

June 2022

# DRDP Technical Report for Early Infancy Through Kindergarten: Interrater Reliability

---

**Authors:**

Min Chen-Gardini,<sup>1</sup> Joshua Sussman,<sup>2</sup> Emily Newton,<sup>1</sup> G. Sam Ruiz Jimenez,<sup>1</sup> Kerry Kriener-Althen,<sup>1</sup> Perman Gochyyev,<sup>2</sup> Karen Draney,<sup>2</sup> Peter Mangione<sup>1</sup>

*This report and related research were supported by contracts with the California Department of Education (CN 20-0190) and the California Department of Social Services (CN 21-7013).*

---

<sup>1</sup> WestEd

<sup>2</sup> University of California – Berkeley Evaluation and Assessment Research Center (UC BEAR)

## Table of Contents

---

<b>The Desired Results Developmental Profile .....</b>	<b>3</b>
DRDP Views for Different Age Levels .....	3
DRDP Sets of Varying Length .....	4
<b>The Current Study .....</b>	<b>5</b>
<b>Method .....</b>	<b>6</b>
Sample and Measures.....	6
Study Procedures .....	7
Analysis Approach.....	8
<b>Results .....</b>	<b>9</b>
Interrater Reliability Evidence at the Domain/Sub-Domain Level .....	9
Interrater Reliability Evidence at the Measure Level.....	10
<b>Discussion and Limitations .....</b>	<b>11</b>
<b>Acknowledgments.....</b>	<b>13</b>
<b>Appendix A: Domain Level Interrater reliability – Infant/Toddler (IT) View.....</b>	<b>14</b>
<b>Appendix B: Domain Level Interrater reliability – Preschool (PS) View.....</b>	<b>16</b>
<b>Appendix C: Domain Level Interrater reliability – Kindergarten (K) View .....</b>	<b>19</b>
<b>Appendix D: Measure Level Interrater reliability – Infant/Toddler (IT) View .....</b>	<b>21</b>
<b>Appendix E: Measure Level Interrater reliability – Preschool (PS) View .....</b>	<b>24</b>
<b>Appendix F: Measure Level Interrater reliability – Kindergarten (K) View .....</b>	<b>29</b>
<b>References .....</b>	<b>35</b>

This report describes a study investigating the interrater reliability of the Desired Results Development Profile (DRDP) 2015 instrument. This interrater reliability study was used to establish the consistency of measurement of the DRDP when completed by two different raters, which is an important aspect of validity evidence for observational assessments (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; National Research Council, 2008).

## The Desired Results Developmental Profile

---

Since 2001, the California Department of Education (CDE) has collaborated with child development and assessment experts from WestEd and the University of California – Berkeley Evaluation and Assessment Research (UC BEAR) Center to provide the DRDP formative child assessment system to publicly funded early care and education programs throughout California. This collaboration has implemented three generations of DRDP instruments for early childhood programs funded by CDE.

The most recent generation of this instrument, the DRDP (2015)<sup>3</sup>, is for infants and toddlers, preschool-, and kindergarten-age children (Draney et al., 2022; DRDP Collaborative Research Group, 2018). The DRDP was developed through a collaboration between the CDE Early Learning Division (CDE EED) and the CDE 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 EED further extended the DRDP developmental continua for use in kindergarten, the Desired Results Developmental Profile—Kindergarten (DRDP-K) (CDE, 2015b). These instruments were created through iterative processes, grounded in child development research literature (WestEd, 2018a; WestEd 2018b), developed through consultation with nationally recognized child development experts, and refined through numerous qualitative and quantitative research studies with early childhood and kindergarten teachers. 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 and other special needs. For details on the development of the DRDP, including the considerations of cultural and linguistic appropriateness, sensitivity, and universal design, please see the *Technical Report for the Desired Results Developmental Profile (2015)* and Draney et al. (2022).

### DRDP VIEWS FOR DIFFERENT AGE LEVELS

The DRDP continuum is presented in three different instrument views, for use in (1) infant/toddler group care settings, (2) preschool classrooms, and (3) kindergarten classrooms. These are known as the DRDP infant/toddler (IT) view, the DRDP preschool (PS) view, and the DRDP kindergarten (K) view. The three views provide for developmentally appropriate assessment for children within each setting. The DRDP is completed by a child’s teacher (whether an infant/toddler care teacher, preschool teacher, or

---

<sup>3</sup> Hereinafter referred to as the DRDP, unless multiple generations of the instrument are being discussed.

kindergarten teacher) and assesses knowledge and skills based on ongoing documentation of teacher’s observations, reports from family members, and examples of children’s work.

The DRDP items, 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 comprehensive IT instrument view has 29 measures within 5 domains; (2) the comprehensive PS view has 56 measures across 10 domains or subdomains; and (3) the comprehensive K view has 55 measures across 11 domains or subdomains.

### DRDP SETS OF VARYING LENGTH

There are also different “sets” of DRDP measures that can be completed depending on the purpose of and time allocated for using the DRDP. The comprehensive set of the DRDP includes all DRDP measures, domains, and subdomains for a view (IT, PS, or K). In addition, three DRDP short forms with fewer domains and measures—referred to as the “fundamental” set, the “essential” set, and the “modified essential” set—are also available to programs and teachers who desire to use a shorter assessment. The measures in the modified essential set were specifically selected to be appropriate for virtual learning, though this set can also be used for in-person programs.

See Table 1 for the number of measures in each set for the three DRDP views. Table 2 contains the comprehensive set of DRDP domains, number of measures per domain per age-level instrument view, and sub-domains available in the PS and K views. Note that domains with asterisks (\*) are included in the fundamental and essential sets.

**Table 1. The number of measures in each set for each view of the DRDP.**

DRDP Sets	Infant/Toddler view	Preschool view	Kindergarten view
<b>Comprehensive</b>	29	56	55
<b>Fundamental</b>	N/A	37	37
<b>Essential</b>	23	29	33

**Table 2. The comprehensive set of 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 view	Preschool (PS) view	Kindergarten (K) view	Subdomains (PS and K views only)
<b>Approaches to Learning—Self-Regulation (ATL-REG)*</b>	5	7	4	N/A
<b>Social and Emotional Development (SED)*</b>	5	5	5	N/A

Domains	Infant/Toddler view	Preschool (PS) view	Kindergarten (K) view	Subdomains (PS and K views only)
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	

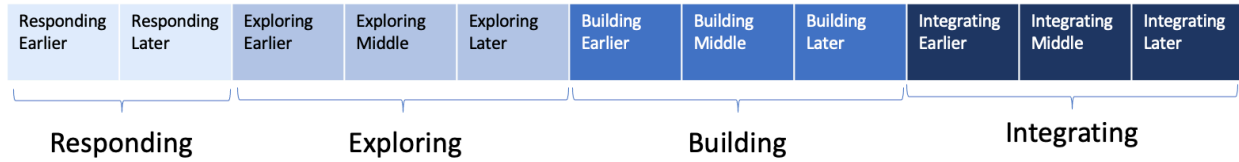
Note. Asterisks (\*) indicate domains included in the fundamental and essential sets. ELD and SPAN are “conditional” domains. ELD is completed only when a child in preschool or kindergarten has a home language other than English. SPAN is completed only when a child is enrolled in a kindergarten classroom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

## The Current Study

In the current study, the DRDP’s interrater reliability was tested at two different levels of data aggregation: (a) the domain and sub-domain level and (b) the measure level. For (a), interrater reliability was tested by comparing domain and sub-domain *developmental levels* between the two raters; domain and sub-domain developmental levels are the units of analysis reported to and used by teachers and families to support children’s learning and development. Domain and sub-domain developmental levels are computed based on the completion of ratings for the relevant measures for each domain/sub-domain. The 11 domain developmental levels were created by partitioning the measurement scale using a psychometric method in the empirical item mapping family of standard setting methods. The full set of domain developmental levels represent key steps in a birth through K continuum of learning and development; different DRDP views and sets reflect different parts of the developmental continuum, though they are all psychometrically linked into a continuous measurement scale. The interrater reliability analyses were applied to the DRDP domains and sub-domains in different views (IT, PS, and K) and different sets of measures (comprehensive, fundamental, essential, and modified essential).

