



**Reliability and Validity of the
Desired Results Developmental Profile *access*
(DRDP *access*):
Results of the 2005–2006 Calibration Study**

Reliability and Validity of the Desired Results Developmental Profile access (DRDP access): Results of the 2005–2006 Calibration Study was developed by the Desired Results *access* Project to support the implementation of the Desired Results Developmental Profile Assessment System based on the guidelines and specifications of the Special Education Division.

© 2008 by the California Department of Education, Special Education Division

All rights reserved

Permission to reproduce for instructional purposes

The Desired Results *access* Project – A Special Project of the Napa County Office of Education is funded by the California Department of Education (CDE), Special Education Division (Contract #CN077059) to assist the CDE with developing and putting into place a system to assess the progress of California’s preschool children with disabilities.

Table of Contents

Executive Summary	5
Reliability and Validity of the Desired Results Developmental Profile <i>access</i> (DRDP <i>access</i>): Results of the 2005–2006 Calibration Study	9
The DRDP <i>access</i>	9
DRDP <i>access</i> Measurement Model.....	11
Calibration Study (2005–2006).....	12
Established Reliability	13
Content Validity.....	14
Discriminative Validity.....	15
Construct Validity.....	15
Sensitivity	20
Usability	20
References	22
Appendix: Glossary	23

Developed by the Desired Results *access* Project – A Special Project of the Napa County Office of Education.
Funded by the California Department of Education, Special Education Division. (March 2008)

Executive Summary

Purpose

This report describes evidence of validity, reliability and scaling of scores from the Desired Results Developmental Profile *access*. The objective is to provide information about the quality of scores obtained from the DRDP *access*. Data for this report was collected during the 2005–2006 calibration of the DRDP *access*. As part of the calibration, 1644 infants, toddlers, and preschoolers throughout the state of California were assessed using the DRDP *access*. For information regarding scoring and interpretation of scores from the DRDP *access*, please see *Understanding the DRDP access Measurement Model* (Desired Results *access* Project, January 2008).

Rationale

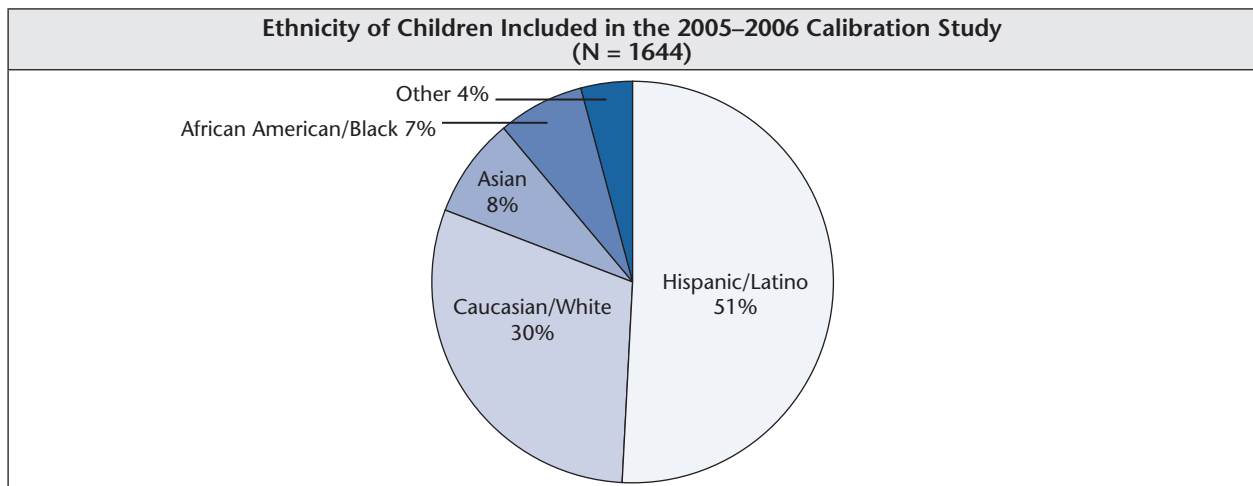
The calibration study was one step in preparing the instrument for use in the Desired Results statewide accountability system. The calibration study was conducted to estimate the statistical properties of the measures, determine if the properties met the expectations of instrument developers, and to evaluate assessors' perceived utility.

Methods

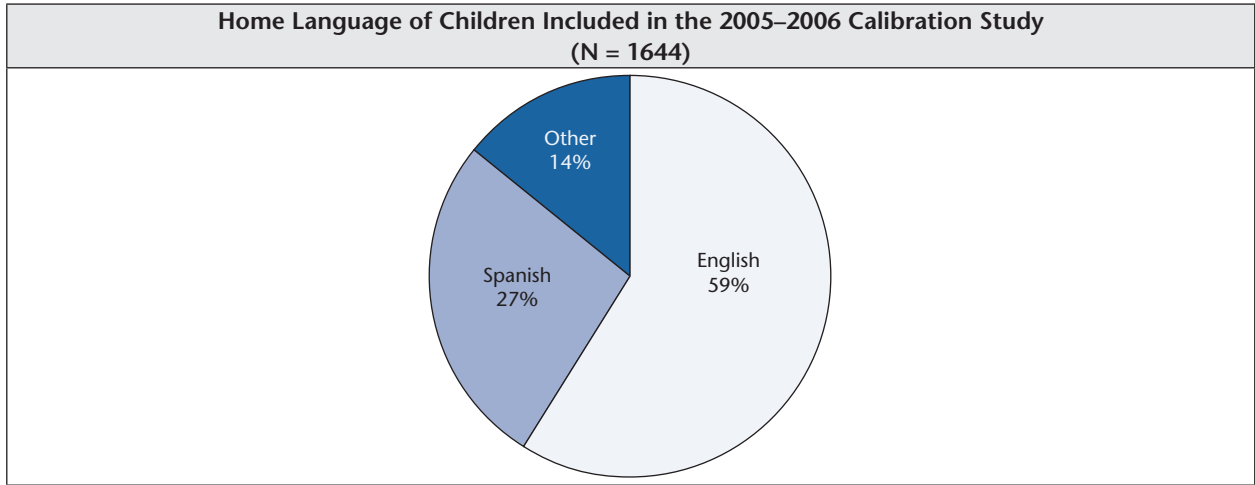
Local Education Agencies (LEAs) were selected throughout the state of California to participate in the study. Participating LEAs represented urban, suburban, and rural settings and included large, small, and moderately sized programs. Early childhood and early childhood special education programs were included. LEAs selected teachers and children who reflected the race and gender of their student populations. Teachers were trained to administer the DRDP *access* by staff of the Desired Results *access* Project. Teachers administered the DRDP *access* to the same set of children in the fall of 2005 and the spring of 2006. The DRDP *access* was administered to 1644 infants, toddlers, and preschoolers. After completing the first administration of the DRDP *access*, teachers completed an evaluation of the instrument.

