DRDP-R©

Technical Report

©2011 by the California Department of Education, Child Development Division
All rights reserved. Permission to reproduce only for instructional purposes.

Acknowledgments
In addition to the BEAR Center staff, the following groups contributed to the development of DRDP-R:

- California Department of Education, Child Development Division
- WestEd Center for Child and Family Studies
- American Institutes for Research
- Mathematica Policy Research
- Sonoma State University – CIHS

Table of Contents

Tables

Figures

Preface

This technical report provides methodological background and psychometric analysis of the Desired Results Developmental Profile – Revised (DRDP-R). The report is intended for those who seek theoretical, technical, and empirical information about the development and validation of the DRDP-R.

The DRDP-R is an embedded observational assessment system designed to track children's progress from birth through age 12, in programs funded by the State of California. The DRDP-R assessment system provides a framework for educational practitioners (e.g., infant/toddler caregivers, pre-school teachers, and before/after school teachers) to observe, document and rate children's development in socio-emotional, cognitive, physical, and behavioral domains. The DRDP-R instruments are designed to collect evidence of children's development using multiple measures of progress in three age groups: Infant/Toddlers, Preschoolers, and School-Age children from kindergarten through 12 years. The DRDP-R provides practitioners, policy makers, and researchers with information about children's socio-emotional, cognitive, physical, and behavioral development.

The first chapter provides an introduction to the DRDP-R, its purpose, design process and format. Chapter 2 discusses the aspects of content validity evidence upon which the DRDP-R was developed. Chapter 3 describes the rationale and methodology of the 2005 Calibration Study of the DRDP-R. Chapter 4 describes the results of the calibration study. Chapter 5 reports on the internal construct validity of the DRDP-R. Chapter 6 describes the Special Study I, which is an analysis of the linkage structure of the DRDP-R instruments across age groups. Chapter 7 discusses the Special Study II, which is an examination of the consistency of ratings across primary and secondary raters. Chapter 8 provides a general discussion of the results. The appendices include examples of the DRDP-R measures, supplemental figures, and tables.

1 Introduction to the Desired Results Developmental Profile – Revised (DRDP-R)

1.1 Overview of DRDP Assessment

In 2000, the California Department of Education (CDE) implemented the *Desired Results Developmental Profile*[1] as the assessment component of the *Desired Results System*[2] for outcomes-based accountability in CDE's state-funded center-based care programs. Concurrently, CDE's Child Development Division (CDD) undertook a series of research studies for the purpose of investigating the validity, reliability, and effectiveness of DRDP as an assessment tool. These studies were a combined effort among CDD staff; researchers and authors in the fields of early-childhood education and developmental psychology; expert practitioners in Infant/Toddler, Preschool, and School-Age care programs; and experts in psychometric measurement and assessment. This series of studies examined DRDP, collected data and feedback from the field, and revised DRDP on the basis of what was learned at each successive phase— the 2002 Pilot Test, the 2003 Field Test, and the 2005 Calibration Study. This process culminated in Fall 2006 with the implementation of a revised, updated, and improved version of DRDP, called DRDP-R (Desired Results Developmental Profile – Revised). For details of the overall Desired Results System and a manual for how the DRDP-R is used, please see *The DRDP-R Users Guide*[3] (CDE, 2008).

From the beginning, the instruments of the original DRDP and of the newer DRDP-R have been intended and designed to measure developmental outcomes at the level of the

---

[1] http://www.cde.ca.gov/sp/cd/ci/documents/drdprps.pdf
[2] http://www.cde.ca.gov/sp/cd/ci/desiredresults.asp
[3]http://www.cde.ca.gov/sp/cd/ci/documents/drdprguide08.doc

individual child based on observations by education professionals. The primary purpose of DRDP assessment results is to provide program staff with information they need to inform and support their decisions on how to make improvements for individual children, classrooms, and programs. DRDP-R results are aggregated by groups before they are shared with CDD. Aggregating the DRDP-R results that are shared outside the program maintains confidentiality for children and teachers[4], and also ensures the level of autonomy necessary for programs to operate effectively, while still providing policy makers with useful information about trends across the overall population served by CDD and CDE.

Further revision and updating of DRDP assessment is an ongoing process to which CDD continues to commit resources. One major motivating factor behind this revision process is CDE's development and implementation of Infant/Toddler (IT) and Preschool (PS) Development and Learning Foundations. Currently, CDE has implemented IT and PS foundations in the areas of Language/Literacy, Mathematics, and Social-Emotional Development; and PS foundations in English Language Development for children who are English learners. Further sets of Foundations will be released in the future. As they are released, DRDP-R will be aligned to these sets of foundations, hence new versions of the DRDP can be expected.

Additional references, support materials, and resources developed for DRDP-R and the Desired Results System as a whole are available through this link: http://www.cde.ca.gov/sp/cd/ci/drdpforms.asp

The Child Development Division of the California Department of Education and the BEAR Center at UC Berkeley have also developed DRDPtech, a computer support system for

4 For simplicity, the term "teachers" is used to refer to anyone who uses the DRDP-R, such as, care - givers, pre-school teachers, youth center staff, administrators, etc.

DRDP-R assessment. When completed, DRDPtech will automatically compile, analyze, and archive results, as well as generate reports to inform program-improvement decisions, share assessment information with parents, and facilitate collaboration between the programs and their Field Services consultants.

1.2 Development of the DRDP-R

The purpose of this technical report is to provide details of the process of developing DRDP-R as a psychometrically valid and reliable tool for measuring developmental outcomes of individual children. Because of the way DRDP-R and its results are intended to be used, however, the process of developing DRDP-R could not solely be an exercise in psychometrics. DRDP-R had to be developed through an intensive partnership of psychometricians with experts in child development—researchers, theoreticians, and program practitioners—as well as with CDD staff.

The design of DRDP-R, like the original DRDP before it, is based on the assumption that a teacher's observations of a child who the teacher knows well are an important and useful resource. That is, DRDP assessment is embedded in ongoing program activities, rather than being tied to separate assessment activities that are undertaken for the purpose of 'direct' assessment. Also like the original DRDP, the design of DRDP-R derives from the four basic components of the Desired Results (DR) system for children:

DR1: Children are personally and socially competent;

DR2: Children are effective learners;

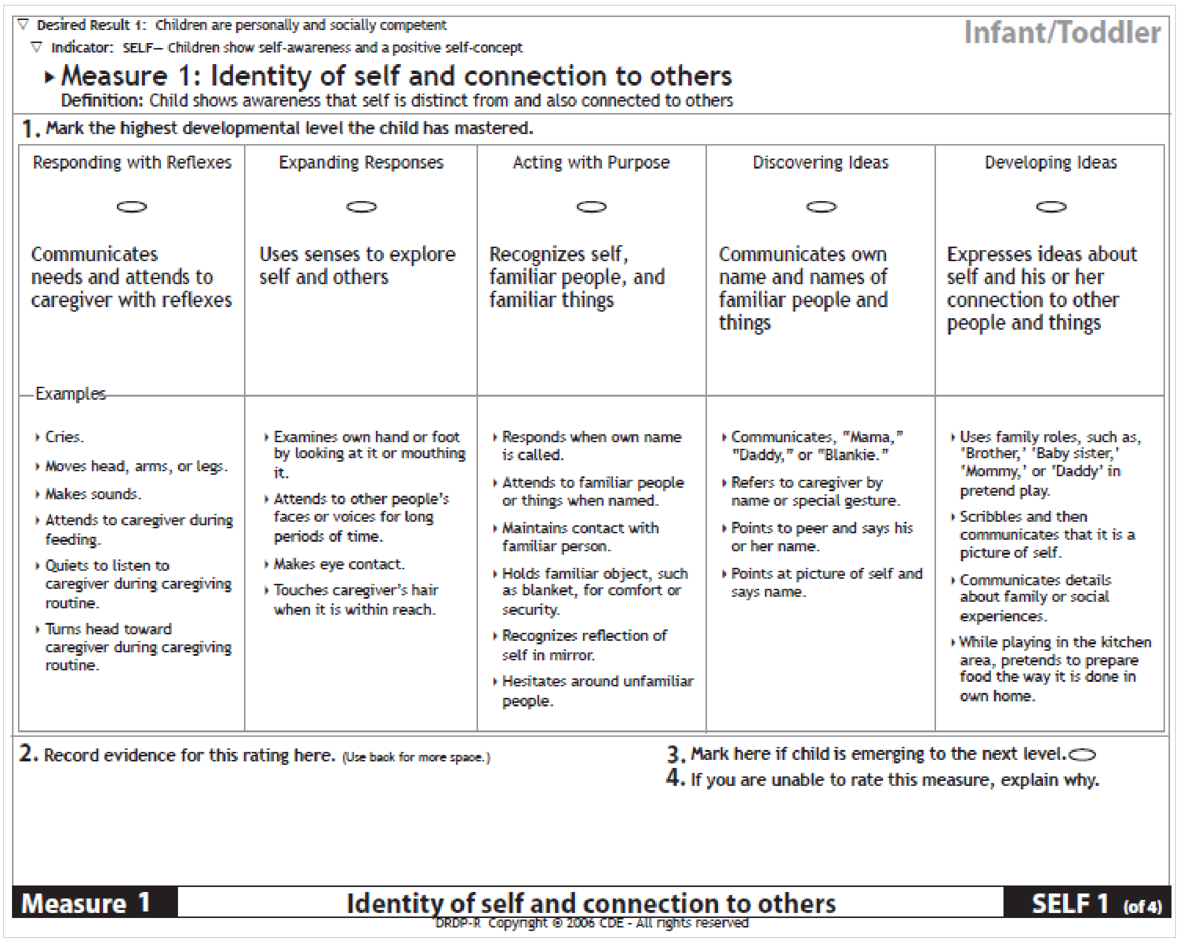
DR3: Children show physical and motor competence;

