had IFSPs or IEPs and therefore, did not have a disability included in this list.

## Instrument

The DRDP (2015) is comprised of 56 items (measures) across two views that are contained within one of eight groupings of measures referred to as developmental domains. The developmental domains, the affiliated domain abbreviation and the number of measures assigned to each domain are shown in Table 5.

**Table 5: Developmental Domains of the DRDP (2015)**

Developmental Domain	Abbreviation	No. Measures
1. Approaches to Learning -Self-Regulation	ATL-REG	6
2. Social-Emotional Development	SED	5
3. Language and Literacy Development	LLD	10
4. English Language Development	ELD	4
5. Cognition, including Math and Science	COG	12
6. Physical Development and Health	PD-HLTH	10
7. History-Social Science	HSS	5
8. Visual and Performing Arts	VPA	4
	Total	56

The calibration version of the DRDP (2015) used for preschool children was comprised of all eight domains while the Infant/Toddler View included five of the eight domains (ATL-REG, SED, LLD, COG, and PD-HLTH).

The data collected for the calibration study were comprised of DRDP (2015) measure ratings assigned by the children’s early care and education teachers and service providers (assessors). Assessors observed children over time in everyday routines and activities and assigned a judgment-based rating of mastery to each measure. An assessor considered a developmental level mastered if the child demonstrated the knowledge, skills, and behaviors defined at that level consistently over time and in different situations or settings. Data were collected in the spring of 2015.

As shown in Table 6, the developmental sequences that comprise the measures on the DRDP (2015) are presented as an ordinal scale, and the number of developmental levels within each measure varies from five to nine, depending on the nature of the developmental sequence for that measure. Ratings are assigned to one of the developmental levels listed below<sup>2</sup>.

**Table 6:** Developmental Rating Levels of the DRDP (2015)

Responding		Exploring			Building			Integrating
Earlier	Later	Earlier	Middle	Later	Earlier	Middle	Later	Earlier

For the purposes of these DIF analyses, the four measures contained in the domain of English Language Development (ELD) were not included. Overall, data representing a total of 52 measures comprising seven of the domains contained on the DRDP (2015) were retained for the analyses.

## Calibration Model of the DRDP (2015)

The DRDP (2015) utilizes Item Response Theory (IRT) modeling, specifically, a Rasch measurement model was used to develop the scaled scores that are assigned based on performance on groups of measures within a domain. The multidimensional structure of the DRDP (2015) applies a multidimensional random coefficients multinomial logit model (MRCML) proposed by Adams, Wilson, and Wang (1997). This one-parameter, item response theory (IRT) approach (i.e., Rasch) integrates the partial credit model (Masters, 1982) and is applied when multiple dimensions are present within a single overarching construct. Under the partial credit model, each measure has a unique rating scale structure that takes into consideration levels assigned on other measures within the domain. The domain-level ratings are converted from ordinal-level values into interval-level values (provided in logits).

For calibration, a marginal maximum likelihood estimation with a Monte Carlo sampling technique for the multiple dimensions was used. Parameter estimates for the measurement model were obtained using the ConQuest 4.5 modeling software (Adams, Wu, and Wilson, 2015), and the expected-a-posteriori (EAP) score estimation method was used to estimate children’s developmental domain scores<sup>3</sup>.

## DIF Analysis Analytic Model

Measure level data derived from the analytic processes were used to estimate children’s scaled scores on the DRDP (2015) and these scores were used to perform all DIF analyses.

The ConQuest software provided the analysis model to understand the performance differences between groups (i.e., children without disabilities versus children with disabilities) at the measure level. Performance differences at the measure level are described here as differential item functioning (DIF).

<sup>2</sup> For more information about the measures and domains contained on the DRDP (2015), refer to Desired Results Developmental Profile (2015): A Developmental Continuum from Early Infancy to Kindergarten Entry (CDE, 2015).

<sup>3</sup> For additional information related to the measurement model and multidimensional domain structure used for the calibration of instrument, see the Desired Results Developmental Profile (2015): Technical Report. Sacramento (CDE, 2018).

To process the model, the ConQuest software identifies all possible combinations of the  $m$  measures and  $d$  division variables and constructs  $m \times d$  generalized items. The model statement requests that ConQuest describe the probability of correct responses to these generalized items using a measure main effect, a division main effect, and an interaction between measure and division.