Participants

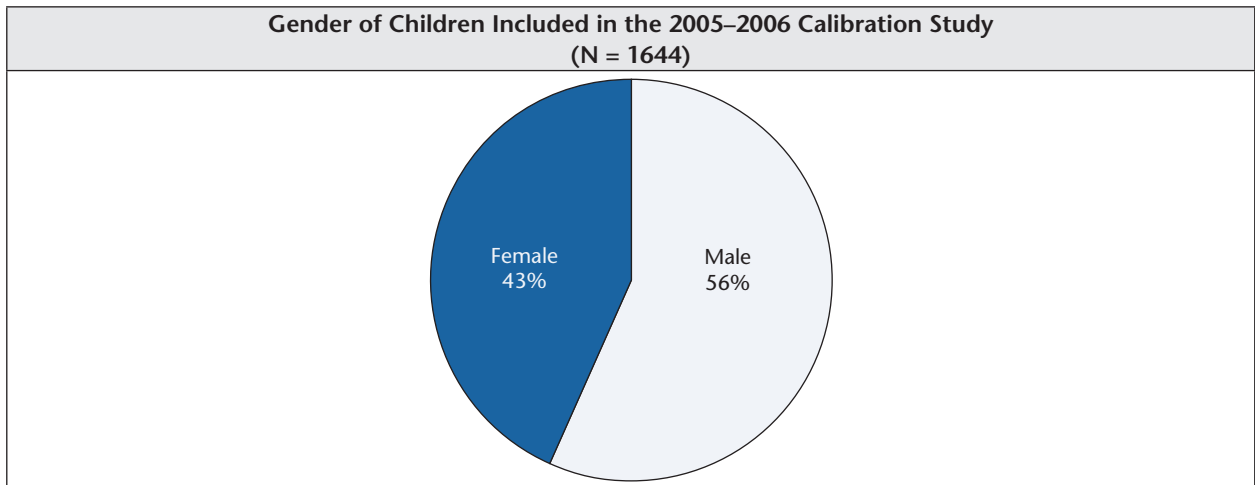
The ethnic representation of the sample matched the ethnic demographics of 3, 4, and 5 year olds in California.



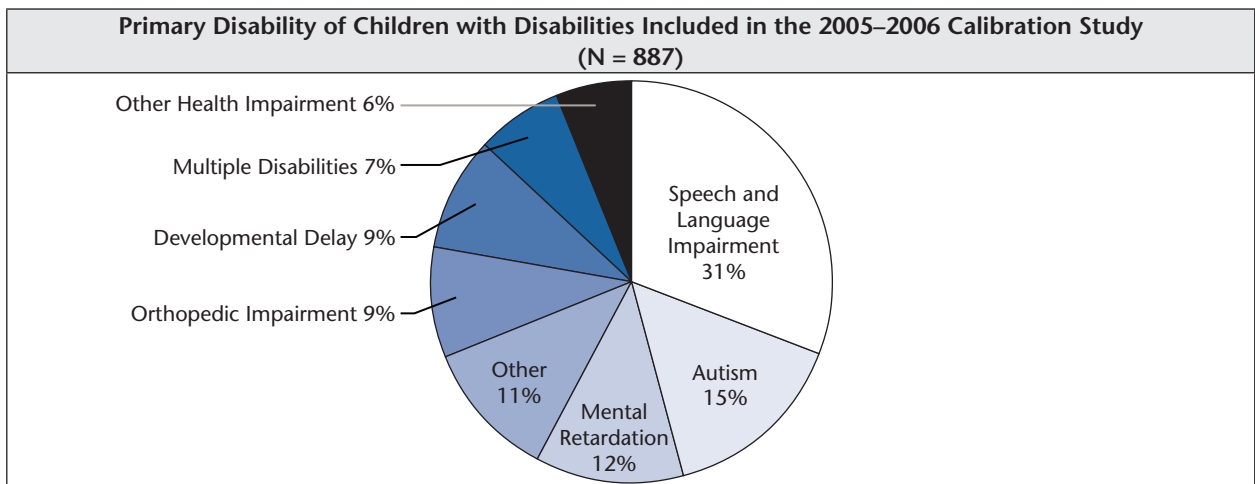
Most of the children in the sample were reported to use English as their home language. The next most frequent language was Spanish.



The sample had a slightly larger proportion of males.