DR4: Children are safe and healthy.

Despite these important aspects in common, DRDP-R is the result of a re-conceptualization of DRDP assessment and hence has a structure and format that is quite different from the original DRDP. The DRDP-R structure was designed to increase the interpretability and meaningfulness of the assessment results and also to improve the accuracy of ratings. The starting point for the team who developed DRDP-R was to examine the sets of specific individual child behaviors (covering all four of the DRs shown above) that the original DRDP instruments provided for teachers to observe. In the original DRDP, teacher observations were recorded for the various individual behaviors in terms of how well a child had mastered each one. That is, for each one of the various behaviors provided in the original DRDP instruments, the teacher was asked to mark whether the child had 'not yet mastered' the behavior, 'partially mastered' the behavior, mastery was 'emerging' for the behavior, or the behavior was 'fully mastered'. See Appendix A for an example page from the original DRDP.

The first step taken in developing DRDP-R was to note which of the behaviors provided for teacher observation in the original DRDP were associated with which of the four DRs. The next step, based on a thorough literature review, and on excellent professional practice, was to identify groups of behaviors that might constitute distinct dimensions of development within a given DR. For example, within DR2 ("Children are effective learners"), behaviors provided in the original DRDP were grouped into early literacy, early mathematics, and overall cognitive development. Borrowing terminology from the original DRDP, these groupings were viewed as indicators of progress toward a given DR, and, hence, were called Indicators for short.

The main difference between the original DRDP and DRDP-R is in the way behaviors are organized and provided to teachers for observation. In both, sets of behaviors are grouped into indicators within DRs. In the original, these constitute individual behaviors that stand alone. A teacher rates a child by marking the extent to which the child has mastered each one. In DRDP-R, however, individual measures of development are presented and a teacher rates a child by marking which of several successive levels of development on that measure the child has mastered. Each Measure is its own page of the DRDP-R. See Figure 1.1 for an example page from DRDP-R. Example pages from the original and revised DRDP formats are presented in Appendixes A and B, respectively.



▽ Desired Result 1: Children are personally and socially competent

▽ Indicator: SELF— Children show self-awareness and a positive self-concept

Infant/Toddler

▸ Measure 1: Identity of self and connection to others

Definition: Child shows awareness that self is distinct from and also connected to others

1. Mark the highest developmental level the child has mastered.

Responding with Reflexes	Expanding Responses	Acting with Purpose	Discovering Ideas	Developing Ideas
⬭	⬭	⬭	⬭	⬭
Communicates needs and attends to caregiver with reflexes	Uses senses to explore self and others	Recognizes self, familiar people, and familiar things	Communicates own name and names of familiar people and things	Expresses ideas about self and his or her connection to other people and things
Examples				
▸ Cries. ▸ Moves head, arms, or legs. ▸ Makes sounds. ▸ Attends to caregiver during feeding. ▸ Quiets to listen to caregiver during caregiving routine. ▸ Turns head toward caregiver during caregiving routine.	▸ Examines own hand or foot by looking at it or mouthing it. ▸ Attends to other people's faces or voices for long periods of time. ▸ Makes eye contact. ▸ Touches caregiver's hair when it is within reach.	▸ Responds when own name is called. ▸ Attends to familiar people or things when named. ▸ Maintains contact with familiar person. ▸ Holds familiar object, such as blanket, for comfort or security. ▸ Recognizes reflection of self in mirror. ▸ Hesitates around unfamiliar people.	▸ Communicates, "Mama," "Daddy," or "Blankie." ▸ Refers to caregiver by name or special gesture. ▸ Points to peer and says his or her name. ▸ Points at picture of self and says name.	▸ Uses family roles, such as, 'Brother,' 'Baby sister,' 'Mommy,' or 'Daddy' in pretend play. ▸ Scribbles and then communicates that it is a picture of self. ▸ Communicates details about family or social experiences. ▸ While playing in the kitchen area, pretends to prepare food the way it is done in own home.

2. Record evidence for this rating here. (Use back for more space.)

3. Mark here if child is emerging to the next level. ⬭

4. If you are unable to rate this measure, explain why.

Measure 1 | Identity of self and connection to others | SELF 1 (of 4)

DRDP-R Copyright © 2006 CDE - All rights reserved

Figure 1.1—Example page from the Infant/Toddler DRDP-R.

Perhaps the most important reason for making this change is that it avoids ratings based on incompletely defined levels of mastery that do not necessarily mean the same thing to all raters. By providing distinct descriptors of successive levels of development within a given dimension and asking which is the highest level mastered according to a clear definition of what mastery is, a rater can more readily understand what the Measure is intended to assess, and can more precisely decide where to place the rating mark. Thus, the intention is to improve both rating reliability and validity. See Wilson (2005) for a discussion of the measurement of progress variables. By presenting observed behaviors in this more overtly developmental format, DRDP-R enhances teacher understanding of the nature and sequence of the progress they can expect their children to be making, thus contributing to the teacher's own professional development. Moreover, in revising the DRDP, the development team relied on input from practitioners, results of previous pilot studies and the most recent theoretical and empirical literature. The guiding principles in creating each measure were:

- The developmental levels within each measure are derived from the scientific research literature in early childhood education and developmental psychology.
- The behaviors described in the measures are typical for children engaged in healthy, age-appropriate, best practice, play-based activities.
- The language used is appropriate and meaningful for the program staff who are expected to use DRDP-R.

The DRDP-R accommodates the complex nature of development by focusing on major milestones that typically occur sequentially, such as the sequence of physical development that the majority of infants and toddlers demonstrate, from rolling over to crawling, followed by walking and then running. By focusing on major milestones of development, and defining "mastery"[5] of a developmental level as that behavior which the child demonstrates repeatedly and on a regular basis, the DRDP-R accommodates children's natural and sometimes non-linear

[5] The DRDP-R defines "mastery" in the following way: A child has mastered a developmental level if she or he typically demonstrates the behaviors in that level's descriptor. Behaviors are considered typical if the child demonstrates them easily and confidently, consistently over time, and in different settings. A child may occasionally behave at a higher or lower level, but mainly demonstrates behaviors representative of one level.

developmental pathways. In addition, since children are rated on discrete developmental constructs at the level that most closely matches their current level of functioning, the instrument reflects children's higher or lower levels of development across a wide range of measures. Thus, the developmental profile of a child who may have more advanced gross motor skills but less advanced fine motor skills will reflect these differential levels of mastery appropriately.