One of the key ways in which DIF is studied is through the use of the Mantel-Haenszel (MH) DIF statistic,  $\hat{D}_i$ , (Holland & Thayer, 1988). The Educational Testing Service (ETS) provides a set of classification rules (Dorans & Holland, 1993) used to evaluate the degree of DIF. However, the DRDP (2015) was constructed under a model grounded in Item Response Theory (Rasch modeling) and MH procedures are most suitable for models developed under classical test theory. In the context of Rasch (1960) modeling, a DRDP (2015) measure would be deemed to exhibit DIF if the response probabilities for that measure cannot be fully explained by the ability of the child and a fixed set of difficulty parameters for that measure (Jin et al., 2017).

Paek and Wilson (2011) present a modified set of classification rules that take into consideration the marginal maximum likelihood estimation context of the Rasch-based modeling approach. The Rasch-based classification rules are based on the item difficulty difference,  $\gamma$ , between the focal group (i.e., children with disabilities) and the reference group (i.e., children without disabilities), which is reflected in the formula  $\gamma = \delta_F - \delta_R$  and described below in Table 7 (for additional discussion of the modified classification rules see Paek and Wilson).

**Table 7:** DIF classifications, rules, and recommended actions

ETS Classification	Rule	Action
A: Trivial DIF	If $ \gamma  \leq 0.426$ or if $H_0: \gamma = 0$ is not rejected below .05 level	None
B: Non-trivial DIF	If $0.426 \leq  \gamma  \leq 0.638$ and if $H_0: \gamma = 0$ is rejected below .05 level	Investigate
C: Large DIF	If $ \gamma  \geq 0.638$ or if $H_0: \gamma = 0$ is rejected below .05 level	Remove

The above classification rules were applied to the measure-level differences between the two groups of children: children with disabilities and children without disabilities. Trivial DIF was defined as being less than or equal to .426. Non-trivial DIF was defined as being less than or equal to .638 but greater than .426. Large DIF was defined as being greater than .638.

## Results

In the current study, a total of 52 measures across seven developmental domains contained on the DRDP (2015)<sup>4</sup> were examined. As shown in Table 9 all items under examination had DIF values below the threshold denoting a non-trivial level of DIF of 0.426 and were given the Group A classification for items exhibiting trivial DIF. No individual measure was shown to have a DIF value exceeding 0.25. Only 3 of 56 measures had a DIF value exceeding 0.20. These measures were COG 1: Spatial Relationships, PD 2: Gross Locomotor Movement Skills, and VPS 2: Music. The table also indicates the directionality of DIF. That is, whether the measure favors the children with disabilities (denoted by an N) and a positive Difference ( $\gamma = \delta_F - \delta_R$ ) or favors children without disabilities (denoted by an ND) and a negative Difference. DIF associated with the three measures with DIF exceeding .20 indicated a slight tendency toward more favorable (later developing) ratings for children without disabilities than for children with disabilities, albeit at a trivial level of DIF. More than half of the DRDP (2015) measures examined (23 of 52) exhibited DIF levels of 0.10 or lower. Additionally, no group of measures across an entire domain (e.g.: ATL-REG or PD) showed a systematic pattern of DIF favoring either children with or without disabilities. Table 8 shows the results of the DIF analyses and the application of the classification rules for all

<sup>4</sup> The four measures contained in the ELD domain were not included in these analyses as this domain does not follow the same developmental progression and as such requires a different analytic approach than employed in these analyses. Overall, data representing a total of 52 measures comprising seven of the eight domains contained on the DRDP (2015) were retained for these analyses.

measures.