Speech and Language Impairment was the most frequent disability in the sample. The sample is representative of the distribution of disabilities for early childhood special education programs.



Results

Reliability

Measures from the DRDP *access* are combined into Indicator Groups to create meaningful subscales. Scores from all subscales demonstrated sufficient reliability, according to standards in the field of statistics and measurement. The table below shows the range of values for each reliability statistic across Indicator Groups. All values are above the criterion value.

Reliability statistic	Range of values (across the 8 Indicator Groups)	Average (across the 8 Indicator Groups)	Criterion value
Person separability: how well children assessed with the DRDP <i>access</i> are differentiated on each subscale (Bond & Fox, 2001).	.90 - .96	.93	.80
Alpha: the degree to which children are rated in the same way across different Measures within each Indicator Group.	.88 - .97	.94	.80
Test-Retest (Pearson r): the stability of scores across time	.86 - .92	.90	.60

Validity

Scores from all subscales demonstrated sufficient validity, according to standards in the field of statistics and measurement.

Content validity. Content validity is the representativeness and relevance of the measures in an Indicator Group to the construct that the Indicator Group is intended to measure. Experts were convened during all stages of development of the DRDP *access* to ensure that the final instrument related closely to the constructs defined by each Indicator Group.

Discriminative validity. Discriminative validity describes how adequately the DRDP *access* differentiates between groups that theoretically should show differences. Scores from the Indicator Groups from the DRDP *access* were compared to scores from the same children on a measure of severity of disability. The two sets of scores ranked children similarly, supporting the discriminative validity of the instrument.

Construct validity. One aspect of the construct validity of scores is the degree to which the factor structure of the items is representative of and consistent with what is currently known regarding a construct (Riese, Waller, & Comrey, 2000). The construct validity of the DRDP *access* for measuring the eight Indicator Groups was assessed in two ways. The average difference between the scores predicted by the measurement model and the scores that were observed was estimated and showed sufficient support for the construct validity of the instrument. In addition, a factor analysis of scores from each Indicator Group supported the construct validity.

Utility

The utility of the DRDP *access* was assessed by surveying assessors about their experiences with the instrument. The results support that the DRDP *access* is easy to use and relevant to the material taught in preschool classrooms.

Reliability and Validity of the Desired Results Developmental Profile *access* (DRDP *access*): Results of the 2005–2006 Calibration Study

The DRDP *access*

The DRDP *access* is an authentic assessment rated by a child’s primary service provider using information gathered from multiple sources. The DRDP *access* includes 48 Measures (assessment items). Each Measure covers a specific aspect of early development (for example, expressive language) and describes a continuum of five to nine distinct levels of development within that area. Assessors rate each Measure by selecting the level of development that best describes the child’s skills, abilities, or knowledge based on the behaviors they observe, conversations with families, and information collected from other service providers and assessments.

Measures from the DRDP *access* are combined into groups to create meaningful subscales. The DRDP *access* was developed to measure three sets of latent variables (i.e., Indicator Groups, Office of Special Education Programs OSEP) Outcomes, and Desired Results): Each set of latent variables (e.g., Desired Results) organizes the 48 Measures (assessment items) included on the DRDP *access* into mutually exclusive item sets that are exhaustive of the 48 Measures included on the assessment (see Figure 1). The eight Indicator Groups are most useful for practitioners because they describe constructs that are relevant to classroom goals and can fit within many different curricula.

Eight Indicator Groups:

1. Self-Concept and Social and Interpersonal Skills
2. Self-Regulation
3. Language
4. Learning and Cognitive Competence
5. Math
6. Literacy
7. Motor Skills
8. Safety and Health

Three child outcomes linked to reporting requirements from the United States Department of Education, Office of Special Education Programs (OSEP):

1. Positive Social-Emotional Skills
2. Knowledge and Skills
3. Takes Appropriate Action to Meet Needs

Four Desired Results linked to California’s early childhood accountability system:

1. Children are Personally and Socially Competent
2. Children are Effective Learners
3. Children Show Physical and Motor Competence
4. Children are Safe and Healthy

OSEP Outcomes	Desired Results	Indicator Groups	Measures				
Positive social-emotional skills	Children are personally and socially competent	Self-Concept and Social and Interpersonal Skills	1 Identity of Self				
			2 Recognition of Own Skills and Accomplishments				
			3 Self-Expression				
			4 Expressions of Empathy				
			5 Interactions with Adults				
			6 Relationships with Familiar Adults				
			7 Developing Friendships				
			8 Building Cooperative Play with Other Children				
			9 Conflict Negotiation				
			10 Awareness of Diversity				
		Self-Regulation	11 Impulse Control				
			12 Seeking Other's Help to Regulate Self				
			13 Responsiveness to Other's Support				
			14 Self-Comforting				
			15 Taking Turns				
Knowledge and skills	Children are effective learners	Language	16 Language Comprehension				
			17 Responsiveness to Language				
			18 Expresses Self Through Language				
			19 Uses Language in Conversation				
Action to meet needs	Children show physical and motor competence	Learning and Cognitive Competence	20 Curiosity and Initiative				
			21 Attention Maintenance and Persistence				
			22 Memory				
			23 Cause and Effect				
			24 Engages in Problem-Solving				
			25 Object and Pretend Play				
			Safety and Health	Children are safe and healthy	Math	26 Number Sense: Understands Quantity and Counting	
						27 Number Sense: Math Operations	
						28 Number Sense: Comparison and Quantity	
	29 Shapes						
	30 Time						
	31 Classification and Matching						
	32 Measurement						
	33 Patterning						
	Motor Skills	Children show physical and motor competence				Literacy	34 Interest in Literacy
							35 Concepts of Print
			36 Letter and Word Knowledge				
			37 Phonological Awareness				
			38 Emerging Writing				
Safety and Health	Children are safe and healthy	Motor Skills	39 Comprehension of Text				
			40 Movement				
			41 Balance				
			42 Grasp/Release and Manipulation				
			43 Eye-Hand Coordination				
			Safety and Health	Children are safe and healthy	Safety and Health	44 Personal Care Routines: Toileting and Hygiene	
						45 Personal Care Routines: Dressing	
						46 Personal Care Routines: Self-Feeding	
						47 Personal Safety	
						48 Understanding Healthy Lifestyle: Eating and Nutrition	

Figure 1. Map showing the relationship of the three sets of subscales to the 48 Measures on the DRDP *access*.