For (b), interrater reliability was tested for each individual DRDP measure for each view. Interrater reliability at the measure level is the agreement in ratings between the two raters.

**Figure 1. DRDP’s Developmental Levels.**



## Method

### SAMPLE AND MEASURES

The sample for the study consisted of 401 infants/toddlers ( $M$  age = 21.9 months,  $SD$  = 9.2), 601 preschool-age children ( $M$  age = 51.3 months,  $SD$  = 7.4), and 99 kindergarten children ( $M$  age = 58.0 months,  $SD$  = 4.8), for a total of 1,101 children. The children were primarily enrolled in state-funded child care programs or public kindergarten classrooms throughout California between fall 2015 and spring 2016. The sample of selected programs and schools included children from throughout the state of California and who were representative of the state’s racial/ethnic diversity. (See Table 3 for children’s demographic characteristics). The teachers participating in the study represented 37 early childhood programs from across California and 11 elementary schools from one district in California’s Central Valley. In total, 152 teachers from infant/toddler settings, 204 teachers from preschool settings, and 51 teachers from kindergarten classrooms participated in the study. They completed the comprehensive set of the age-appropriate DRDP view (IT, PS, or K) for each participating child.

**Table 3. Demographic characteristics of children from infant/toddler, preschool, and kindergarten settings in the DRDP interrater reliability study. Percent of children by gender, dual language learner status, IEP status, and race/ethnicity.**

Demographic Category	All Children (N=1,101)	Infant/Toddler (N=401)	Preschool (N=601)	Kindergarten (N=99)
Male	48.1	47.6	48.1	50.5
Female	50.5	52.4	49.3	49.5
No Information Available About Gender	1.5	0	2.7	0
Dual Language Learners	46.2	n/a	47.6	37.4
IEP or IFSP	1.8	1.7	1.8	n/a
Hispanic or Latino/a/x	52.5	49.6	57.7	32.3

<b>White (Not Hispanic / Latino/a/x)</b>	19.0	21.7	14.4	36.4
<b>Black/African American</b>	4.7	5.2	4.3	5.1
<b>Asian or Pacific Islander</b>	5.9	5.7	5.8	7.1
<b>Native American</b>	1.8	1.5	2.0	2.0
<b>Multiple Race/Ethnicities</b>	5.3	8.5	3.2	5.1
<b>No Information Available About Race/Ethnicity</b>	10.8	7.7	12.6	12.1

*Note.* Dual language learner information is not collected with the infant/toddler view of the DRDP.

## STUDY PROCEDURES

The study sample was recruited from early childhood education programs and school districts throughout California that were using the DRDP instrument between fall 2016 and spring 2017. 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 and encouraged them to participate to ensure sufficient representation of various subpopulations of children in the study (e.g., young infants, kindergarten-age children, dual language learners). We used an active, opt-in consent process.

Rater pairs consisted of a primary rater (most often the teacher in a group care setting who regularly completed the DRDP for that child) and a secondary rater (other teachers, assistant teachers, or program administrators who had some familiarity with using the DRDP and who knew the child well or cared for the child a minimum of 10 hours per week). School/site administrators identified teachers who met the above criteria to take part in the study. WestEd staff and program administrators at participating sites worked together to identify secondary raters who fit the criteria outlined above. In infant/toddler settings, primary raters were lead teachers or program administrators, and secondary raters were typically teachers, assistant teachers, or program directors who also knew the child well. In preschool care settings, lead teachers typically acted as primary raters for children in their classroom while aides or other site personnel who typically assisted the lead teachers with completion of the DRDP acted as secondary raters. Lead teachers acted as secondary raters for children in each other’s infant/toddler or preschool classrooms in settings where aides or other site personnel were not familiar with the DRDP. In these scenarios, lead teachers would arrange at least 10 hours of observation time in each other's classrooms in order to complete the DRDP for the children in the study.

In kindergarten settings, rater pairs consisted of the lead classroom teachers as the primary raters and assistant classroom teachers<sup>4</sup> as the secondary raters. Secondary raters were asked to complete DRDP

---

<sup>4</sup> Unlike early childhood settings, most kindergarten settings have a single teacher who works with the same group of children in a classroom. For this study, we capitalized upon a unique situation where

assessments for at least three children in the interrater reliability study. All raters had received training in using the DRDP. However, in the kindergarten settings, the secondary raters were less experienced teachers and had prior experience assisting primary raters with the DRDP during its previous administration. We describe the implications of rater experience in the discussion section.

In total, 91 pairs of raters from infant/toddler settings, 93 pairs from preschool settings, and 50 pairs from kindergarten classrooms provided data for the study. Of the raters who participated in the study, 177 completed a follow-up survey of rater characteristics, training, and experience. Of these, 74 participated as a primary rater only, 56 participated as a secondary rater only, and 47 participated as a primary rater for some children and a secondary rater for other children. Most raters (87.5 percent) had completed ratings for at least one child with the DRDP prior to the study, with 80.7 percent having rated five or more children with the DRDP. Most of the raters (95.5 percent) reported that they and their rating partner had worked directly with the children they were rating for at least two months, and 58 percent reported that they had been working with the children for at least seven months. Though not all raters completed the survey, the evidence from those who did complete suggests that most raters had experience with the DRDP and knew well the children they were rating.

To complete the comprehensive view of the DRDP, rating pairs independently recorded documentation about the developmental competencies they observed children demonstrating over a six-week period during primary raters' usual planned DRDP assessment completion dates (after fall enrollment or during the winter or spring assessment periods). Following the six-week period of observation and documentation, teachers made rating determinations for each measure and entered the ratings into the online software system over a two-week period. Rater pairs were provided with instructions to complete the DRDP independently from each other. After the close of the data entry period in 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.

## ANALYSIS APPROACH

We assessed interrater reliability in these ordinal rating data using three metrics: percent absolute agreement, weighted percent agreement, and Cohen's weighted kappa ( $\kappa$ ; Cohen, 1968). First, percent absolute agreement is the percent of measures for which two raters agree. Percent agreement treats all disagreements equally, regardless of how close or far apart they are. Second, we used weighted percent agreement, which sums the percentage of ratings that are in absolute agreement and those that agree partially, with adjacent agreements partially contributing to the agreement statistic.<sup>5</sup> This metric

---

transitional kindergarten and kindergarten classrooms in one large district had hired a set of assistant teachers, each to support two classrooms.

<sup>5</sup>Weighted percent agreement is the mean of  $1 - (k - l)^2 / (q_{\max} - q_{\min})^2$  across all pairs of raters, where  $k$  and  $l$  are two ratings and  $q$  is the set of categories used to rate subjects.



accounts for the magnitude of the disagreement, but it does not account for chance agreement. Hence, third, Cohen's weighted  $\kappa$  was used, which accounts for chance agreement and the size or magnitude of the disagreements (i.e., partial agreements). The weighted  $\kappa$  statistic compares observed agreement between two raters on an ordinal scale while accounting for the agreement expected by chance alone and allows disagreements to contribute to the measure of agreement as a function of how disparate the disagreements are.<sup>6</sup> Fleiss et al. (2003) provided rough rules of thumb for interpreting weighted  $\kappa$ : > 0.75 signifies excellent agreement, and < 0.40 signifies poor agreement.

These three metrics—percent absolute agreement, weighted percent agreement, and weighted Cohen's kappa—were used to analyze interrater reliability for agreement between domain scaled scores on the comprehensive, essential, and modified essential sets for the DRDP IT, PS, and K views. These same metrics were used to calculate measure (i.e., item) level interrater reliability for all measures from the IT, PS, and K views.