The DRDP-R assessment system was designed to measure the achievement of DRs for children from birth through 12 years of age. To cover this broad age range, there are separate DRDP-R instruments for three age levels corresponding to three different levels of educational organization used by CDE:

Infant/Toddler (**IT**) – birth through 2 years old;

Preschool (**PS**) – 3 and 4 years old;

School-Age (**SA**) – 5 through 12 years old.

One of the main goals of the development team was to ensure that each of the instruments targets important and observable behaviors for the age range it was designed for, and was readily interpretable by program staff working with children of that age. Therefore, although the DRDP-Rs are designed to measure similar sets of underlying constructs, the focus of each instrument is age-specific. In other words, discontinuities between age groups exist for some of the DRDP-R measures. However, there are contexts where a perspective across pairs of these instruments is needed, for example, for children at about the intersection points between instruments (e.g., entering preschool before age 3, or leaving preschool after age 5). Similarly, a cross-instrument (wider age-groups) perspective is needed for assessment of children receiving special education services (see the DRDP-Access[6]). Therefore, another goal of the team was to ensure constructs are measured well between age-consecutive measures. Throughout the DRDP-R development process, special attention was given to ensuring an appropriate articulation of the different instruments across age groups.

1.3 DRDP-R Format and Function

The DRDP-R is comprised of ten indicators for CDE's four DRs. Every measure on the DRDP-R is associated with one of the indicators and, in turn, one of the four DRs. Every indicator has multiple measures that define it more specifically. To support reliability of measurement, and ease of interpretation of results, the ten indicators were grouped into six Domains of development. The groupings of Indicators into the six Domains measured by DRDP-R are shown with their shorthand abbreviations:

Domain: Self and Social Development: SELF/SOC
- Indicator: *SELF – Children show self-awareness and a positive self-concept.*
- Indicator: *SOC– Children demonstrate effective social and interpersonal skills.*

Domain: Self Regulation and Self Care: REG/SH
- Indicator: *REG – Children demonstrate effective self-regulation of their behavior.*
- Indicator: *SH – Children show an emerging awareness and practice of safe and healthy behavior.*

Domain: Language and Literacy: LANG/LIT
- Indicator: *LANG – Children show growing abilities in communication and language.*
- Indicator: *LIT – Children demonstrate emerging literacy skills.*

Domain: Cognitive Development: LRN/COG
- Indicator: *LRN – Children are interested in learning new things.*
- Indicator: *COG – Children show cognitive competence and problem-solving skills through play and daily activities.*

---

[6] http://www.cde.ca.gov/sp/se/sr/drassessment.asp

Domain: Mathematical Development: MATH

Indicator: *MATH – Children show interest in real-life mathematical concepts.*

Domain: Motor Development: MOT

Indicator: *MOT – Children demonstrate an increased proficiency in motor skills.*

Each indicator is assessed using multiple measures. As described above, each measure describes a range of developmental levels that the teacher may choose from in characterizing the progress of an individual child for a single aspect of development within the age range of the instrument. The measure gives a name for each developmental level, a brief description of what it means for a child to be at that level, and several example behaviors that might be observed to indicate a child is at that level. See the example of a DRDP-R Measure in Figure 1.1.

The descriptors for each developmental level (the text below the oval response option) explicate the type of behavior expected from a child at this level. The examples below each descriptor provide the teacher with a sample of observable behaviors a child might exhibit in their program setting. One important consideration is that this list of examples is not intended to be exhaustive, and is provided only as a sample. In some circumstances it may be appropriate to develop different sets of exemplars for different contexts. For each measure, teachers can choose whether to rate the child at one of the developmental levels, rate the child as "not yet at first level"[7], or opt not to rate the child due to lack of evidence. Teachers can also provide comments and documentation to support their rating in the form of "anecdotal records."[8]

Teachers are required by the CDD to complete the age-specific DRDP-R instrument twice a year for every child under their care. They are trained to observe and document

[7] This option is missing from the IT instrument because the first level is the most basic one possible (i.e., reflexive)

[8] Anecdotal records could include brief descriptions of a child's actions or words, drawings, photos of a child engaged in a relevant activity, etc.

children's behaviors during regular daily activities. They use anecdotal notes and checklists to reflect on each DRDP-R measure and select the most appropriate level for each child on every measure[9]. Teachers are required to undertake professional development in order to gain skills using DRDP-R.[10]

In addition to assisting teachers in tracking an individual child's development, the DRDP-R is intended to serve as a training tool for practitioners. The measures include all core domains of development, provide examples of observable children's behavior during typical daily activities in high quality programs, and form a research-based framework for development. An individual child's unique developmental pathway can be documented as teachers select the most appropriate level for each of the measures. The recurrent use of the instrument enables the teacher to track changes in developmental levels as children grow in their care over time, as well as to become increasingly familiar with the measures.

Completing the DRDP-R takes time and effort, and time management can be difficult for teachers who interact with many children. With time, teachers develop an understanding of the DRDP-R indicators and measures, which helps them to quickly identify the most appropriate developmental level for each measure. Teachers and administrators report that they find the DRDP-R useful for reflecting on their children's progress and for planning ahead.

[9] Teachers are encouraged to confer with other teachers who interact with each child to help them assign ratings that most accurately reflect the child's current developmental level.
[10] Training materials can be found at http://www.wested.org/desiredresults/training/ .

2 Content Information for the DRDP-R

2.1 Introduction

Evidence based on (test) content provides a foundation for investigating the remaining aspects of validity evidence (Wilson, 2005, p. 157). Validity evidence based on test content is an important factor in ensuring that the content of an assessment is in alignment with the underlying construct it is designed to measure. Messick (1994) defines the content aspect of construct validity to include "evidence of content relevance, representativeness, and technical quality" (p. 11). To build evidence regarding the content of an assessment, the researcher needs to analyze "the relationship between a test's content and the construct it is intended to measure" (AERA, APA, NCME, 1999, p. 11). To produce this evidence, the following steps or "building blocks" were followed: "(a) definition of the construct (and a visual representation thereof as a construct map); (b) a description of a set of items that comprise the instrument—the items' design, (c) a strategy for coding the responses into an outcome space, and for scoring them—the outcome space, and (d) a technically calibrated version of the construct—the Wright map" (Wilson, 2005, p. 156). To begin the investigation of the evidence based on test content, an exhaustive review of empirical research findings in the early childhood literature was completed regarding the substance of the DRDP-R.

To examine the evidence based on content validity for the DRDP-R, it is important to recall the structure of the DRDP-R. The DRDP-R is based on ten indicators for CDE's four Desired Results (DR). Each measure on the DRDP-R addresses one of the indicators while each indicator relates to one of the four DRs. All indicators have multiple measures that define them more specifically. The indicators are grouped into the six indicator sets or domains measured by DRDP-R. The six domains are (1) Self and Social Development, (2) Self Regulation and Self

Care, (3) Language and Literacy, (4) Cognitive Development, (5) Mathematical Development, and (6) Motor Development. An extensive literature review was conducted to develop the measures so that they (1) would accurately represent and demonstrate the Indicators, and (2) be based on leading research in the field of child development. Authors in the fields of early-childhood education and developmental psychology and expert practitioners in Infant/Toddler, Preschool, and School-Age care programs were a part of the development of the DRDP-R.

2.1.1 *Domain: Self and Social Development (SELF/SOC)*

Self Concept (SELF) and Social Interpersonal Skills (SOC) are the indicators that help define the domain of Self and Social Development. In the DRDP materials, Self Concept is summarized as: Children show self-awareness and a positive self-concept. The Social Interpersonal Skills indicator is summarized as: Children demonstrate effective social and interpersonal skills. DRDP-R measures for this domain are summarized in Table 2.1.