Figure 1 shows a map of each of the three sets of subscales to the 48 items on the DRDP *access*. This technical report will focus on the reliability and validity of scores in the eight Indicator Group metric.

DRDP *access* Measurement Model

The measurement model used with the DRDP *access* is based on the work of George Rasch (1960, 1966) and is typically referred to as the Rasch measurement model. The potential utility of Rasch measurement models in early childhood assessment has been recognized for many years (Snyder & Sheehan, 1992), and their usefulness in assessment for early childhood accountability systems has recently been highlighted (Meisels, 2006). The Rasch measurement model is probabilistic, that is statistical estimates are made to address the following question: “When a child with a certain ‘ability’ level is observed using the DRDP *access* what is the probability or likelihood that the child will get credit for an item at a particular level on a DRDP *access* Measure?” The probability of getting credit at a particular level is based on the difference between the “ability” of the child and the difficulty level of the item.

A mathematical equation is used to define the probabilistic relationship between item difficulty and person ability. Equation 1 is used to define the probability that a child would score at a particular level of a Measure on the DRDP *access*.

$$P(\theta) = \frac{e^{\theta-B}}{1 + e^{\theta-B}} \quad (1)$$

Equation 1 states that the probability that a person with ability (θ) will pass an item with difficulty B is equal to 2.72 (e) raised to the power of the person’s ability minus the difficulty of the item divided by 1 + 2.72 raised to the power of the person’s ability minus the difficulty of the item (B). The probability that a person will pass an item is also referred to as the predicted score for that person. For example, if a person had ability of 1 and the item difficulty was 4, the equation could be solved $(e^{(1-4)} / (1 + e^{(1-4)})) = 0.05$. This result means that a person with an ability of 1 has a 5% chance of passing an item with a difficulty of 4 or a predicted score of .05. What would be the probability of a person with an ability of 6 passing the same item? Using the equation results in $(e^{(6-4)} / (1 + e^{(6-4)})) = .88$. This result means that a person with ability of 6 has an 88% chance of passing an item with a difficulty of 4 or a predicted score of .88. This relationship between item difficulty, person ability, and probability of passing an item is demonstrated in Figure 2. The difficulty of the item in Figure 2 is 0. Note that as ability increases the probability of correctly answering the item also increases.

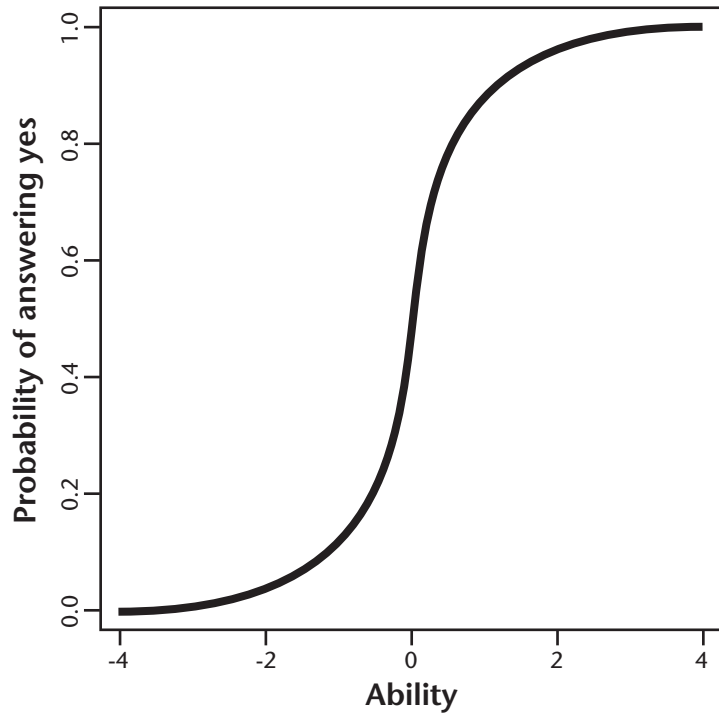


Figure 2. The item characteristic curve for an item with a difficulty of 0.

Calibration Study (2005–2006)

Sample description. In order to scale the items on the DRDP *access* a calibration study was conducted beginning in the fall of 2005. Sampling was conducted at the level of the Local Education Agency (LEA). Participating districts represented urban, suburban, and rural settings and included large, small, and moderately-sized programs. LEA's were instructed to select primary service providers and students who were reflective of the race and gender of their student populations. The DRDP *access* was administered to 1644 infants, toddlers, and preschoolers in the state of California for the purpose of defining the ability and item parameters and determining the fit of the observed data to the Rasch model. The participants included both typically developing children and children with disabilities. The scores from this sample were also used to calculate traditional reliability and validity coefficients. See Table 1 for the sample demographics.

Table 1. Sample Demographics for the 2005–2006 Calibration Study.