While both domain level and measure level interrater reliability are important for any assessment, domain level agreement is especially important for the DRDP because domain ratings are often used for reporting a child's progress to parents and teachers and for aggregate research and reporting.

## Results

---

This section presents evidence of interrater reliability for DRDP domain and measure ratings, focusing on the weighted Cohen's kappa statistic. Note that the analysis results using the other two metrics (i.e., percent absolute agreement and weighted percent agreement) are not reported in this section as they had similar patterns to the weighted Cohen's kappa. Instead, all those results, along with the detailed results for the weighted Cohen's kappa, are included in Appendices A-F.

### INTERRATER RELIABILITY EVIDENCE AT THE DOMAIN/SUB-DOMAIN LEVEL

The evidence of interrater reliability between pairs of raters for DRDP domain developmental levels is presented by DRDP views. We first report the results for the DRDP IT view, including the comprehensive set, the essential set, and the modified essential set. Then we report the results for the DRDP PS view and the DRDP K view. Within each view, results are presented for the comprehensive set, the fundamental set, the essential set, and the modified essential set.

*The DRDP IT View:* First, the weighted  $\kappa$  were excellent for the comprehensive set of the DRDP IT view, ranging between 0.77-0.88 for pairs of raters across the domains and sub-domains. Similarly, strong

---

<sup>6</sup>When specifying weights, we used quadratic weights, also known as the Fleiss-Cohen weights. With ordinal data, weighted  $\kappa$  is asymptotically equivalent to the intraclass coefficient (ICC) from a two-way random effects ANOVA (Fleiss & Cohen, 1973), which is another common method for calculating interrater agreement between two ratings from multiple sets of raters.

relationships between raters were observed for the essential and modified essential sets of domain and sub-domain ratings for the DRDP IT view (0.77-0.88 and 0.78-0.82, respectively).

*The DRDP PS View:* The interrater reliability for the comprehensive set of the DRDP PS view was moderate to excellent, ranging from 0.68-0.81 for pairs of raters across the domains and sub-domains. Similarly moderate to excellent reliability was observed for the other sets of the DRDP PS view, ranging between 0.69-0.81 (fundamental set), 0.73-0.76 (essential set), and 0.72-0.80 (modified essential set), for pairs of raters across the domains and sub-domains.

*The DRDP K View:* Interrater reliability for the comprehensive set of the DRDP K view was moderate, ranging between 0.40-0.68 for pairs of raters across the domains and sub-domains. Similar patterns of interrater reliability were observed for the other sets of the DRDP K view, ranging between 0.51-0.68 (fundamental set), 0.51-0.68 (essential set), and 0.55-0.69 (modified essential set), for pairs of raters across the domains and sub-domains. The discussion section explores the reasons why interrater reliability was lower for the DRDP K view compared to the DRDP IT and PS views.

## INTERRATER RELIABILITY EVIDENCE AT THE MEASURE LEVEL

The evidence of interrater reliability for DRDP measures between pairs of raters is presented in this section by DRDP views and domains. We first report the results for the five domains for the DRDP IT view: (1) Approaches to Learning—Self-Regulation (ATL-REG), (2) Social and Emotional Development (SED), (3) Language and Literacy Development (LLD), (4) Cognition, including Math and Science (COG), and (5) Physical Development—Health (PD-HLTH). Next, we present the results for the DRDP PS and K views, each containing the five domains listed above and three additional domains: (1) English Language Development (ELD), (2) History—Social Science (HSS), and (3) Visual and Performance Arts (VPA).

*The DRDP IT View:* Across five domains, the measure level interrater reliability, again expressed as weighted  $\kappa$ , was excellent for the measures in the DRDP IT view. The weighted  $\kappa$ s ranged between 0.70-0.77 for ATL-REG, 0.74-0.80 for SED, 0.74-0.83 for LLD, 0.72-0.79 for COG, and 0.79-0.87 for PD-HLTH.

*The DRDP PS View:* For the DRDP PS view, the measure level weighted  $\kappa$ s were excellent for six domains, ranging between 0.70-0.73 for SED, 0.69-0.77 for LLD, 0.69-0.76 for ELD, 0.60-0.70 for PD-HLTH, 0.64-0.75 for HSS, and 0.60-0.71 for VPA. The weighted  $\kappa$ s were moderate to high for measure level ratings for ATL-REG and COG as they ranged between 0.46-0.70 and 0.48-0.77, respectively.

*The DRDP K View:* For four domains of the DRDP K view, moderate weighted  $\kappa$ s were observed. Specifically, the measure level weighted  $\kappa$ s ranged between 0.43-0.61 for SED, 0.41-0.67 for LLD, 0.46-0.70 for ELD, and 0.50-0.66 for HSS. It is noteworthy that, for the other four domains, while most measures had moderate to high weighted  $\kappa$ s between pairs of kindergarten raters, one or two measures were found to have low weighted  $\kappa$  within each domain. Specifically, in COG, the weighted  $\kappa$  for “Measurement” was 0.33, and the remaining measures ranged between 0.43-0.63. In PD-HLTH, the weighted  $\kappa$  for “Gross Locomotor Movement Skills” was 0.27, and the remaining measures ranged between 0.50-0.69. In VPA, the weighted  $\kappa$  for “Visual Art” was 0.35, and the remaining measures

ranged between 0.48-0.56. And in ATL-REG, the weighted  $\kappa$ s for “Curiosity and Initiative in Learning” and “Self-Control of Feelings and Behavior” were 0.22 and 0.15, and the remaining measures ranged between 0.49-0.58.

## Discussion and Limitations

---

The interrater reliability of the DRDP for infant/toddler, preschool, and kindergarten settings was studied as recommended in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). We used the weighted Cohen’s kappa statistic to analyze the agreement between pairs of raters at the domain and measure level for the three DRDP views (IT, PS, and K). Overall, we observed moderate to high interrater reliability for the DRDP IT and PS views at both the domain and measure levels. For the DRDP K view, we observed moderate interrater reliability for all the domains and most measures. In sum, the results show adequate to strong interrater reliability across the DRDP views, measures, and domains/sub-domains.

Lower interrater reliability was demonstrated in the DRDP K view than in the IT and PS views. We recognize that the characteristics of the kindergarten rater sample—in particular the secondary raters—may have played a role in the lower observed interrater reliability in the kindergarten sample. Our sampling strategy for kindergarten raters capitalized upon a unique situation where kindergarten classrooms in one large district had hired a set of assistant teachers, with one assistant teacher splitting time between two different kindergarten classrooms. Primary kindergarten classroom teachers in the district served as the primary raters for the K view whereas the assistant teachers served as secondary raters. The secondary kindergarten raters were less experienced with the DRDP and/or with using observations to reflect on children’s learning and development compared to primary raters for the K view and compared to secondary raters for the IT and PS views. The secondary kindergarten raters received DRDP-K training and collaborated with the primary raters to complete the DRDP during one prior administration. For most secondary kindergarten raters, the winter assessment period was the first time completing the K view independently. A follow-up analysis showed that the secondary kindergarten raters systematically rated children at earlier developmental levels than the primary kindergarten raters. The differences in interrater reliability between K view (lower) compared to the IT and PS views (both higher) may be, at least in part, related to the differences in the raters’ training and experience. We recognize the possibility that novice assessors may be more conservative when they are completing the DRDP on their own for the first time, as evidenced by the secondary kindergarten raters systematically choosing earlier developmental levels compared to the more experienced primary kindergarten raters.