Table 2.1 – DRDP-R Self and Social Development (SELF/SOC) Measures

Infant Toddler	Preschool	School Age
▸ identity of self and connection to others ▸ recognition of ability ▸ self expression ▸ empathy ▸ interaction with adults ▸ relationships with familiar adults ▸ relationships with familiar peers ▸ interactions with peers ▸ and awareness of diversity	▸ identity of self ▸ recognition of own skills and accomplishments ▸ expressions of empathy ▸ building cooperative relationships with adults ▸ developing friendships ▸ building cooperative play with other children ▸ awareness of diversity in self and others	▸ identity of self and connection to others ▸ self-esteem ▸ empathy ▸ interactions with adults ▸ friendship ▸ conflict negotiation ▸ awareness of diversity (appreciation of differences and similarities)

Identity of self is assessed at each of the three age levels of the DRDP-R. Self-concept is defined as an individual's theory of self (Harter, 1990, 1998, 1999). This theory of self includes a set of beliefs about one's own characteristics, such as the attributes, abilities, attitudes, and values that an individual believes define who he or she is. Because it is a theory, one's self-concept is continually modified and changed with experience. By age 2, some toddlers are already using first person pronouns (e.g., I, me, my) to refer to the self, and second person pronouns (e.g., you) to refer to a companion (Lewis & Brooks-Gunn, 1979). Toddlers progress to be able to describe themselves in concrete and physical terms, such as describing physical characteristics (age, size, and gender), possessions, and the physical actions they can perform (Damon & Hart, 1982). Preschoolers usually describe themselves with concrete terms such as name, physical appearance, possessions, and everyday behavior. By age 3 ½, some can also describe themselves in terms of typical beliefs, emotions, and attitudes. Preschoolers do not yet make explicit reference to internal dispositions. During the preschool years, children gain more and more understanding of which social categories they belong to (Sigelman & Shafer, 1991). Children's struggles with one another over objects seem to be positive efforts at forming boundaries between self and others. Older children generally move beyond the concrete categorizations of the self.

Ideas related to self esteem are aspects of the self-concept and social interpersonal skills domain that involve judgments about one's own worth and the feelings associated with those judgments. Younger children usually rate their own ability as extremely high and underestimate the difficulty of a task (Curry & Johnson, 1990; Stipek, Roberts, & Sanborn., 1984; Stipek, Recchia, & McClintic, 1992). A high sense of self-esteem contributes greatly to a child's initiative during a period in which they must master many new skills (Durkin, 1995; Erikson, 1963; Hartup, 1983; Rubin, 1980). Contrary to this, even a little criticism can undermine a child's self-esteem and enthusiasm for learning.

Empathy is thought to contribute to an orientation towards others' feelings and needs, which is incorporated in moral reasoning and reflected in social behavior (Eisenberg, Spinrad, & Sadovsky, 2006; Hoffman, 1990, 2000). From an early age, empathy is related to prosocial behavior. With the rudimentary development of role-taking abilities, the period of "empathy for another's feelings" emerges as early as 2 to 3 years of age (Hoffman, 1990, 2000). At this stage, children are increasingly aware of other people's feelings, and differences between other people's perspectives and their own. Thus, prosocial actions reflect an awareness of the other person's needs. Moreover, with the development of language, children begin to empathize with a wider range of emotions than previously. With increasing age, children are more likely to respond to others' distress with empathy and prosocial behaviors (Eisenberg et al., 2006).

Young children's relations with adults and peers are, in many ways, affected by earlier parent-child interactional routines in families (Parke & Ladd, 1992). Children seek, in adult caretakers and peers, the emotional bonds and feelings of security they first established in families (Ladd, 1992). During the preschool years, children are eager to master new skills, use language to ask questions to seek new meanings, and enlist others in work and play interactions (Puckett & Black, 2005). Their social circle is expanding rapidly, and they actively seek interactions with others. Being engaged in cooperative efforts with others is particularly enjoyable. Although children are not so dependent on adult guidance and supervision to find meaningful activities, contain impulses, follow rules, or avoid prohibited behavior, they are effectively using adults as resources when their own capacities are exceeded, or when disappointment, sadness, or other strong emotions are beyond a level they can manage (Elicker, Englund, & Sroufe, 1992). Children also attempt to gain control over their lives by resisting and challenging adult rules and authority (Corsaro, 2005). They challenge adult rules in the family from the first year of life. Such activity becomes more widespread and sophisticated when children discover common interests in later school settings. In these settings, children produce a wide set of practices in which they both mock and evade adult authority.

Naturally occurring conflict is an opportunity for children to develop social, emotional, intellectual, and moral skills by working through their disagreements (Sandy & Cochran, 2000). Conflict serves different purposes according to the level at different levels of early child development. During the 2nd and 3rd years, it corresponds with children’s developing autonomy. The increasing assertiveness of the child is to be desired rather than socialized into compliance with parental demands. Between the ages of 3 and 7, constructive conflict management helps to coordinate play (Sandy & Cochran, 2000). By the time children have reached 3 or 4 years of age, they attempt to justify their viewpoints during conflict using rights of entitlement. Compromise, negotiation, and conciliation are more frequently observed in older children during conflict situations (Dunn & Herrera, 1997; Ross & Conant, 1992).

Between ages 3 and 5, children become aware of racial categories but do not always accurately classify themselves (Spencer & Markstrom-Adams, 1990). Children prefer classmates who are similar to themselves in terms of both race and sex (Ramsey & Myers, 1990). Children at this point are not forming generalized negative attitudes toward other races, but are dealing with their own developing self-concepts and racial identities (Puckett and Black, 2005). Young children can make global distinctions between social class differences (rich and poor) and have a few ideas about these discrepancies (Ramsey, 1991). Preschoolers are aware of differences characterizing their peers with disabilities and are able to explain the differences by referring to the degree of immaturity, the occurrence of accidents, and the use of adaptive equipment such as a wheelchair (Diamond, 1993). Opportunities to play with classmates who have disabilities facilitate children's sensitivity to the needs of others (Diamond, 2001).

2.1.2 *Domain: Self Regulation and Self Care (REG/SH)*

The Self Regulation and Self Care domain includes two indicator sets. The first indicator, Self Regulation (REG), is summarized as: Children demonstrate effective self-regulation of their behavior. The second indicator, Self Care (SH) is summarized as: Children show an emerging awareness and practice of safe and healthy behavior. DRDP-R measures for this domain are summarized in Table 2.2.

Table 2.2 – DRDP-R Self Regulation and Self Care (REG/SH) Measures

Infant Toddler	Preschool	School Age
▸ impulse control ▸ seeking other's help to regulate self ▸ responsiveness to other's support ▸ self comforting ▸ attention maintenance ▸ personal care routines ▸ safety	▸ impulse control ▸ taking turns ▸ shared use of space and materials ▸ personal care routines ▸ personal safety ▸ understanding healthy lifestyle	▸ impulse control ▸ follows rules ▸ personal care routines ▸ safety ▸ understands healthy lifestyle ▸ exercise and fitness

Measures to examine impulse control appear for each of the three age groups (Infant/Toddler, Preschool, and School-Age) assessed using the DRDP-R. Emotion regulation is one of the most challenging aspects of emotional development (Shonkoff & Phillips, 2000). Both positive and negative emotions require regulation, and this usually happens in a real-life context that can be overwhelming for children and lead to being frustrated, upset, or embarrassed. Self-regulation not only decreases the number of disappointments, frustrations and hurt feelings in young children, but it also eases them into having more positive relations with others (Shonkoff & Phillips, 2000). This relationship between emotion and relationship with others has critical implications for children's social development. Self-regulation facilitates "emotional self-efficacy" (Saarni, 1990, 1999); decreases outburst, increases attention, and helps manage stressful situations (Garber et al., 1991); and it's a prerequisite for the learning to comply with external and internalized standards of conduct (Zahn-Waxler & Radke-Yarrow, 1990; Zahn-Waxler, Robinson, & Emde, 1992). Infants show rudimentary skills to manage their own emotional experience, relying mostly on their ability to enlist the help of others. For example, this is observable in the comfort-seeking strategies of infants (Thompson, 1990). Research on self-control examines the child's emerging ability to comply with a request, to inhibit or delay an activity, and to monitor behavior according to situational demands (Kopp, 1982). The ability to exercise self-control increases from 18 to 30 months and becomes more stable across time and across situations (Vaughn, Kopp, & Krakow, 1984). For example, toddlers (2½ year olds) are capable of making active efforts to avoid or ignore emotionally arousing situations, using strategies such as engaging in encouraging or reassuring self-talk, and changing or substituting frustrated goals (Braungart & Stifter, 1991; Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986; Buss & Goldsmith, 1998; Calkins & Johnson, 1998; Cummings, 1987; Grolnick, Bridges, & Connell, 1996; Stein & Levine, 1989, 1990). By the time they are ready to enter school, children's regulatory repertoires have become increasingly proficient and flexible. For example, children learn that their interpretations of events can affect how they react and that they can camouflage their emotions when needed (Harris, 1993). Children's regulatory capacities have also been related to increasing early conscience and moral behavior, as well as to diminishing impulsive and negative behavior (Kochanska, Padavich, & Koenig, 1996; Kochanska, Murray, & Coy, 1997).