Demographic variable	Children with disabilities (n = 887)	Typically developing children (n = 757)
	Percent in each row	Percent in each row
Ethnicity		
Hispanic/Latino	43.8	55.0
Caucasian/White	33.6	25.0
Asian	8.4	6.7
African American/Black	6.4	7.4
Other	2.4	5.9
Home language		
English	62.6	52.7
Spanish	23.5	30.8
Other	12.9	15.3
Gender		
Male	64.9	48.6
Female	35.1	51.4
Disability		
Speech	30.6	—
Autism	14.6	—
Mental Retardation	12.2	—
Orthopedic Impairment	9.2	—
Developmental Delay	8.6	—
Multiple Disabilities	7.0	—
Other Health Impaired	6.1	—
Other	11.4	—
Age in months	95% CI around the mean 41.47–43.43	95% CI around the mean 50.13–51.49

Established Reliability

Reliability refers to the stability of subscale score estimates across Measures, contexts, people, and time. Three different techniques were used to test the reliability of the Indicator Group subscales. Alpha and person separation were computed using the entire sample of 1644 infants, toddlers, and preschoolers. Alpha is used to estimate the internal consistency of a test (Cronbach, 1951). Internal consistency is the extent to which children are rated in the same way across different Measures within each Indicator Group. Alpha scores above .8 are indicative of good internal consistency. As shown in Table 2, the internal consistency of item scores in the DRDP *access* was very high (.92 – .97). How well children who were assessed with the DRDP *access* could be differentiated on each subscale or person separation was then estimated (Bond & Fox, 2001). The results were also high for person separation (.90 – .96) as shown in Table 2. Next, the stability of scores across time

was examined by calculating the correlation between scores in the same group of children given at two different time points. This analysis included only children with disabilities although we are currently collecting similar information on typically developing children. As part of the 2005–2006 calibration study, 707 children with disabilities were assessed at two time points (fall, 2005 and spring, 2006). The mean length of time between the two assessments was 5.5 months (minimum = 4 months; maximum = 8 months). Test-retest score reliability was also excellent ($r=.86$ – $r=.92$). The results of all reliability calculations are included in Table 2.

Table 2. Reliability Statistics for the DRDP *access*.

Reliability Statistics Calculated from the 2005–2006 Calibration Sample			
Indicator Group	Alpha	Person separation	Test-Retest reliability
Self-Concept and Social and Interpersonal Skills	.97	.96	.91
Self-Regulation	.88	.90	.86
Language	.96	.93	.90
Learning and Cognitive Competence	.96	.95	.90
Math	.95	.95	.91
Literacy	.93	.94	.92
Motor Skills	.96	.92	.92
Safety and Health	.92	.90	.90

Content Validity

Content validity describes the degree to which the Measures and subscales of the DRDP *access* are representative of important developmental domains for young children. There were several iterative steps to ensure the content validity of the DRDP *access*. Experts were convened to develop the items on the DRDP *access* from the Measures on the Infant/Toddler DRDP and the Preschool DRDP. This process resulted in the version of the DRDP *access* that was used for calibration. When the item analysis of the calibration data was complete, a panel of both content and analysis experts was convened to review the item level data. Panel participants reviewed both the item analysis and the transcripts of comments that were provided by the assessors. All participants were trained in the Rasch measurement model expectations. When the analysis identified deviations from model expectations, these were discussed in the context of the items and comments from the assessors. Deviations from the Rasch model that were reviewed include the following: 1) weighted mean square outside of the defined parameters, 2) unevenly spaced thresholds, 3) levels with an insufficient number of children to establish stable parameter estimates, and 4) Measures that assessors were unable to rate because of the severity of a child’s disability. This information was used to revise the Measures on the DRDP *access* to ensure that they were representative of important developmental domains in early childhood and thus content valid.

Discriminative Validity

Discriminative validity describes how adequately the DRDP *access* differentiates between groups that theoretically should show differences. The ABILITIES Index (Simeonsson & Bailey, 1991) was completed in addition to the DRDP *access* for children with disabilities in the calibration study. The ABILITIES Index is a functional measure designed to evaluate the severity of disability across several domains of functioning (e.g., audition, behavior, intelligence, language). Higher scores on the ABILITIES index reflect more significant impairments or disabilities. Higher scores on the DRDP *access* reflect more advanced development with respect to an Indicator Group and less severe disabilities. Theoretically, one would expect that the DRDP *access* would discriminate children with more severe disabilities from children with less severe disabilities. This expectation was tested by calculating the correlation between scores on the ABILITIES Index and each of the subscales of the DRDP *access*. A strong negative correlation would support the discriminative validity of a subscale.

The ABILITIES Index and the DRDP *access* were administered to a sample of 396 children with disabilities during the fall of 2005 and the spring of 2006. The ABILITIES Index scores were averaged across the two administrations. The averaged ABILITIES index was correlated with the Time 1 DRDP *access* Indicator Group subscale scores. The obtained Pearson-product moment correlation coefficients are presented in Table 3 and ranged from $-.697$ to $-.726$. These correlations support the discriminative validity of the subscale scores.

Table 3. Discriminate Validity of the DRDP *access*.