At the measure level, we observed low interrater reliability on five out of 51 measures<sup>7</sup> of the K view. We reason that the content areas for some of these measures may be less familiar to novice

---

<sup>7</sup>The DRDP comprehensive K view contains 55 measures. One domain with four measures, Language and Literacy Development in Spanish, was omitted from the current study because it is designed to be used in kindergarten

kindergarten teachers (e.g., Curiosity and Initiative in Learning, Self-Control of Feelings and Behavior) or that kindergarten settings may provide fewer opportunities to observe children in these activities (e.g., Gross Locomotor Movement Skills, Measurement, Visual Art), especially when the assistant teachers serving as secondary raters support two different classrooms and may not be present with children for the entire day. Additional exploration is needed, such as interviewing teachers and assistants about their experiences with these measures.

Another limitation relates to the small size and diversity of the sample for the K view. The smaller sample of kindergarten children and raters came from a single district in California's Central Valley, while the infant/toddler and preschool samples were drawn from multiple districts. This leads to reasonable questions about the representativeness of the sample and the generalizability of the results. However, the limitation is also acceptable given the rare opportunity for encountering naturally occurring rating pairs in kindergarten classrooms, the importance of interrater reliability, the state of the literature and need for information about interrater reliability with observational assessments in kindergarten, and the provisional conclusions described above about novice raters.

In terms of the rigor of the current analysis, it is noteworthy that the results in this report are based on weighted Cohen's kappa, which is more stringent in addressing agreement based on chance than the percent agreement and percent adjusted agreement statistics used in many early childhood research analyses of interrater reliability. The weighted  $\kappa$  statistic showed moderate to high interrater reliability for both domain and measure levels for the IT and PS views, and moderate interrater reliability for all the domains and most measures in the K view. In addition, we calculated percent exact agreement and weighted percent agreement (reported in the appendices). These statistics, in particular weighted percent agreement, showed very strong interrater agreement at both domain and measure levels across the various views and sets of the DRDP (see Appendices A-F). The weighted percent agreement results showed that even when exact agreement between raters was not present, the two ratings were close to one another, which is a desirable property in an observational developmental assessment. Finally, to place the findings in context, the results reported here for the three DRDP views were quite strong compared to the interrater reliability measured by percent exact agreement for other instruments using a master coder (i.e., Soderberg et al., 2013).

In conclusion, this study demonstrated that the DRDP has moderate to high interrater reliability for DRDP IT, PS, and K views across developmental domains, as well as for the majority of individual measures. The findings of this study, along with other evidence on the validity of the DRDP (Draney et al., 2022; Chen-Gaddini et al., 2022), strengthen the argument that the DRDP is a reliable and valid

---

classrooms where Spanish is the language of instruction. All kindergarten classrooms in the current sample used English as the primary language of instruction.

assessment of children’s learning and development for use in early infancy through kindergarten settings.

## Acknowledgments

---

The authors would like to thank Zsofia Tallai, Jeannie Nguyen, Alec Jimenez, Cherry Hanna, and Alex Dang-Lozano for data and report preparation; Sara Miller for recruiting study participants; and Amy Reff and Katie Ranftle for copyedit support. The authors also express their gratitude and appreciation to the numerous infant/toddler, preschool, and kindergarten teachers and administrators who participated in DRDP research studies, including the current study.

## Appendix A: Domain Level Interrater reliability – Infant/Toddler (IT) View

The tables in this appendix show interrater reliability evidence at the domain/sub-domain level, across various sets, for the DRDP IT view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table A. 1. Interrater reliability evidence for each domain and sub-domain of the comprehensive set for the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.62 [.55, .69]	.99 [.92, 1.00]	.78 [.70, .85]
<b>Social and Emotional Development (SED)</b>	.67 [.60, .74]	.99 [.92, 1.00]	.83 [.76, .90]
<b>Language and Literacy Development (LLD)</b>	.62 [.55, .69]	.98 [.92, 1.00]	.81 [.73, .88]
<b>Cognition, including Math and Science (COG)</b>	.67 [.60, .74]	.99 [.92, 1.00]	.77 [.69, .85]
<b>Physical Development—Health (PD-HLTH)</b>	.66 [.59, .73]	.99 [.92, 1.00]	.88 [.81, .95]
<b>Physical Development (PD)</b>	.67 [.60, .74]	.99 [.92, 1.00]	.88 [.81, .94]
<b>Health (HLTH)</b>	.64 [.57, .71]	.99 [.92, 1.00]	.85 [.78, .92]

**Table A. 2. Interrater reliability for each domain and sub-domain of the essential set for the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.63 [.56, .70]	.99 [.92, 1.00]	.77 [.70, .85]
<b>Social and Emotional Development (SED)</b>	.63 [.56, .70]	.99 [.92, 1.00]	.81 [.74, .88]
<b>Language and Literacy Development (LLD)</b>	.62 [.55, .69]	.98 [.91, 1.00]	.81 [.73, .88]
<b>Cognition, including Math and Science (COG)</b>	.64 [.57, .71]	.97 [.91, 1.00]	.78 [.71, .86]
<b>Physical Development (PD)</b>	.67 [.60, .74]	.99 [.92, 1.00]	.88 [.81, .94]

**Table A. 3. Interrater reliability evidence for each domain of the modified essential set for the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED)</b>	.65 [.58, .72]	.99 [.92, 1.00]	.82 [.74, .89]
<b>Language and Literacy Development (LLD)</b>	.62 [.55, .69]	.98 [.91, 1.00]	.81 [.73, .88]
<b>Cognition, including Math and Science (COG)</b>	.64 [.57, .71]	.97 [.91, 1.00]	.78 [.71, .86]

## Appendix B: Domain Level Interrater reliability – Preschool (PS) View

The tables in this appendix show interrater reliability evidence at the domain/sub-domain level, across various sets, for the DRDP PS view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table B. 1. Interrater reliability evidence for each domain and sub-domain of the comprehensive set for the DRDP PS view**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.53 [.45, .61]	.98 [.88, 1.00]	.69 [.59, .80]
<b>Social and Emotional Development (SED)</b>	.60 [.52, .69]	.99 [.89, 1.00]	.77 [.67, .87]
<b>Language and Literacy Development (LLD)</b>	.62 [.54, .71]	.99 [.90, 1.00]	.81 [.71, .90]
<b>Literacy (LIT)</b>	.63 [.55, .72]	.97 [.87, 1.00]	.75 [.65, .85]
<b>Language (LANG)</b>	.59 [.51, .68]	.99 [.89, 1.00]	.79 [.69, .89]
<b>Cognition (COG)</b>	.57 [.48, .65]	.99 [.89, 1.00]	.77 [.68, .87]
<b>Math (MATH)</b>	.59 [.51, .68]	.98 [.89, 1.00]	.75 [.66, .85]
<b>Science (SCI)</b>	.62 [.53, .70]	.99 [.89, 1.00]	.77 [.67, .86]
<b>Physical Development—Health (PD-HLTH)</b>	.58 [.49, .66]	.98 [.89, 1.00]	.73 [.62, .83]
<b>Physical Development (PD)</b>	.59 [.51, .68]	.98 [.89, 1.00]	.74 [.65, .84]
<b>Health (HLTH)</b>	.63 [.55, .72]	.99 [.89, 1.00]	.77 [.67, .87]
<b>History—Social science (HSS)</b>	.68 [.59, .77]	.97 [.88, 1.00]	.72 [.62, .82]
<b>Visual and Performing Arts (VPA)</b>	.66 [.57, .75]	.97 [.87, 1.00]	.68 [.56, .80]
<b>English Language Development (ELD)</b>	.65 [.50, .80]	.96 [.79, 1.00]	.72 [.53, .92]