Important self regulation skills assessed on the DRDP-R include: following rules (School Age); taking turns and sharing space (Preschool); and seeking others' help to regulate self, responding to others' support, self comforting, and attention maintenance (Infant/Toddler). Sharing is a regular practice in every child's life, beginning as soon as the child interacts with other children – often during the first year of life (Damon, 1988). During infancy, children discover by chance that other infants share an interest in toys and that joint play with the same toy is more fun and more interesting than solitary play. They especially delight in the symmetrical rhythm of turn-taking with toys and other objects (Piaget, 1962). Two- and three-year-olds spontaneously give gifts and share their toys with other children and with unfamiliar adults (Steinberg & Meyer, 1995). At these ages, gift giving is a way to begin and maintain social contact, and the gifts may be everyday objects such as pieces of wood or stone. Most early sharing during the first three years of life is done for the fun of the social play ritual, from unquestioning deference to an authority figure's demands, or out of mere imitation. It is around the fourth year that the combination of natural empathic awareness and reasoned adult encouragement leads the child to develop a firm sense of obligation to share with others (Damon, 1988). The advent of this perspective makes the child a more reliable and consistent sharer, even in the absence of authority figures. Children now believe that they should share, but not necessarily that they should be as generous with others as with themselves. Nor do their actions always live up to their beliefs, particularly when the object of contention is highly desirable. By the beginning of the elementary school years, children begin to express more genuinely objective notions of fairness with some regularity (Damon, 1988). In the preschool and early elementary school year, children immensely enjoy simply doing things together.

The self regulatory and self care areas addressed at each of the three age groups are: personal care routines and safety. Understanding healthy lifestyles is addressed on the Preschool and School-Age instruments, while exercise and fitness is addressed on the School Age DRDP-R. The DRDP-R is intended to measure how well the child embraces the safe and healthy choices he or she is exposed to in the program, and makes them his or her own. Therefore, these measures are not based on a developmental theory about safe and healthy life choices. They are based on sound standards of healthy and safe practice in child care programs, and assess children's independence in following these practices. Thus, the measure focuses on the most observable and salient routines: hand-washing, and nose-blowing. Hand-washing is the most important way to reduce the spread of infection. Research has shown that unwashed or improperly washed hands are the primary carriers of infections (Hawks, Ascheim, Giebink, Gravile, & Solnit, 1994). Research has also shown that implementing a hand-washing training program reduces the number of incidences of various illnesses such as diarrhea and colds (Donowitz, 1996; Kotch, 1990; Niffenegger, 1997; Roberts et al., 2000; Soto & Belanger, 1994). Children are expected to "learn to blow or wipe their noses with disposable, one-use tissues and then discard them in a plastic-lined, covered, hands-free trash container. After blowing the nose, they shall wash their hands" (American Academy of Pediatrics [AAP], American Public Health Association [APHA], & National Resource Center for Health and Safety in Child [NRCHSCC], 2002, standard 3.024).

Understanding healthy lifestyles is addressed on the Preschool and School-Age DRDP-Rs while exercise and fitness is addressed only on the School Age instrument. The levels on the instruments describe how the responsibility to keep a healthy life style (food, rest, and health) develops from dependence on the adult to child self-reliance. Children are expected to learn how to sit appropriately while eating (to reduce the risk of food aspiration), develop skills and coordination in handling food and utensils, and to become able to serve food themselves (Graves, Suitor, & Holt, 1997; Lally et al., 1995; Endres & Rockwell, 1994), 1994; Pipes & Trahms, 1995; USDA, 1994; USDA, 1995). They also need to be able to make nutritionally-sound choices when eating. For example, by learning about new food, children increase their knowledge of the world around them, as well as the likelihood that they will choose a more varied, better balanced diet in later life (Birch, 1990). Nutrition, in turn, is a vital component of good health (AAP et al., 2002). Nutrition has also been found to be related to other indicators of children's well being. For example, Espinosa and colleagues (1992) found that nourished children are more active, happy, and show more leadership behavior, whereas poorly nourished children appeared more anxious while exercising or engaging in fitness activities.

2.1.3 *Domain: Language and Literacy (LANG/LIT)*

The two indicators in the Language and Literacy domain are simply called (1) Language (LANG) and (2) Literacy (LIT). For the DRDP-R, the Language indicator is summarized as: Children show growing abilities in communication and language. The Literacy indicator is summarized as: Children demonstrate emerging literacy skills. DRDP-R measures for this domain are summarized in Table 2.3

Table 2.3 – DRDP-R Language and Literacy (LANG/LIT) Measures

Infant Toddler	Preschool	School Age
‣ language comprehension ‣ responsiveness to language ‣ communication needs ‣ feelings and interests ‣ reciprocal communication ‣ interest in literacy ‣ recognition of symbols	‣ comprehends meaning ‣ follow increasingly complex instructions ‣ expresses self through language ‣ uses language in conversation ‣ interest in literacy ‣ letter and word knowledge	‣ pursuit of understanding ‣ expression of oral language ‣ interest in literacy ‣ decoding (word recognition and use) ‣ writing ‣ comprehension of written materials

- emerging writing
- concepts of print
- phonological awareness

Aspects of language regarding a child's comprehension of meaning centers on two areas: content (vocabulary, basic language concepts, and meaning) and form (grammar or syntax). Taken together, these represent a single measure of overall oral language development. Children's understanding and use of words – that is, their vocabulary knowledge – provides children an important language tool they will use to access background knowledge, express ideas, and acquire new concepts. Vocabulary undergoes a rapid growth during the early years (Bates, Bretherton, & Snyder, 1988; Hoff, 2005; Owens, 1996). One element of vocabulary development is the attainment of a core group of terms that represent basic language concepts. These concepts include concepts about family (e.g., mother, father, sister), colors (e.g., red, orange, blue), size of objects (e.g., small, big, huge), location (e.g., in, on, under, above), and an additional aspect of early vocabulary and linguistic concept development: categorization (Hoff, 2005; Owens, 1996). Often, children's language must be contextualized, or supported by the immediate context. As children's vocabulary and language concepts expand, they can be more decontextualized in their language use and comprehension.

This movement from the concrete and contextualized to the abstract and decontextualized plays a critical role in the development of academic language (also called *literate language;* see Curenton & Justice, 2004) and the vocabulary used to produce and comprehend the relatively abstract content of written language (Charity, Scarborough, & Griffin, 2004; Dickinson & Snow, 1987; Dickinson & Tabors, 1991; Snow, 1983). In the first five years of life, children's language development involves not only the acquisition of vocabulary and basic language concepts, but also the development of an adult-like syntax, or grammar. Grammar refers to the structural organization of phrases and sentences, or essentially the way words are strung together in a linear order to make meaning. Most accounts of language acquisition (Chomsky, 1957) assert that children have an innate propensity towards learning the grammatical rules that govern their language, which includes the organization of basic sentences (e.g., subject + verb + object: Juan drew the picture) and the joining of clauses and phrases to elaborate the basic sentence structure (e.g., subject + verb + object and subject + verb + object: Juan drew the picture and he is hanging it). The idea of following instructions merges content, form, and use (function or pragmatics). It is a way to observe whether children are understanding language in multiple ways. It is also part of preschool's daily language exchanges and one that will not only help children have a positive participation in preschool but also when they enter the school system, where following directions is part of most classroom's participant structures (Phillips, 1971).