Correlations between the averaged total score on the ABILITIES Index and time 1 scores on the DRDP <i>access</i>	
Indicator Group	ABILITIES Index score
	Pearson correlation
Self-Concept and Social and Interpersonal Skills	$-.726$
Self-Regulation	$-.720$
Language	$-.726$
Learning and Cognitive Competence	$-.723$
Math	$-.668$
Literacy	$-.691$
Motor Skills	$-.704$
Safety and Health	$-.697$

Construct Validity

One aspect of the construct validity of scores is the degree to which the factor structure of the items is representative of and consistent with what is currently known regarding a construct (Riese, Waller, & Comrey, 2000). The Measures on the DRDP *access* were assigned to the eight Indicator Groups based on the relation between the content of the Measure and what is known about the Indicator. For example, the *Movement* Measure was

assigned to the *Motor Skills* Indicator because how a child moves is an important piece of information about his or her motor skills. Two methods were used to determine how well the Measures assigned to each Indicator “stick together.” The first method was factor extraction using principle components analysis. The second method was the calculation of weighted mean square for each Measure in each Indicator Group.

Factor extraction. Factor extraction was used to identify the number of factors that were needed to accurately account for the common variance between the Measures on each of the eight subscales. To ensure sample homogeneity only scores collected on children with disabilities in the calibration sample were used. Some children with disabilities were excluded from the analysis due to missing data. The final sample for the factor analysis included 708 children with disabilities. The number of factors needed to accurately account for common variance between Measures was determined using a scree plot augmented by parallel analysis. The parallel analysis was conducted by creating 1000 sets of random data simulating each of the eight Indicator Groups. Factors were extracted from each set using principle components analysis. The eigenvalues for the random data were averaged across the 1000 data sets to get a best estimate of the eigenvalue associated with each of the factors.

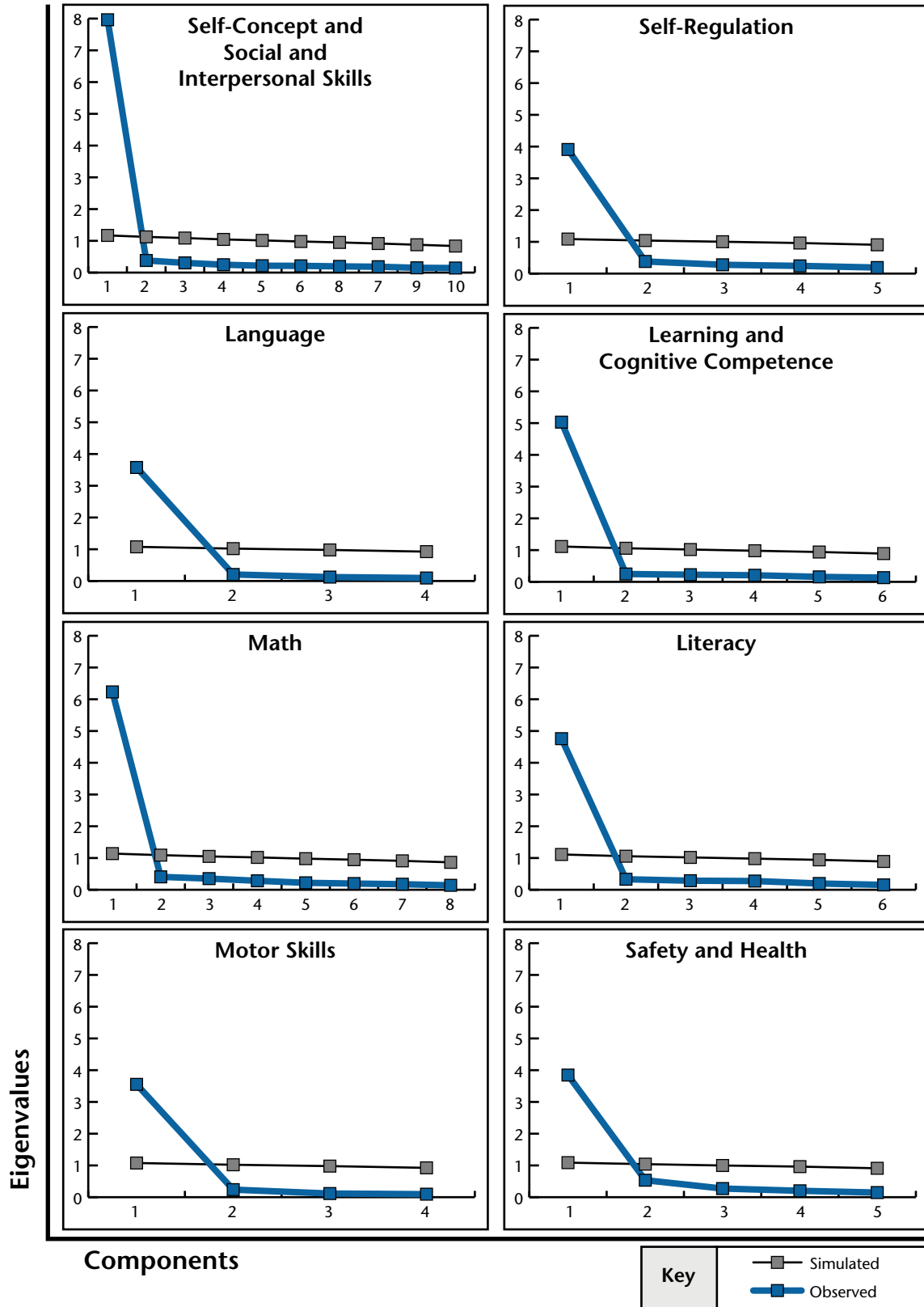


Figure 3. Plot of eigenvalues for each of the eight Indicator Groups.

The subscale met the criteria for unidimensionality if there was a visible “elbow” between the first and third factor of the observed data and if all components except the first component had eigenvalues lower than the corresponding eigenvalue from the random data (Reise, Waller, & Comrey; 2000). See Figure 3 for each of the eight scree plots. Each of the eight Indicator Groups met the criteria for unidimensionality.