**Table B. 2. Interrater reliability evidence for each domain of the fundamental set for the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.53 [.45, .61]	.98 [.88, 1.00]	.69 [.59, .80]
<b>Social and Emotional Development (SED)</b>	.60 [.52, .69]	.99 [.89, 1.00]	.77 [.67, .87]
<b>Language and Literacy Development (LLD)</b>	.62 [.54, .71]	.99 [.90, 1.00]	.81 [.71, .90]
<b>Cognition, including Math and Science (COG)</b>	.62 [.53, .70]	.99 [.89, 1.00]	.77 [.68, .87]
<b>Physical Development—Health (PD-HLTH)</b>	.58 [.49, .66]	.98 [.89, 1.00]	.73 [.62, .83]
<b>English Language Development (ELD)</b>	.65 [.50, .80]	.96 [.79, 1.00]	.72 [.53, .92]

**Table B. 3. Interrater reliability evidence for each domain and sub-domain of the fundamental set for the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.54 [.46, .62]	.97 [.88, 1.00]	.70 [.60, .81]
<b>Social and Emotional Development (SED)</b>	.60 [.51, .68]	.98 [.89, 1.00]	.76 [.66, .85]
<b>Language and Literacy Development (LLD)</b>	.58 [.50, .67]	.99 [.89, 1.00]	.80 [.70, .89]
<b>English Language Development (ELD)</b>	.65 [.50, .80]	.96 [.79, 1.00]	.72 [.53, .92]
<b>Physical Development (PD)</b>	.59 [.51, .68]	.98 [.89, 1.00]	.74 [.65, .84]
<b>Math (MATH)</b>	.59 [.50, .67]	.98 [.89, 1.00]	.76 [.66, .85]

**Table B. 4. Interrater reliability evidence for each domain and sub-domain of the modified essential set for the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED)</b>	.58 [.50, .66]	.98 [.89, 1.00]	.74 [.64, .84]
<b>Language and Literacy Development (LLD)</b>	.58 [.50, .67]	.99 [.89, 1.00]	.80 [.70, .89]
<b>English Language Development (ELD)</b>	.65 [.50, .80]	.96 [.79, 1.00]	.72 [.53, .92]
<b>Math (MATH)</b>	.61 [.52, .69]	.98 [.89, 1.00]	.77 [.68, .87]

## Appendix C: Domain Level Interrater reliability – Kindergarten (K) View

The tables in this appendix show interrater reliability evidence at the domain/sub-domain level, across various sets, for the DRDP K view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table C. 1. Interrater reliability evidence for each domain and sub-domain of the comprehensive set for the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.39 [.30, .49]	.94 [.92, .95]	.51 [.40, .62]
<b>Social and Emotional Development (SED)</b>	.52 [.42, .62]	.96 [.95, .97]	.68 [.59, .78]
<b>Language and Literacy Development (LLD)</b>	.48 [.38, .59]	.96 [.95, .97]	.67 [.58, .77]
<b>Literacy (LIT)</b>	.40 [.31, .50]	.96 [.94, .97]	.64 [.55, .73]
<b>Language (LANG)</b>	.40 [.30, .50]	.95 [.92, .97]	.60 [.49, .71]
<b>Cognition (COG)</b>	.51 [.40, .61]	.96 [.91, 1.0]	.61 [.50, .72]
<b>Math (MATH)</b>	.54 [.44, .64]	.95 [.93, .98]	.58 [.44, .71]
<b>Science (SCI)</b>	.45 [.34, .55]	.95 [.90, 1.0]	.59 [.48, .70]
<b>Physical Development—Health (PD-HLTH)</b>	.54 [.44, .64]	.96 [.95, .98]	.59 [.45, .73]
<b>Physical Development (PD)</b>	.48 [.38, .59]	.95 [.93, .96]	.56 [.41, .70]
<b>Health (HLTH)</b>	.66 [.56, .75]	.97 [.96, .98]	.61 [.50, .73]
<b>History—Social science (HSS)</b>	.56 [.45, .67]	.92 [.86, .98]	.40 [.22, .58]
<b>Visual and Performing Arts (VPA)</b>	.55 [.45, .65]	.88 [.84, .92]	.46 [.30, .63]
<b>English Language Development (ELD)</b>	.46 [.29, .63]	.96 [.94, .98]	.64 [.38, .90]

**Table C. 2. Interrater reliability evidence for each domain of the fundamental set for the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.39 [.30, .49]	.94 [.92, .95]	.51 [.40, .62]
<b>Social and Emotional Development (SED)</b>	.52 [.42, .62]	.96 [.95, .97]	.68 [.59, .78]

<b>Language and Literacy Development (LLD)</b>	.48 [.38, .59]	.96 [.95, .97]	.67 [.58, .77]
<b>Cognition, including Science and Math (COG)</b>	.51 [.40, .61]	.96 [.91, 1.0]	.61 [.50, .72]
<b>Physical Development—Health (PD-HLTH)</b>	.54 [.44, .64]	.96 [.95, .98]	.59 [.45, .73]
<b>English Language Development (ELD)</b>	.46 [.29, .63]	.96 [.94, .98]	.64 [.38, .90]

**Table C. 3. Interrater reliability evidence for each domain and sub-domain of the essential set for the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG)</b>	.39 [.30, .49]	.94 [.92, .95]	.51 [.40, .61]
<b>Social and Emotional Development (SED)</b>	.51 [.40, .60]	.96 [.95, .97]	.68 [.58, .77]
<b>Language and Literacy Development (LLD)</b>	.42 [.32, .52]	.96 [.94, .98]	.63 [.53, .73]
<b>English Language Development (ELD)</b>	.46 [.29, .63]	.96 [.94, .98]	.64 [.38, .90]
<b>Physical Development (PD)</b>	.48 [.38, .59]	.95 [.93, .96]	.56 [.41, .70]
<b>Math (MATH)</b>	.54 [.44, .64]	.95 [.93, .98]	.58 [.44, .71]

**Table C. 4. Interrater reliability evidence for each domain and sub-domain of the modified essential set for the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED)</b>	.55 [.45, .65]	.96 [.95, .97]	.69 [.60, .79]
<b>Language and Literacy Development (LLD)</b>	.42 [.32, .52]	.96 [.94, .98]	.63 [.53, .73]
<b>English Language Development (ELD)</b>	.46 [.29, .63]	.96 [.94, .98]	.64 [.38, .90]
<b>Math (MATH)</b>	.46 [.36, .57]	.95 [.92, .97]	.55 [.43, .68]

## Appendix D: Measure Level Interrater reliability – Infant/Toddler (IT) View

The tables in this appendix show interrater reliability evidence at the measure level for the DRDP IT view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table D. 1. Interrater reliability evidence for each measure in the Approaches to Learning—Self-Regulation domain of the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning–Self-Regulation (ATL-REG) 1</b> Attention Maintenance	.64 [.57, .71]	.96 [.90, 1.00]	.73 [.65, .82]
<b>ATL-REG 2</b> Self-Comforting	.62 [.55, .69]	.97 [.90, 1.00]	.70 [.62, .78]
<b>ATL-REG 3</b> Imitation	.61 [.54, .68]	.97 [.90, 1.00]	.76 [.69, .84]
<b>ATL-REG 4</b> Curiosity and Initiative in Learning	.64 [.57, .71]	.98 [.92, 1.00]	.77 [.70, .85]
<b>ATL-REG 5</b> Self-Control of Feelings and Behavior	.62 [.55, .69]	.97 [.91, 1.00]	.72 [.64, .80]

**Table D. 2. Interrater reliability evidence for each measure in the Social and Emotional Development domain of the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED) 1</b> Identity of Self in Relation to Others	.67 [.60, .74]	.98 [.92, 1.00]	.80 [.73, .87]
<b>SED 2</b> Social and Emotional Understanding	.59 [.52, .66]	.98 [.92, 1.00]	.74 [.66, .82]
<b>SED 3</b> Relationships and Social Interactions With Familiar Adults	.62 [.55, .69]	.97 [.91, 1.00]	.77 [.69, .85]
<b>SED 4</b>	.62 [.55, .69]	.98 [.92, 1.00]	.78 [.70, .85]