Table 8: DRDP (2015) DIF Analyses Findings

Measure*	Measure Name	$\delta_R$ (ND)	$\delta_F$ (D)	Difference, $\gamma = \delta_F - \delta_R$	Standard Error	Favors**	Interpre- tation
ATL-REG 1	Attention Maintenance	0.049	-0.049	-0.10	0.01	ND	Trivial
ATL-REG 2	Self-Comforting	-0.006	0.006	0.01	0.02	D	Trivial
ATL-REG 3	Curiosity and Initiative in Learning	0.018	-0.018	-0.04	0.02	ND	Trivial
ATL-REG 4	Self-Control of Feelings and Behavior	0.044	-0.044	-0.09	0.01	ND	Trivial
ATL-REG 5	Engagement and Persistence	0.014	-0.014	-0.03	0.01	ND	Trivial
ATL-REG 6	Shared Use of Space and Materials	0.038	-0.038	-0.08	0.01	ND	Trivial
ATL-REG 7	Imitation	0.025	-0.025	-0.05	0.01	ND	Trivial
SED 1	Identity of Self in Relation to Others	-0.021	0.021	0.04	0.01	D	Trivial
SED 2	Social and Emotional Understanding	-0.042	0.042	0.08	0.01	D	Trivial
SED 3	Relationships and Social Interactions with Familiar Adults	0.005	-0.005	-0.01	0.01	ND	Trivial
SED 4	Relationships and Social Interactions with Peers	-0.053	0.053	0.11	0.01	D	Trivial
SED 5	Symbolic and Sociodramatic Play	-0.081	0.081	0.16	0.01	D	Trivial
LLD 1	Understanding of Language (Receptive)	-0.022	0.022	0.04	0.01	D	Trivial
LLD 2	Responsiveness to Language	0.009	-0.009	-0.02	0.01	ND	Trivial
LLD 3	Communication and Use of Language (Expressive)	-0.048	0.048	0.10	0.01	D	Trivial
LLD 4	Reciprocal Communication and Conversation	-0.095	0.095	0.19	0.01	D	Trivial
LLD 5	Interest in Literacy	-0.022	0.022	0.04	0.01	D	Trivial
LLD 6	Comprehension of Age-Appropriate Text	-0.051	0.051	0.10	0.01	D	Trivial
LLD 7	Concepts About Print	0.022	-0.022	-0.04	0.01	ND	Trivial
LLD 8	Phonological Awareness	-0.011	0.011	0.02	0.01	D	Trivial
LLD 9	Letter and Word Knowledge	0.087	-0.087	-0.17	0.01	ND	Trivial
LLD 10	Emergent Writing	-0.023	0.023	0.05	0.01	D	Trivial
COG 1	Spatial Relationships	0.126	-0.126	-0.25	0.02	ND	Trivial
COG 3	Cause and Effect	0.063	-0.063	-0.13	0.01	ND	Trivial
COG 4	Classification	0.03	-0.03	-0.06	0.01	ND	Trivial
COG 5	Number Sense of Quantity	-0.015	0.015	0.03	0.01	D	Trivial
COG 6	Number Sense of Math Operations	-0.051	0.051	0.10	0.01	D	Trivial
COG 7	Measurement	0.021	-0.021	-0.04	0.01	ND	Trivial
COG 8	Patterning	0.058	-0.058	-0.12	0.01	ND	Trivial
COG 9	Shapes	0.073	-0.073	-0.15	0.01	ND	Trivial
COG 10	Inquiry Through Observation and Investigation	-0.009	0.009	0.02	0.01	D	Trivial
COG 11	Documentation and Communication of Inquiry	-0.072	0.072	0.14	0.01	D	Trivial
COG 12	Knowledge of the Natural World	-0.025	0.025	0.05	0.01	D	Trivial
PD 1	Perceptual-Motor Skills and Movement Concepts	0.034	-0.034	-0.07	0.01	ND	Trivial
PD 2	Gross Locomotor Movement Skills	0.122	-0.122	-0.24	0.01	ND	Trivial
PD 3	Gross Motor Manipulative Skills	0.017	-0.017	-0.03	0.01	ND	Trivial
PD 4	Fine Motor Manipulative Skills	-0.024	0.024	0.05	0.01	D	Trivial
HLTH 1	Safety	-0.031	0.031	0.06	0.01	D	Trivial
HLTH 2	Personal Care Routines: Hygiene	-0.015	0.015	0.03	0.01	D	Trivial
HLTH 3	Personal Care Routines: Self-Feeding	-0.026	0.026	0.05	0.01	D	Trivial