Measures on the DRDP-R focus on production (as well as comprehension, as previously mentioned). Production of more complex vocabulary, grammatical structures, and terms tends to happen later than understanding of those same things. Children's first words, which typically emerge at the end of the first year, most often reflect objects and persons that are very meaningful to them (e.g., doggie, mama, cup; Anglin, 1995) and that are present in their immediate environment. From this base, a child's initial "core group" of vocabulary (Bloom & Lahey, 1978) changes and evolves over time: from an emphasis on personal experience and concrete labels to those that represent a more "general" or abstract understanding of the world and the events and relationships within it (Wehren, DeLisi, & Arnold, 1981).

Language use is defined as: the way in which language is used for social and communicative purposes. This definition is used to focus the measures addressing language use. It relates to the child's development of language as a tool for communication purposes. As children develop as users of language, they are able to negotiate with peers for social purposes, request and question information from others, and participate in extended language-based interactions that focus on both lower level and higher level cognitive content. Two aspects are emphasized in the measures addressing language use: on-topic conversational participation and use of language for various communication purposes. In addition, the focus moves from contextualized language in the lower levels to more decontextualized language in the higher levels addressed within the DRDP-R measures.

The "Interest in Literacy" measure (so named for each age group's DRDP-R) combines two important aspects of literacy: children's interest in literacy and children's ability to understand stories based on written texts. Research indicates that children's interest in literacy has a positive influence in children's comprehension of text. Interest in and motivation towards reading describes the child's affect towards literacy activities (Alexander & Filler, 1976; Mathewson, 1994; McKenna, Kear, & Ellsworth, 1995). Motivation theorists propose that an individual's beliefs, motivation, and purposes have a high influence on their decisions about which activities to do, how long to do them, and how much effort to put into them (Bandura, 1997; Eccles, Wigfield, & Schiefele, 1998; Pintrich & Schunk, 1996). Motivation towards reading influences individual's engagement with reading and literacy activities by facilitating their entry into a "psychological state of interest" (Krapp, Hidi, & Renniger, 1992), in which individuals demonstrate increased attention, cognitive functioning, and persistence in different literacy tasks, as well as an increase in their affective investment (Hidi, 1990; Krapp et al, 1992).

The processes activated in such state of interest facilitate individual’s comprehension and recall of the information to which they are exposed during book reading and other literacy activities (e.g., Anderson, 1982; Asher, 1979,1980; Bernstein, 1955; Estes & Vaughan, 1973; Hidi, 2001; Hidi & Baird, 1986, 1988; Kintsch, 1980; Schank, 1979; Schraw, Bruning, & Svoboda, 1995).

Reading comprehension is influenced greatly by language comprehension, and in large part these draw upon the same developmental processes (Perfetti, Van Dyke, & Hart, 2001). Just as children move from understanding simple phrases and directions to comprehending more detailed information, they also progress from remembering isolated aspects of simple stories to understanding parts of more complex literacy events. Children’s development of narrative thinking goes through a series of stages that ultimately help them to make sense of stories and the world around them (Paris & Paris, 2003), and constitutes an important foundation for learning to read (Burns, Griffin, & Snow, 1999; Whitehurst & Lonigan, 1998). At the early stages, preschoolers construct narrative scripts, which involve primitive accounts of story plots. These scripts usually focus on the description of familiar events and routine activities, such as going to a birthday party or visiting the doctor. Over time, children construct narrative schemas, which include knowledge about the main elements of stories (such as characters and settings) and about the sequence of events (such as time order and causal progression). In the last, and perhaps most difficult stage, pre-K children come to understand and relate to characters’ internal responses, such as their mental processes and experiences. This ability to understand characters’ internal thinking also helps children to develop a sense of perspective by which they can emphasize with the experiences and reaction of characters in a story, and helps children develop the ability to recognize both the external and internal features of narratives (Paris & Paris, 2003).

The ability to recognize first symbols and then letters is a basic step in the process of learning to read and write. Knowledge of the alphabet is related to children's reading ability in that knowing the names of letters facilitates children's ability to decode text and to apply the alphabetic principle to word recognition. For most children the name of the letters is what helps them connect the sounds in words and letters in print (Durrell, 1980). The order of learning the alphabet letters seems facilitated by both environmental and developmental influences. An important environmental influence is exposure to the individual letters of the alphabet. Studies show that children learn first the letters that are most familiar to them, such as the letters in their own names and the letters that occur earlier in the alphabet string (Treiman & Broderick, 1998). Studies also show that features of certain letters makes them more amenable to learning. For instance, letters that contain their sound in their name (B and F, for instance) are learned earlier than those letters that do not (Q and W, for example) (Treiman & Broderick, 1998). Finally, children learn those letters earlier that map onto earlier-acquired phonemes, such as B and D, which are acquired earlier than, say R and L (Justice, Pence, Bowles, & Wiggins, 2005).

As children become aware of the name of letters, they also start to be able to identify printed words. Ehri (1995) has defined children's developments in word recognition as a series of transitions as they move from pre-alphabetic readers (learning words by sight and using salient contextual cues for word recognition), to partial alphabetic (applying some phonetic information, such as the sound corresponding to the first letter in a word, to recognize the word), to full alphabetic (reading a word using the alphabetic principle). For children who are younger than school age, their word recognition is primarily of a pre-alphabetic type, meaning that they can recognize some words but that they rarely examine the alphabetic or phonetic structures of the word to arrive at its meaning.

Learning to write (emerging writing or writing skills) involves cognitive, social, and physical development. Children who grow up in an urban environment from a very young age notice the writing in their surroundings. At first they begin to differentiate between writing and other kinds of visual representation such as drawing (Bissex, 1980; Harste, Woodward, & Burke, 1984; Ferreiro & Teberosky, 1982). They differentiate between the tools for writing and the tools for drawing ("I need to get a pencil to write my name"). Their writing looks different from their drawing (Ferreiro & Teberosky, 1982; Harste et al., 1984) often being linear in form rather than circular. Children's emergent writing abilities are demonstrated in the preschool classroom with activities such as pretending to write and learning to write one's name (Whithurst & Lonigan, 2001). Children's writing tends to follow a well-documented path. Initially, children demonstrate a global form of writing. They tend to treat writing from a pictographic perspective, which is usually demonstrated by using drawings as writing or using idiosyncratic scribble-like markings that only have meaning for the child. Children later move to a stage in which they use letter-like forms to write, and in many cases these marks resemble characteristics of real writing (e.g., longer words are represented by longer strings of letter-like symbols). For many children, this process moves along to the next stage of using actual letters to write, even when there is no connection between the true spelling of what they want to write and what they produce; that is, they produce nonphonetic strings of letters (Ferreiro & Teberosky, 1982; Sulzby, 1986, 1987).

This stage is followed by a period in which children produce phonetic spelling, also called "invented spelling." In this stage, children use letter-like symbols to represent the parts of words that they hear and attempt to match letters to sounds or syllables, usually from a phonological rather than an orthographic perspective (Ferreiro & Teberosky, 1982). Throughout this early stage of learning to write, children come to realize that writing carries meaning; people should be able to read what you write (Clay, 1977; Kress, 1994; Harste et al., 1984). They also learn that people write for different purposes (Taylor & Dorsey-Gaines, 1988; Heath, 1983; Scheiffelin & Cochran-Smith, 1984; Ferreiro & Teberosky, 1982; Teale, 1987). Finally, children have the physical experience of actually writing and drawing, where they begin to develop effective (or not-so-effective) ways of handling writing implements.

Concepts of print and comprehension of written materials are important elements in the development of a child's literacy. Children need to understand that print carries meaning – that there is a meaning or message encoded. This basic understanding is critical to all later development (Purcell-Gates, 1996). Children will need knowledge of concepts about print, which emphasize five areas of development: print interest, print functions, print conventions, print forms, and print part-to-whole relationships (Justice & Ezell, 2004), which are described next.