<b>Relationships and Social Interactions With Peers</b>			
<b>SED 5</b>			
<b>Symbolic and Sociodramatic Play</b>	.63 [.56, .70]	.98 [.92, 1.00]	.77 [.70, .85]

**Table D. 3. Interrater reliability evidence for each measure in the Language and Literacy Development domain of the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Language and Literacy Development (LLD) 1</b>			
<b>Understanding of Language (Receptive)</b>	.61 [.54, .68]	.98 [.92, 1.00]	.80 [.72, .87]
<b>LLD 2</b>			
<b>Responsiveness to Language</b>	.61 [.54, .68]	.98 [.92, 1.00]	.79 [.71, .86]
<b>LLD 3</b>			
<b>Communication and Use of Language</b>	.64 [.57, .71]	.99 [.92, 1.00]	.83 [.76, .91]
<b>LLD 4</b>			
<b>Reciprocal Communication and Conversation</b>	.62 [.55, .70]	.98 [.92, 1.00]	.78 [.70, .87]
<b>LLD 5</b>			
<b>Interest in Literacy</b>	.58 [.51, .64]	.98 [.92, 1.00]	.74 [.66, .82]

**Table D. 4. Interrater reliability evidence for each measure in the Cognition domain of the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Cognition (COG) 1</b>			
<b>Spatial Relationships</b>	.64 [.56, .71]	.96 [.90, 1.00]	.74 [.65, .82]
<b>COG 2</b>			
<b>Classification</b>	.64 [.57, .71]	.98 [.92, 1.00]	.78 [.70, .85]
<b>COG 3</b>			
<b>Number Sense of Quantity</b>	.70 [.63, .77]	.99 [.92, 1.00]	.79 [.72, .87]
<b>COG 8</b>			
	.60 [.53, .67]	.98 [.92, 1.00]	.76 [.68, .83]

<b>Cause and Effect</b>			
<b>COG 9</b> <b>Inquiry Through Observation and Investigation</b>	.65 [.58, .72]	.98 [.91, 1.00]	.73 [.65, .82]
<b>COG 11</b> <b>Knowledge of the Natural World</b>	.63 [.56, .70]	.98 [.91, 1.00]	.72 [.64, .81]

**Table D. 5. Interrater reliability evidence for each measure in the Physical Development—Health domain of the DRDP IT view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen’s kappa
<b>Physical Development—Health (PD-HLTH) 1</b> <b>Perceptual-Motor Skills and Movement Concepts</b>	.68 [.61, .75]	.98 [.92, 1.00]	.81 [.73, .88]
<b>PD-HLTH 2</b> <b>Gross Locomotor Movement Skills</b>	.69 [.62, .76]	.99 [.92, 1.00]	.87 [.80, .94]
<b>PD-HLTH 3</b> <b>Gross Motor Manipulative Skills</b>	.65 [.58, .72]	.98 [.92, 1.00]	.82 [.75, .90]
<b>PD-HLTH 4</b> <b>Fine Motor Manipulative Skills</b>	.58 [.51, .65]	.98 [.91, 1.00]	.79 [.71, .87]
<b>PD-HLTH 5</b> <b>Safety</b>	.62 [.55, .69]	.99 [.92, 1.00]	.80 [.73, .87]
<b>PD-HLTH 6</b> <b>Personal Care Routines: Hygiene</b>	.62 [.55, .69]	.98 [.92, 1.00]	.80 [.72, .88]
<b>PD-HLTH 7</b> <b>Personal Care Routines: Feeding</b>	.62 [.55, .70]	.98 [.92, 1.00]	.81 [.74, .88]
<b>PD-HLTH 8</b> <b>Personal Care Routines: Dressing</b>	.62 [.55, .69]	.98 [.91, 1.00]	.79 [.71, .87]

## Appendix E: Measure Level Interrater reliability – Preschool (PS) View

The tables in this appendix show interrater reliability evidence at the measure level for the DRDP PS view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table E. 1. Interrater reliability evidence for each measure in the Approaches to Learning—Self-Regulation domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG) 1</b> Attention Maintenance	.69 [.59, .79]	.97 [.86, 1.00]	.56 [.41, .71]
<b>ATL-REG 2</b> Self-Comforting	.62 [.52, .72]	.95 [.84, 1.00]	.46 [.31, .60]
<b>ATL-REG 3</b> Imitation	.68 [.58, .78]	.97 [.86, 1.00]	.58 [.45, .71]
<b>ATL-REG 4</b> Curiosity and Initiative in Learning	.52 [.44, .60]	.98 [.89, 1.00]	.70 [.60, .80]
<b>ATL-REG 5</b> Self-Control of Feelings and Behavior	.51 [.43, .59]	.97 [.88, 1.00]	.69 [.58, .80]
<b>ATL-REG 6</b> Engagement and Persistence	.49 [.41, .57]	.96 [.87, 1.00]	.63 [.53, .74]
<b>ATL-REG 7</b> Shared Use of Space and Materials	.56 [.48, .65]	.96 [.87, 1.00]	.64 [.54, .75]

**Table E. 2. Interrater reliability evidence for each measure in the Social and Emotional Development domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED) 1</b> Identity of Self in Relation to Others	.52 [.44, .60]	.98 [.88, 1.00]	.72 [.63, .82]
<b>SED 2</b> Social and Emotional Understanding	.54 [.46, .62]	.98 [.89, 1.00]	.72 [.62, .82]
<b>SED 3</b>	.58 [.50, .66]	.98 [.89, 1.00]	.70 [.58, .81]



<b>Relationships and Social Interactions With Familiar Adults</b>			
<b>SED 4</b>			
<b>Relationships and Social Interactions With Peers</b>	.57 [.48, .65]	.98 [.89, 1.00]	.72 [.62, .82]
<b>SED 5</b>			
<b>Symbolic and Sociodramatic Play</b>	.62 [.53, .70]	.98 [.89, 1.00]	.73 [.63, .83]

**Table E. 3. Interrater reliability evidence for each measure in the Language and Literacy Development domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Language and Literacy Development (LLD) 1</b>			
<b>Understanding of Language (Receptive)</b>	.55 [.47, .63]	.98 [.89, 1.00]	.75 [.66, .84]
<b>LLD 2</b>			
<b>Responsiveness to Language</b>	.56 [.48, .64]	.98 [.89, 1.00]	.71 [.61, .82]
<b>LLD 3</b>			
<b>Communication and Use of Language</b>	.57 [.49, .66]	.99 [.89, 1.00]	.77 [.67, .86]
<b>LLD 4</b>			
<b>Reciprocal Communication and Conversation</b>	.60 [.52, .69]	.98 [.89, 1.00]	.77 [.67, .87]
<b>LLD 5</b>			
<b>Interest in Literacy</b>	.56 [.47, .64]	.98 [.89, 1.00]	.70 [.60, .80]
<b>LLD 6</b>			
<b>Comprehension of Age-Appropriate Text</b>	.61 [.52, .69]	.97 [.87, 1.00]	.77 [.66, .88]
<b>LLD 7</b>			
<b>Concepts About Print</b>	.51 [.43, .59]	.97 [.87, 1.00]	.69 [.59, .79]
<b>LLD 8</b>			
<b>Phonological Awareness</b>	.58 [.50, .66]	.97 [.88, 1.00]	.71 [.61, .81]
<b>LLD 9</b>			
<b>Letter and Word Knowledge</b>	.54 [.46, .62]	.96 [.87, 1.00]	.73 [.62, .83]
<b>LLD 10</b>			
<b>Emergent Writing</b>	.56 [.47, .64]	.97 [.88, 1.00]	.74 [.65, .84]