Measure*	Measure Name	$\delta_R$ (ND)	$\delta_F$ (D)	Difference, $\gamma = \delta_F - \delta_R$	Standard Error	Favors**	Interpretation
HLTH 4	Personal Care Routines: Dressing	-0.024	0.024	0.05	0.01	D	Trivial
HLTH 5	Active Physical Play	0.025	-0.025	-0.05	0.01	ND	Trivial
HLTH 6	Nutrition	-0.054	0.054	0.11	0.01	D	Trivial
HSS 1	Sense of Time	-0.059	0.059	0.12	0.01	D	Trivial
HSS 2	Sense of Place	0.052	-0.052	-0.10	0.01	D	Trivial
HSS 3	Ecology	-0.06	0.06	0.12	0.01	D	Trivial
HSS 4	Conflict Negotiation	0.003	-0.003	-0.01	0.01	ND	Trivial
HSS 5	Responsible Conduct as a Group Member	0.068	-0.068	-0.14	0.01	ND	Trivial
VPA 1	Visual Art	-0.062	0.062	0.12	0.01	D	Trivial
VPA 2	Music	0.106	-0.106	-0.21	0.01	ND	Trivial
VPA 3	Drama	-0.091	0.091	0.18	0.01	D	Trivial
VPA 4	Dance	0.054	-0.054	-0.11	0.01	ND	Trivial

\* DRDP measure numbers reflect the order of measures as they appear in the current DRDP (2015), rather than the measure sequence that was used during the calibration study.

\*\* ND = Children without disabilities (EESD Sample); D= Children with disabilities (SED Sample)

## Review of Measures with Larger DIF Values

While only trivial amounts of DIF were detected in these analyses, it is instructive to review some of the measures that exhibited a larger amount of DIF than other items. The table below presents a summary of measures with larger values of DIF ( $|\gamma| > 0.14$ ). The measures varied in terms of which group an item favored with four of nine measures favoring children without disabilities and the remaining five measures favoring children with disabilities. Three domains had more than one measure with larger DIF values represented: COG (three measures), LLD (two measures), and VPA (two measures). However, the observed DIF across the measures within a domain did not all favor one group. It will be useful to monitor these measures over time to determine if DIF levels remain consistent. Table 9 lists the nine measures that exhibited the largest amount of DIF in these analyses.

Table 9: DRDP (2015) Measures with larger DIF

Measure	Measure Name	Difference, $\gamma = \delta_F - \delta_R$	Standard Error	Favors*	Interpretation
SED 5	Symbolic and Sociodramatic Play	0.16	0.01	D	Trivial
LLD 4	Reciprocal Communication and Conversation	0.19	0.01	D	Trivial
LLD 9	Letter and Word Knowledge	-0.17	0.01	ND	Trivial
COG 1	Spatial Relationships	-0.25	0.02	ND	Trivial
COG 9	Shapes	-0.15	0.01	ND	Trivial
COG 11	Documentation and Communication of Inquiry	0.14	0.01	D	Trivial
PD 2	Gross Locomotor Movement Skills	-0.24	0.01	ND	Trivial
VPA 2	Music	-0.21	0.01	ND	Trivial
VPA 3	Drama	0.18	0.01	D	Trivial

\*ND = Children without disabilities (EESD Sample); D= Children with disabilities (SED Sample)

## Discussion

The DIF analyses performed in the present study sought to examine to what degree DIF existed on any measures of the



DRDP (2015) for children in CDE-funded programs when comparing assessment results for children with disabilities (i.e., with IFSPs and IEPs) and children without disabilities. Overall, no measures contained on the DRDP (2015) were shown to have a DIF value exceeding the established threshold denoting a non-trivial level of DIF. Per the classification rules, no further action is required to address potential bias of any measures contained on the DRDP (2015) for these two subgroups. No observable trends were observed in the direction of findings that suggested items tended to be rated in a manner that consistently favored the level of performance of either group (even at the trivial DIF level).

## Limitations

All data for the present analyses utilized data collected using the calibration version of the DRDP (2015) during the spring 2015 calibration study. The use of a calibration version of the DRDP (2015) and reliance on a study sample of participants presents potential limitations to interpretations drawn from these analyses. At the time of calibration study, the instrument had not been formally deployed into the field for use. As a result, many assessors were likely unfamiliar with this version of the DRDP instrument. Additionally, the number of children with disabilities assessed during the calibration study was limited to a sample of approximately 1,500 children. Future DIF analyses should be based on more current assessment results that would reflect teachers' increased familiarity and experience with the instrument and use larger samples of children with disabilities (i.e.: IFSPs and IEPs) participating in CDE-funded programs. Accessing data from the state-wide administrations of the DRDP (2015) could produce a larger sample size, and allow a deeper exploration of DIF across additional disaggregated groups of children.