Print interest refers to children's coming to view print as an object warranting attention as a distinct type of environmental stimuli (Whitehurst & Lonigan, 2001). Print function refers to children's awareness that print carries messages that can serve multiple purposes (Gunn, Simmons, & Kameenui, 1995, van Kleeck, 1990). It also refers to children's ability to recognize that changes in the function of print are commonly related to socio-cultural features of the context in which print is used. Print conventions describe children's growing knowledge of the ways in which print is organized, and how this organization changes for various genres. The specific knowledge of print associated with this period includes an initial knowledge of the alphabet and letter recognition (the graphic aspects only – not the sounds associated with the letters), and a sense of the conventions associated with print. These conventions include directionality, and for English orthography it pertains to the left-to-right and top-to-bottom organization of print in books or other print media (Clay, 2002). It also refers to the way books are organized (title, author, front and back) and the way they should be handled (Clay, 2002).

Print forms describes children's understanding that words, letters and other print units have distinct names and are used in specific, organized ways. As children come to understand the intentional aspects of print—basic conventions and the alphabet being the building blocks—their further development is marked by an understanding of the concept of a word (Beers & Henderson, 1977; Henderson & Beers, 1980). Knowledge that the word is the basic unit of meaning in the reading and writing process is a critical transition point (Ferreiro & Teberosky, 1982), one that if not mastered may prevent children from moving further in their literacy development (Clay, 1993).

Print part-to-whole relationships describe children's growing knowledge of the combinatorial properties of print units (but this ability is too complex to be included at the preschool level). These areas are not to be conceptualized as stages with a fixed sequential nature, but rather as areas of development that interrelate and influence one another as children develop their sophisticated understanding of print forms and functions (Justice & Ezell, 2004).

Phonological awareness is generally defined as an individual's sensitivity to the sound (or phonological) structure of spoken language. Spoken language is made up of different phonological units, which differ in their linguistic complexity, including words, syllables, intra-syllabic units (onsets, rimes), and sounds (phonemes). Essentially, phonological awareness is an individual's ability to detect and/or manipulate these units of spoken language independent of meaning. The development of phonological awareness occurs on various fronts: (1) type of sound unit, (2) type of operation performed on those sound units, (3) position and context of the smaller sound unit within the larger sound unit (e.g., phoneme within a word), and (4) kind of supports provided when administering phonological awareness tasks or tests.

The development of phonological awareness typically moves along a continuum in which children progress from a sensitivity to larger concrete units of sound to a sensitivity to smaller abstract units of sound (Adams, 1990; Anthony, Lonigan, & Burgess, 2002; Fox & Routh, 1975; Goswami & Bryant, 1990; Liberman, Shankweiler, Fischer, & Carter, 1974; Lonigan, 2006; Lonigan, Burgess, & Anthony, 2000; Lonigan, Burgess, Anthony, & Barker, 1998; MacLean, Bryant, & Bradley, 1987; Treiman, 1992).

Children demonstrate their phonological awareness through three types of operations—detection, analysis, and synthesis (Anthony, Lonigan, & Burgess, 2003). Detection is the ability to match similar sounds. Analysis is the ability to segment words or syllables into smaller units. Synthesis is the ability to combine smaller segments into syllables and words. Children's phonological awareness performance usually progresses from detection to analysis to synthesis of units of sound, although this development does not occur in discrete stages, but rather are overlapping abilities (Anthony et al., 2003).

In addition to the variation in terms of the unit of sound and the type of operation performed on the units of sound, there also are other variations that make phonological awareness tasks more difficult. These include the position of a phoneme in a word or syllable and the context in which the phoneme occurs.

Another source of variation relates to the amount and kind of supports provided to children to enable them to perform these tasks. For example, when asking children to delete the onset of a word, providing pictures of stimuli can reduce the difficulty level of the task by helping children remember the different words, thus enhancing children's performance relative to having them perform this task without picture stimuli (Anthony et al., 2003). As phonological awareness progresses, a child's ability to comprehend written material increases.

2.1.4 *Domain: Cognitive Development (LRN/COG)*

The Cognitive Development domain is comprised of two indicator sets: Learning (LRN) and Cognitive Competence (COG). The Learning indicator is summarized as: Children show cognitive competence and problem-solving skills through play and daily activities. The Cognitive Competence indicator is summarized as: Children show cognitive competence and problem-solving skills through play and daily activities. DRDP-R measures for this domain are summarized in Table 2.4.

Table 2.4 – DRDP-R Cognitive Development (LRN/COG) Measures

Infant Toddler	Preschool	School Age
‣ memory ‣ understanding cause and effect ‣ problem solving ‣ symbolic play ‣ curiosity	‣ curiosity and initiative ‣ engagement and persistence ‣ memory and knowledge ‣ cause and effect ‣ engagement in problem-solving skills ‣ socio-dramatic play	‣ pursuit of understanding ‣ task persistence ‣ memory/knowledge ‣ cause and effect relationships ‣ problem-solving ‣ demonstration of inventiveness or inventive play

Memory (IT), memory and knowledge (PS), or memory/knowledge (SA) relates to the child's capacity to store, retrieve, and use information regarding events, objects, and actions. The literature in this area focuses on two aspects of young children's memory: event and autobiographical memory.

Research has demonstrated that even quite young children have general, temporal knowledge about familiar, recurring events (see Nelson, 1986, for a review). For example, children as young as 3 years of age can recreate a visit to the grocery store and report familiar and recurrent actions common in their temporal order. Research has shown that both children and adults find it easier to recall events when they follow a logical sequence, rather than an arbitrary one (Bauer & Mandler, 1989; Bauer & Shore, 1987; Ratner, Smith, & Dion, 1986). Children seem to be able to form spatial and temporal expectations that are general enough to include variations in the specific instantiations of the event (Arehart & Haaf, 1989; Slackman & Nelson, 1984; Smith, deSaint-Victor, & Arehart, 1988). Farrar and Goodman (1990) studied children's recall of a play event in which children were exposed to the same event three times (the scripted event) and a similar event with somewhat different activities and props once (the deviation event). Children tended to be able to recall more about the scripted event than the deviation one, and this was especially true for younger children.

Even though young children have been found to be able to demonstrate their ability to remember events, there are several differences in the way younger and older children recall these events. Younger children recall more component activities of an event under behavioral reenactment than they do verbally (Price & Goodman, 1990). Older children show greater hierarchical organization of event knowledge than do younger children (Ratner et al., 1986). Also, as discussed earlier, while both younger and older children sequence familiar events in their canonical older (Fivush & Mandler, 1985), older children sequence less familiar events better than do younger children (Fivush & Mandler, 1985; Price & Goodman, 1990). For example, in the Fivush and Mandler study, children were asked to both construct forward and backward sequences from unorganized displays of pictures and to reconstruct previously seen sequences. The youngest children could neither create nor reconstruct backward sequences, but the oldest children could do both. Five-year-olds were transitional; they could reconstruct previously seen sequences even if backward, but they could not create backward sequences.

Younger children are still forming their general event representations; that is, they are more likely than older children to be confused about what event instantiations (i.e., inclusions) occur during both standard and deviation (i.e., with changes to the standard structure) visits (Farrar & Goodman, 1990).

In sum, regardless of what event is being reported, older children's reports of events are more elaborate and more complex than younger children's (Fivush & Slackman, 1986; Nelson & Gruendel, 1986). For example, Friedman (1990, 1993) showed that the ability to order familiar daily events increases over the preschool period, and that children's understanding of sequence, duration, and distance of events begins during the preschool years but continues to develop in later childhood. Importantly, these developmental differences hold even when younger and older children have equal amounts of experience with the event (Farrar & Goodman, 1990; Myers & Bluhm, 1985, Price & Goodman, 1990). Finally, Fivush, Kuebli, and Clubb (1992) found that 5-year-olds were easily able to non-verbally reenact a variable event with new objects event after just a single experience. Three-year olds, in contrast, had more difficulty. After one experience with the event, they seemed confused when confronted with new objects and were unable to reenact the component activities of the event very well. After several experiences, however, they seemed to understand that certain components of this event changed from experience to experience and were easily able to reenact the event with new objects. This pattern suggests that younger children's representations were more concrete; they were more tied to the specific objects and activities encountered during the first experience. Only with repeated experience that provided explicit information about variability were they able to extend their event knowledge to novel variations.