**Table E. 4. Interrater reliability evidence for each measure in the English Language Development domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>English Language Development (ELD) 1</b>			
Comprehension of English (Receptive English)	.73 [.58, .89]	.96 [.80, 1.00]	.69 [.46, .93]
<b>ELD 2</b>			
Self-Expression in English (Expressive English)	.73 [.58, .88]	.97 [.80, 1.00]	.73 [.53, .94]
<b>ELD 3</b>			
Understanding and Response to English Literacy Activities	.70 [.54, .85]	.97 [.80, 1.00]	.74 [.54, .94]
<b>ELD 4</b>			
Symbol, Letter, and Print Knowledge in English	.68 [.53, .83]	.97 [.81, 1.00]	.76 [.59, .92]

**Table E. 5. Interrater reliability evidence for each measure in the Cognition, Including Math and Science domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Cognition (COG) 1</b>			
Spatial Relationships	.71 [.61, .82]	.96 [.86, 1.00]	.48 [.30, .65]
<b>COG 2</b>			
Classification	.55 [.47, .63]	.98 [.88, 1.00]	.67 [.56, .77]
<b>COG 3</b>			
Number Sense of Quantity	.60 [.51, .68]	.98 [.89, 1.00]	.77 [.67, .87]
<b>COG 4</b>			
Number Sense of Math Operations	.57 [.49, .65]	.97 [.87, 1.00]	.74 [.64, .84]
<b>COG 5</b>			
Measurement	.58 [.50, .66]	.96 [.87, 1.00]	.67 [.56, .78]
<b>COG 6</b>			
Patterning	.62 [.53, .70]	.97 [.87, 1.00]	.73 [.62, .84]

<b>COG 7</b> Shapes	.58 [.50, .67]	.96 [.86, 1.00]	.63 [.53, .74]
<b>COG 8</b> Cause and Effect	.57 [.49, .66]	.98 [.89, 1.00]	.70 [.60, .80]
<b>COG 9</b> Inquiry Through Observation and Investigation	.57 [.48, .65]	.98 [.89, 1.00]	.74 [.64, .84]
<b>COG 10</b> Documentation and Communication	.61 [.53, .70]	.97 [.88, 1.00]	.77 [.67, .87]
<b>COG 11</b> Knowledge of the Natural World	.58 [.49, .66]	.98 [.89, 1.00]	.73 [.63, .83]

**Table E. 6. Interrater reliability evidence for each measure in the Physical Development—Health domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Physical Development—Health (PD-HLTH) 1</b> Perceptual-Motor Skills and Movement Concepts	.55 [.47, .63]	.98 [.88, 1.00]	.67 [.56, .78]
<b>PD-HLTH 2</b> Gross Locomotor Movement Skills	.62 [.53, .71]	.98 [.89, 1.00]	.69 [.58, .79]
<b>PD-HLTH 3</b> Gross Motor Manipulative Skills	.54 [.46, .62]	.98 [.88, 1.00]	.69 [.58, .80]
<b>PD-HLTH 4</b> Fine Motor Manipulative Skills	.52 [.44, .60]	.98 [.88, 1.00]	.70 [.61, .79]
<b>PD-HLTH 5</b> Safety	.58 [.50, .67]	.98 [.89, 1.00]	.70 [.59, .80]
<b>PD-HLTH 6</b> Personal Care Routines: Hygiene	.58 [.49, .66]	.98 [.88, 1.00]	.63 [.51, .75]
<b>PD-HLTH 7</b> Personal Care Routines: Feeding	.60 [.52, .68]	.98 [.89, 1.00]	.60 [.48, .71]
<b>PD-HLTH 8</b> Personal Care Routines: Dressing	.65 [.57, .74]	.98 [.89, 1.00]	.69 [.58, .79]

<b>PD-HLTH 9</b> Active Physical Play	.55 [.47, .63]	.96 [.87, 1.00]	.63 [.51, .74]
<b>PD-HLTH 10</b> Nutrition	.57 [.48, .65]	.96 [.87, 1.00]	.68 [.57, .78]

**Table E. 7. Interrater reliability evidence for each measure in the History—Social Science domain of the DRDP PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen’s kappa
<b>History—Social science (HSS) 1</b> Sense of Time	.60 [.52, .68]	.97 [.88, 1.00]	.75 [.64, .85]
<b>HSS 2</b> Sense of Place	.60 [.51, .68]	.97 [.87, 1.00]	.73 [.64, .83]
<b>HSS 3</b> Ecology	.55 [.47, .63]	.96 [.87, 1.00]	.70 [.59, .80]
<b>HSS 4</b> Conflict Negotiation	.58 [.50, .67]	.96 [.87, 1.00]	.71 [.60, .81]
<b>HSS 5</b> Responsible Conduct as a Group Member	.55 [.47, .63]	.95 [.86, 1.00]	.64 [.52, .76]

**Table E. 8. Interrater reliability evidence for each measure in the Visual and Performing Arts domain of the PS view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen’s kappa
<b>Visual and Performing Arts (VPA) 1</b> Visual Art	.59 [.51, .68]	.97 [.87, 1.00]	.71 [.60, .83]
<b>VPA 2</b> Music	.57 [.49, .66]	.95 [.86, 1.00]	.60 [.46, .73]
<b>VPA 3</b> Drama	.53 [.45, .61]	.96 [.86, 1.00]	.67 [.56, .78]
<b>VPA 4</b> Dance	.56 [.48, .64]	.95 [.86, 1.00]	.62 [.50, .74]

## Appendix F: Measure Level Interrater reliability – Kindergarten (K) View

The tables in this appendix show interrater reliability evidence at the measure level for the DRDP K view. In each table, the 95 percent confidence intervals are included in brackets following the estimate.

**Table F. 1. Interrater reliability evidence for each measure in the Approaches to Learning—Self-Regulation domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Approaches to Learning—Self-Regulation (ATL-REG) 4</b> Curiosity and Initiative in Learning	.06 [.01, .11]	.86 [.83, .89]	.22 [.13, .31]
<b>ATL-REG 5</b> Self-Control of Feelings and Behavior	.09 [.03, .15]	.93 [.91, .94]	.15 [.09, .21]
<b>ATL-REG 6</b> Engagement and Persistence	.31 [.22, .41]	.93 [.92, .95]	.49 [.36, .62]
<b>ATL-REG 7</b> Shared Use of Space and Materials	.45 [.35, .55]	.96 [.95, .97]	.58 [.47, .69]

**Table F. 2. Interrater reliability evidence for each measure in the Social and Emotional Development domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Social and Emotional Development (SED) 1</b> Identity of Self in Relation to Others	.38 [.29, .48]	.94 [.93, .96]	.61 [.51, .72]
<b>SED 2</b> Social and Emotional Understanding	.45 [.35, .55]	.96 [.95, .97]	.58 [.47, .69]
<b>SED 3</b> Relationships and Social Interactions with Familiar Adults	.53 [.43, .63]	.96 [.94, .97]	.58 [.45, .70]
<b>SED 4</b> Relationships and Social Interactions with Peers	.42 [.33, .52]	.95 [.93, .96]	.43 [.27, .58]
<b>SED 5</b> Symbolic and Sociodramatic Play	.48 [.38, .59]	.95 [.94, .97]	.61 [.48, .74]