## Future Studies

DIF analyses of the DRDP (2015) focusing on additional subgroups of children will be conducted. Analyses that include assessment data from children with specific identified disabilities such as autism spectrum disorders and speech and language delays have been planned. DIF analyses could also be performed on a larger sample that would include DRDP (2015) assessment results gathered from all children with IFSPs and IEPs participating in CDE-funded programs over at least a three-year period. This larger sample would allow for examination of assessment data from groups of children with less frequently occurring disabilities, including children with low incidence disabilities such as children who are deaf, hard of hearing, or with visual or orthopedic impairments.

## Conclusions

The purpose of this study was to examine measurement bias, the degree to which DIF existed on measures of the DRDP (2015) for children in CDE-funded programs when comparing assessment results for children with disabilities (i.e., with IFSPs and IEPs) and children without disabilities. For the purposes of the present analyses, levels of DIF identified across measures of the DRDP (2015) were found to be of a trivial level when comparing children with disabilities and those without. No particular domains showed any pattern of elevated DIF. The results of these DIF analyses suggest that the DRDP (2015) does function as a universal measure for all children, including those with and without disabilities, and shows little evidence of bias for these two subgroups in any of the measures present on the DRDP (2015).

## References

---

- Adams, R. J., Wilson, M., & Wang, W. C. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21, 1 -23.
- Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). ACER ConQuest: Generalized Item Response Modelling Software [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.
- American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.
- California Department of Education (CDE). (2008). *California Preschool Learning Foundations, Volume 1*. Sacramento, CA: CDE Press.

- California Department of Education (CDE). (2009). California Infant/Toddler Learning and Development Foundations. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2010a). California Preschool Curriculum Framework, Volume 1. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2010b). California Preschool Learning Foundations, Volume 2. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2011). California Preschool Curriculum Framework, Volume 2. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2012). California Preschool Learning Foundations, Volume 3. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2013). California Preschool Curriculum Framework, Volume 3. Sacramento, CA: CDE Press.
- California Department of Education (CDE). (2015). Desired Results Developmental Profile (2015) [DRDP (2015)]: An Early Childhood Developmental Continuum. Sacramento, CA: CDE.
- California Department of Education (CDE). (2018). Desired Results Developmental Profile (2015): Technical Report. Sacramento, CA: CDE.
- de Ayala, R. J. (2009). The theory and practice of item response theory. New York, NY: Guilford
- Desired Results Access Project. (2015). 2014-2015 *Interrater Agreement Study Report*. Rohnert Park, CA: Napa County Office of Education. Retrieved from <http://draccess.org/sites/default/files/pdfs/DRDP2015InterRaterStudyReport.pdf>
- Dorans, N., & Holland, P. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, N.J.: Erlbaum.
- Holland, P., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, N.J.: Erlbaum.
- Jin, H., Shin, H. J., Hokayem, H., Qureshi, F., & Jenkins, T. (2017). Secondary Students' Understanding of Ecosystems: a Learning Progression Approach. *International Journal of Science and Mathematics Education*, 1-19 [online]. Retrieved from: <https://doi.org/10.1007/s10763-017-9864-9>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika*, 72(4), 461-473.
- Paek, I. and 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(6), 1023-1046.
- U.S. Department of Education, Office of Special Education Programs, (2005).
- Family and child outcomes for early intervention and early childhood special education. Retrieved from: [https://ecoutcomes.fpg.unc.edu/sites/ecoutcomes.fpg.unc.edu/files/resources/ECO\\_Outcomes\\_4-13-05.pdf](https://ecoutcomes.fpg.unc.edu/sites/ecoutcomes.fpg.unc.edu/files/resources/ECO_Outcomes_4-13-05.pdf)
- U.S. Department of Health and Human Services (USDHHS), Administration for Children & Families, Office of Head Start. (2015). *The Head Start Early Learning Outcomes Framework*. (Publication No. HHSP233201000415G). Retrieved from: <https://eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-early-learning-outcomes-framework>
- Wang, W. C., Yao, G., Tsai, Y. J., Wang, J. D., & Hsieh, C. L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research*, 15(4), 607-620.