Mean square analysis. As part of the Rasch analysis used to calibrate the instrument, the unidimensionality of each of the eight Indicator Groups was estimated using the weighted mean square. The weighted mean square is an estimate of the fit of the observed scores to the scores predicted by the measurement model. The scores are predicted for each Measure based on a transformation of the child’s total score from the Indicator Group that the Measure belongs to. For example, a child’s rating for the *Movement* Measure would be predicted based on his or her combined score from all of the *Motor Skills* Measures. The measurement model used for the DRDP *access* assumes that each subscale is unidimensional. If the observed data fit the predicted data it is assumed that the assessment meets the assumptions of the measurement model, including the assumption of unidimensionality.

The mean square for each item is the difference between observed and predicted scores squared and summed across people (Wright & Masters, 1982). For the current analysis, the weighted mean square was used. The weighted mean square is weighted by the distance of the test-taker’s ability from the mean of the sample. Scores that are very far from the mean of the sample are given less weight to reduce the impact of outliers on the test statistics.

Figure 4 shows the item difficulty and weighted mean square for each Measure on the DRDP *access* scaled using Indicator Groups as the subscales. Item misfit is defined as a weighted mean square greater than 1.33 or less than .73. For example, the *Eating and Nutrition* Measure met the criteria for misfit with a weighted mean square of 1.45 and a T-score of 10.9. The fit of the majority of Measures to the Indicator Group subscales is evidence that this model correctly classifies score variance.

		Item Difficulty	Weighted MS
Outcome 1: Desired Result 1: Indicator 1	1 Identity of Self	-0.647	0.84
	2 Recognition of Own Skills and Accomplishments	-0.721	0.91
	3 Self-Expression	-0.498	0.89
	4 Expressions of Empathy	0.921	1.03
	5 Interactions with Adults	-1.200	0.85
	6 Relationships with Familiar Adults	-0.899	0.88
	7 Developing Friendships	-0.364	0.89
	8 Building Cooperative Play with Other Children	0.155	1.01
	9 Conflict Negotiation	0.819	0.77
	10 Awareness of Diversity	0.570	0.79
Indicator 2	11 Impulse Control	-0.433	1.07
	12 Seeking Other's Help to Regulate Self	-1.803	0.96
	13 Responsiveness to Other's Support	-0.704	0.91
	14 Self-Comforting	-0.666	1.21
	15 Taking Turns	0.730	0.84
Outcome 2: Desired Result 1: Indicator 3	16 Language Comprehension	-0.874	0.94
	17 Responsiveness to Language	-0.578	0.81
	18 Expresses Self Through Language	-0.419	0.77
	19 Uses Language in Conversation	0.168	0.77
Desired Result 2: Indicator 4	20 Curiosity and Initiative	-0.340	0.85
	21 Attention Maintenance and Persistence	-0.092	1.07
	22 Memory	-0.348	0.77
	23 Cause and Effect	-0.577	0.88
	24 Engages in Problem Solving	-0.255	0.89
	25 Object and Pretend Play	-0.292	0.92
Indicator 5	26 Understands Quantity and Counting	0.708	0.92
	27 Math Operations	1.613	0.89
	28 Comparison of Quantity	1.324	0.74
	29 Shapes	0.826	0.73
	30 Time	0.249	0.84
Indicator 6	31 Classification and Matching	1.836	0.89
	32 Measurement	1.574	0.89
	33 Patterning	0.441	0.81
	34 Interest in Literacy	0.250	1.05
	35 Concepts of Print	1.222	0.66
	36 Letter and Word Knowledge	0.298	0.9
	37 Phonological Awareness	1.459	0.96
	38 Emerging Writing	0.934	0.92
Outcome 3: Desired Result 3: Indicator 7	39 Comprehension of Text	1.408	0.75
	40 Movement	-1.264	0.98
	41 Balance	-1.259	0.99
	42 Grasp/Release and Manipulation	-0.700	0.77
Outcome 3: Desired Result 4: Indicator 8	43 Eye-Hand Coordination	-0.943	0.85
	44 Toileting and Hygiene	-0.147	0.77
	45 Dressing	-0.088	0.82
	46 Self-Feeding	-1.156	0.79
	47 Personal Safety	0.314	0.8
	48 Eating and Nutrition	-0.551	1.45

Figure 4. Weighted mean square for each of the Measures on the DRDP *access*.
Note. Indicator 1 = Self-Concept and Social and Interpersonal Skills; Indicator 2 = Self-Regulation;
Indicator 3 = Language; Indicator 4 = Learning and Cognitive Competence; Indicator 5 = Math;
Indicator 6 = Literacy; Indicator 7 = Motor Skills; and Indicator 8 = Safety and Health.

Sensitivity

Because the DRDP *access* will be used to measure change in the subscales across time it is important that scores on each of the subscales change over time. As part of the 2005 – 2006 calibration study, 707 children with disabilities were assessed at two time points (fall, 2005 and spring, 2006). The mean length of time between the two assessments was 5.5 months (minimum = 4 months; maximum = 8 months). To test whether there was change in the scores across time, the mean difference between the Time 1 and Time 2 scores was examined and a t-statistic was calculated to measure the significance of the mean difference. The paired-t comparisons of children’s scores at these two time points are shown in Table 4. All t-statistics are statistically significant at the .001 level, and all have large effect sizes (Cohen, 1988).

Table 4. Mean Difference by Indicator Group Between Two Assessment Time Points.