**Table F. 3. Interrater reliability evidence for each measure in the Language and Literacy Development domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Language and Literacy Development (LLD) 1</b> Understanding of Language (Receptive)	.45 [.35, .55]	.95 [.94, .96]	.61 [.52, .71]
<b>LLD 2</b> Responsiveness to Language	.38 [.29, .48]	.95 [.94, .96]	.63 [.53, .73]
<b>LLD 3</b> Communication and Use of Language	.40 [.31, .50]	.93 [.91, .95]	.49 [.35, .63]
<b>LLD 4</b> Reciprocal Communication and Conversation	.38 [.28, .48]	.92 [.89, .95]	.45 [.30, .60]
<b>LLD 5</b> Interest in Literacy	.43 [.34, .53]	.92 [.90, .94]	.41 [.27, .56]
<b>LLD 6</b> Comprehension of Age-Appropriate Text	.45 [.35, .55]	.95 [.93, .96]	.57 [.45, .68]
<b>LLD 7</b> Concepts About Print	.47 [.37, .57]	.95 [.93, .97]	.60 [.47, .73]
<b>LLD 8</b> Phonological Awareness	.53 [.43, .63]	.95 [.94, .97]	.67 [.56, .78]
<b>LLD 9</b> Letter and Word Knowledge	.33 [.24, .43]	.94 [.92, .95]	.65 [.55, .76]
<b>LLD 10</b> Emergent Writing	.40 [.31, .50]	.94 [.93, .96]	.53 [.41, .66]

**Table F. 4. Interrater reliability evidence for each measure in the English Language Development domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>English Language Development (ELD) 1</b>			
<b>Comprehension of English (Receptive English)</b>	.49 [.32, .66]	.96 [.94, .98]	.70 [.49, .91]
<b>ELD 2</b>			
<b>Self-Expression in English (Expressive English)</b>	.51 [.34, .68]	.96 [.94, .98]	.66 [.38, .93]
<b>ELD 3</b>			
<b>Understanding and Response to English Literacy Activities</b>	.35 [.19, .51]	.92 [.89, .95]	.46 [.20, .72]
<b>ELD 4</b>			
<b>Symbol, Letter, and Print Knowledge in English</b>	.41 [.24, .57]	.93 [.90, .97]	.57 [.29, .85]

**Table F. 5. Interrater reliability evidence for each measure in the Cognition: Math sub-domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Cognition (COG): Math (MATH) 1</b>			
<b>Classification</b>	.40 [.31, .50]	.94 [.93, .96]	.48 [.34, .62]
<b>COG:MATH 2</b>			
<b>Number Sense of Quantity</b>	.46 [.36, .56]	.95 [.93, .96]	.61 [.49, .72]
<b>COG:MATH 3</b>			
<b>Number Sense of Math Operations</b>	.44 [.34, .54]	.94 [.92, .96]	.57 [.44, .70]
<b>COG:MATH 4</b>			
<b>Measurement</b>	.43 [.34, .53]	.91 [.89, .94]	.33 [.14, .51]
<b>COG:MATH 5</b>			
<b>Patterning</b>	.39 [.29, .49]	.94 [.91, .96]	.49 [.36, .61]
<b>COG:MATH 6</b>			
<b>Shapes</b>	.39 [.29, .49]	.93 [.90, .96]	.49 [.34, .64]

**Table F. 6. Interrater reliability evidence for each measure in the Cognition: Science sub-domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Cognition (COG): Science (SCI) 1</b> Cause and Effect	.39 [.29, .50]	.93 [.88, .97]	.45 [.30, .60]
<b>COG: SCI 2</b> Inquiry Through Observation and Investigation	.40 [.30, .51]	.93 [.88, .99]	.43 [.28, .58]
<b>COG: SCI 3</b> Documentation and Communication	.43 [.33, .54]	.93 [.88, .99]	.55 [.41, .69]
<b>COG: SCI 4</b> Knowledge of the Natural World	.42 [.32, .52]	.95 [.90, 1.00]	.63 [.53, .73]

**Table F. 7. Interrater reliability evidence for each measure in the Physical Development sub-domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Physical Development (PD) 1</b> Perceptual-Motor Skills and Movement Concepts	.39 [.30, .49]	.97 [.96, .98]	.51 [.35, .66]
<b>PD 2</b> Gross Locomotor Movement Skills	.47 [.37, .57]	.97 [.95, .99]	.27 [.08, .47]
<b>PD 3</b> Gross Motor Manipulative Skills	.57 [.47, .67]	.99 [.98, .99]	.56 [.42, .70]
<b>PD 4</b> Fine Motor Manipulative Skills	.46 [.36, .56]	.95 [.93, .97]	.50 [.33, .66]

**Table F. 8. Interrater reliability evidence for each measure in the Health sub-domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Health (HLTH) 1</b>	.42 [.33, .52]	.95 [.94, .96]	.56 [.44, .68]



<b>Safety</b>			
<b>HLTH 2</b> Active Physical Play	.44 [.34, .54]	.95 [.94, .97]	.61 [.50, .72]
<b>HLTH 3</b> Nutrition	.46 [.36, .56]	.97 [.96, .98]	.63 [.52, .73]
<b>HLTH 4</b> Knowledge of Wellness	.55 [.45, .65]	.96 [.94, .97]	.69 [.58, .80]

**Table F. 9. Interrater reliability evidence for each measure in the History—Social Science domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>History—Social science (HSS) 1</b> Sense of Time	.42 [.31, .52]	.94 [.88, 1.00]	.56 [.41, .72]
<b>HSS 2</b> Sense of Place	.49 [.39, .60]	.95 [.90, 1.00]	.62 [.49, .76]
<b>HSS 3</b> Ecology	.52 [.42, .63]	.93 [.89, .97]	.51 [.37, .65]
<b>HSS 4</b> Conflict Negotiation	.48 [.38, .59]	.93 [.91, .96]	.50 [.33, .68]
<b>HSS 5</b> Responsible Conduct as a Group Member	.52 [.42, .62]	.96 [.95, .97]	.66 [.54, .79]

**Table F. 10. Interrater reliability evidence for each measure in the Visual and Performing Arts domain of the DRDP K view.**

	Perfect percent agreement	Weighted percent agreement	Weighted Cohen's kappa
<b>Visual and Performing Arts (VPA) 1</b> Visual Art	.43 [.34, .53]	.93 [.91, .95]	.35 [.17, .52]
<b>VPA 2</b> Music	.45 [.35, .55]	.94 [.92, .96]	.48 [.35, .62]
<b>VPA 3</b>	.46 [.36, .56]	.94 [.92, .96]	.56 [.42, .70]

<b>Drama</b>			
<b>VPA 4</b>	.47 [.37, .57]	.95 [.93, .96]	.56 [.44, .67]
<b>Dance</b>			

---

## References

---

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- 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.
- Chen-Gaddini, M., Sussman, J., Newton, E., Jimenez, S. R., Kriener-Althen, K., Gochyyev, P., Draney, K., & Mangione, P. (2022). *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.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220.
- Draney, K., Sussman, J., Gochyyev, P., Kriener-Althen, K., Newton, E., & Mangione, P. (2021). *DRDP technical report for early infancy through kindergarten: Structural validity and reliability information for the Desired Results Development Profile*. Berkeley Evaluation and Assessment Research Center.
- DRDP Collaborative Research Group. (2018). *Technical report for the Desired Results Developmental Profile (2015)*. Report prepared for the California Department of Education.
- Fleiss, J. L., & Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, *33*(3), 613–619.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed). J. Wiley.
- Koch, G. G. (1982). Intraclass correlation coefficient. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences: Vol. 4* (pp. 213–217). John Wiley & Sons.
- National Research Council. (2008). *Early childhood assessment: Why, what, and how*. The National Academies Press.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.
- Soderberg, J., Stull, S., Cummings, K., Nolen, E., McCutchen, D., Joseph, G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills*

(*WaKIDS*). Report prepared for the State of Washington Office of Superintendent of Public Instruction.

WestEd. (2018a). *DRDP (2015) research summaries*. Report prepared for the California Department of Education. Retrieved on June 22, 2022, from <https://www.desiredresults.us/research-summaries-drdp-2015-domain>

WestEd. (2018b). *DRDP-K (2015) research summaries*. Report prepared for the California Department of Education and the Illinois State Board of Education. Retrieved on June 22, 2022, from [https://drdpk.org/res\\_sum.html](https://drdpk.org/res_sum.html)