Indicator Group	Mean difference (T2 – T1)	SD	95% Confidence interval		Paired-t	Cohen’s d
			Lower bound	Upper bound		
Self-Concept and Social and Interpersonal Skills	9.9	10.7	9.1	10.7	24.6	1.8
Self-Regulation	10.4	13.2	9.4	11.3	21.0	1.6
Language	9.1	11.1	8.3	10.0	21.9	1.6
Learning and Cognitive Competence	10.4	11.4	9.6	11.2	24.2	1.8
Math	10.2	11.2	9.4	11.0	24.3	1.8
Literacy	9.4	10.2	8.6	10.2	24.4	1.8
Motor Skills	9.8	10.5	9.0	10.5	24.6	1.9
Safety and Health	8.7	11.4	7.9	9.6	20.3	1.5

Usability

The practical utility for the DRDP *access* has been established through multiple focus groups and surveys of assessors. The most recent written survey was conducted with 204 assessors who administered the DRDP *access* in the 2005–2006 calibration study. The survey consisted of 39 items that combined rating scales, forced choice, and open-ended items.

The 95% confidence intervals around the mean rating for items on the questionnaire addressing acceptability are presented in Table 5. Most assessors (85.7%) agreed that the instrument measured the things that they teach the children in their caseload. Most assessors (60.1%) agreed that the DRDP *access* would capture children’s progress if it were administered in the fall and again in the spring.

Table 5. Descriptive Statistics About the Practical Utility of the DRDP *access*.

Item	95% CI around mean 1 = Strongly disagree, 10 = Strongly agree
The Measures were meaningful	7.39 – 7.97
The meaning of Descriptors was clear	7.62 – 8.12
The Examples were clear	7.58 – 8.14
The overall structure of the instrument was easy to understand	8.01 – 9.27
The format/layout was easy to use	8.22 – 8.52
The Descriptors were easy to observe during the daily routine of our program	7.26 – 7.88
The DRDP <i>access</i> accurately captured the Developmental Levels for the children observed	6.73 – 7.07

In addition to these data, the practical utility of the DRDP *access* has been evaluated using iterative processes throughout its development and initial calibration. Earlier versions of the instrument have been reviewed by early childhood special education providers; researchers; faculty with expertise in development, as well as, cultural, linguistic, and ability diversity; and families of young children with disabilities. This stakeholder input informed the development of the Measures, layout of the instrument, examples used for items on every Measure, scoring, as well as, the development of support materials.

In conclusion, the scores from the DRDP *access* are reliable and valid for the measurement of developmental growth in young children with disabilities and typically developing children between the ages of birth to five years. The Desired Results *access* Project continues to assess the validity and reliability of the assessment. The project is currently testing the equivalence of item parameters across important subgroups of the calibration sample. The project is also collecting longitudinal data on typically developing children to better specify norming parameters. Please contact the Desired Results *access* Project at info@draccess.org if you are interested in participating in future studies using the DRDP *access* or have suggestions for types of validity and reliability that should be tested.

References

- Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, New Jersey: Lawrence Erlbaum.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Desired Results access Project (2008, January). *Understanding the DRDP access Measurement Model*. Retrieved March 20, 2008, from <http://www.draccess.org/assessors/UnderstandingDRDPaccessMM.html>
- Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16(3), 297–334.
- Meisels, S. (2006, March). *Accountability in early childhood: No easy answers* (Occasional Paper No. 6). Chicago: Erikson Institute, Herr Research Center for Children and Social Policy.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49–57.
- Reise, S.P., Waller, N.G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12, 287–297.
- Simeonsson, R.J., & Bailey, D.B. (1991). *The ABILITIES Index*. Retrieved April 1, 2005 from <http://www.fpg.unc.edu/~publicationsoffice/fpgpdfs/AbilitiesIndex.pdf>
- Snyder, S., & Sheehan, R. (1992). The Rasch measurement model: An introduction. *Journal of Early Intervention*, 16(1), 87–95.
- Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.

Appendix: Glossary

Term	Definition
Ability	Ability is the term used to describe scaled scores on an Indicator Group. Scale scores are raw scores that have been transformed to a new “scale” so that equal intervals are maintained between each unit on the scale.
Calibrations on the DRDP <i>access</i>	Raw scores from the DRDP <i>access</i> subscales are totaled and transformed to maintain an equal distance between each point on the scale. These transformed scores are called calibrations. Raw subscale scores are computed by summing the ratings for all of the Measures on a subscale. This total is transformed into a natural logarithm (Bond & Fox, 2001). This transformation creates an equal distance between each of the total scores. For the DRDP <i>access</i> the mean scale score for each Indicator Group is 200 and the range is from 100-300. Children’s scores are reported in units of one on either side of the mean.
Item difficulty	Item difficulty is the ability or scale score required to have a 50% chance of passing an item. Item difficulty parameters are used to predict how a person will score on an item. The correctness of these predictions is measured by the magnitude of the person and item residuals. Person residuals are the difference between predicted and observed scores across all of the items administered to that person. Item residuals are the difference between predicted and observed scores across all of the people that were administered the item.
Weighted mean square	The weighted mean square is a way of summarizing the residuals. Because of an interest in the magnitude of residuals as opposed to the direction (positive vs. negative), each of the residuals was squared before the residuals were added together. To summarize item residuals across people, the residuals are summed across all of the people. The summed residuals are then divided by the observed variance of the residuals. An item with a weighted mean square greater than 1.33 or lower than .73 is considered misfit. The criteria for the weighted mean square were selected because the items are partial credit and the DRDP <i>access</i> needs to be very sensitive to change (Bond & Fox, 2001).

The Desired Results *access* Project
A Special Project of the Napa County Office of Education

www.draccess.org
info@www.draccess.org

© 2008 California Department of Education, Special Education Division