Autobiographical memory incorporates many different concepts and skills—language ability, narrative understanding, temporal concepts, self concepts and consciousness, and social psychological concepts, and each of these individual and social processes follows a variable course of development. There are three critical aspects of autobiographical memory (Nelson & Fivush, 2004): (1) There is a gradual emergence of autobiographical memory across the preschool years rather than a point before which there are no autobiographical memories and after which there are; (2) language is a fundamental social cultural tool in the development of an autobiographical memory system; and (3) there are cultural, gender, and individual differences in autobiographical memory across the life span that need to be explained.

Autobiographical narratives involve at least three orderings of time, and in a full realization of a life story, a third ordering. The *first* refers to the order of the sequence within the event recalled, including settings, plans, goals, actions, outcomes, achievements, and the temporal and causal relations among them. Research indicates that very young children have a good command of sequence of familiar routines, or scripts (Fivush & Mandler, 1985; Nelson, 1986), and are sensitive to order—especially causal order—in brief, newly learned action sequences (Bauer & Mandler, 1989). The ability to order familiar daily events increases over the preschool period (Friedman, 1990, 1993). Children's understanding of sequence, duration, and distance of events begins during the preschool years but continues to develop in later childhood.

The *second* ordering relates to the specific time in the past in which the event is placed. For a young child who has no external measures of time (e.g., days, weeks, months, and years), this can be achieved primarily by nominal days; for example, "my birthday" or "Christmas," or times of the year such as "last summer." Use of labels of this kind indicates that the child is conceiving of an event as having happened at a particular time in the past different from the present. Research indicates that this use of nominal days to recall past events can be performed by children as young as two (Nelson & Ross, 1980). However, the acquisition of relative time markers, such as *yesterday* and *tomorrow,* is typically a late achievement, often not acquired until late in the 5th year. At the outset of their use, *yesterday* and *tomorrow* may be used for any day not today, or *yesterday* may be used for any time in the past (Harner, 1982), such as when a 2-year-old says "yesterday did that," referring to an event from earlier that day (Nelson, 1989a). It is not that children do not have a sense of past and future, but rather than the sense is related to particular activities.

The *third* temporal ordering of autobiographical memory involves the placing of memories in a life span relation, usually in relation to an external sequence, such as school years or jobs – an ability that goes beyond the capacity of the preschool child. Ongoing activities can be considered within the extended present. Thus, for example, having a meal may be comprehensible as a present "now." Actions that are completed within the activity may be referred to with the past tense, and those that are anticipated with the future tense. But beyond the bounds of the activity (e.g., after finishing the meal, or the playtime before the meal), actions would not qualify as "now." Memory of what has happened before "now" is simply general knowledge without specific reference to a point in the past (James, 1890).

From early on, understanding cause and effect plays an important role in interpreting, representing, and remembering events in the physical world (Goswami, 1998). Causal reasoning can be studied from three different perspectives: reasoning about causes and effects, reasoning on the basis of causal principles, and reasoning related to causal chains. All these aspects of reasoning ultimately affect children's ability to engage in scientific reasoning. Research indicates that children can predict or infer the final state of an object that has been changed by causal transformation, and can also infer the kind of transformation that relates two object states (Gelman, Bullock, & Meck, 1980). Several causal principles apply to the research of children's reasoning development. The priority principle asserts that causes precede their effects. This notion seems to be present by at least age 3 (Bullock & Gelman, 1979), and it is well established by age 5 with 3-, 4-, and 5-year-olds correctly attributing causality at rates of 75%, 94%, and 100%, respectively.

Another important causation principle, which seems to be present by age 3, is the *covariation* principle, which states that if an effect has a number of potential causes, then the true cause will be the one that regularly and predictably covaries with the effect (Shultz & Mendelson, 1975). Research about children's development of causal chains reasoning indicates that very young children can use information about three-term causal chains to reason about event sequences, whether the reasoning is measured via a prediction task (Baillargeon, Gelman, & Meck, 1981, reported in Bullock, Gelman, & Baillargeon, 1982) or via a problem-solving task (Shultz, Pardo, & Altman, 1982). However, research with hiding and finding experiments indicates that children are not able to understand the causal implications of the event sequence before the age of 4 (Sophian & Somerville, 1988).

Crick and Dodge (1994) reformulated earlier views (Piaget, 1952) regarding children's ability to problem solve. Their model includes the following steps:

1. Encoding of external and internal cues (children selectively attend to particular situational and internal cues),
2. Interpretation and mental representation of those cues (children interpret the selected cues),
3. Clarification or selection of a goal (children select a goal for the situation or continue with a preexisting goal),

4. Access or construction of a response (children access from memory possible responses to the situation or construct new behaviors in response to immediate social cues in a novel situation),
5. Response selection (children evaluate the previously accessed or constructed responses and select the most positively evaluated response for enactment), and
6. Behavioral enactment (children enact the chosen behavior).

In their overview of studies linking toddlers' thinking to older children's thinking, Chen and Siegler (2000, p.87) concluded that children use "several problem-solving strategies from the beginning of learning; they continue to use less advantageous strategies even after they learned a more advantageous one; they chose among strategies in fairly adaptive ways from the beginning of learning; their choices became increasingly adaptive with problem-solving experience." More specifically, research shows that when tasks are properly contextualized and related to children's play interests, and if younger children are overtly encouraged to use good reasoning strategies, they show meta-cognitive capabilities (Cullen, 1992; Schneider & Bjorklund, 1992; Slawinski & Best, 1995).

There is a growing body of evidence suggesting that high-quality symbolic play is an important facilitator of perspective taking and later abstract thought and higher-level cognition, and that there are clear links between pretend play and social and linguistic competence (Bergen, 2002). Although there are many ways to classify children's play, a common approach employs two broad categories: pretend and construction play (or social versus object play) (Parten, 1932; Piaget, 1959, 1962). Pretend play is the ‘‘voluntary transformation of the here and now, the you and me, and the this or that, along with any potential action that these components of a situation might have’’ (Garvey, 1990, p. 82). Every pretend act involves certain features, several of which are defining and necessary, such as having a pretender who is an animate being, a reality that is pretended about, and guidance by a mental representation (Lillard, 1993a, 1998b). In pretend play, children carry out action plans, take on roles, and transform objects as they express their ideas and feelings about the social world (Garvey, 1984). Action plans are blueprints for the ways in which actions and events are related and sequenced (e.g., family-related themes). Roles are identities children assume in play. There are various types of roles: functional (e.g., taking a trip requires passengers and a driver), relational (e.g., mother, father, and baby play in themes related to familiar home activities), character roles both stereotypic (e.g., nurse, teacher, doctor) and fictional (e.g., drawn from books and television, such as superheroes), and peripheral (e.g., imaginary friend, which is discussed but not enacted) (Garvey, 1984).

Objects influence the roles children assume: for example, household implements trigger family-related roles and action plans, but capes stimulate superhero play. Perceptually bound younger children may be aided by the provision of realistic objects (Fein, 1981). While some pretend play is solitary or shared with adults, preschoolers' pretend or socio-dramatic play is often shared with peers in the school or neighborhood. To implement and maintain pretend play episodes, a great deal of shared meaning must be negotiated among children. Play procedures may be talked about explicitly, or signaled subtly in role-appropriate action or dialogue. Players often make rule-like statements to guide behavior ("You have to finish your dinner, baby").

Construction play with symbolic themes is also popular with preschoolers, who use blocks and miniature cars and people to create model situations related to their experience. A kind of play with motion, rough, and tumble play, is popular in the preschool years. In this type of play groups of children run, jump, and wrestle. Action patterns call for these behaviors to be performed at a high pitch. Adults may worry that such play will become aggressive, and they should probably monitor it. Children who participate in this play become skilled in their movements, distinguish between real and feigned aggression, and learn to regulate each other's activity (Garvey, 1984). Research indicates that children's play interactions increase not only in amount, but also in complexity, with the age of the child. Preschool children, in particular, show major gains in the amount of peer interaction from 3 to 5 years. These gains are due in part to their growing social circle as well as to their increasing physical, cognitive, and socio-emotional capabilities (Hestenes & Carroll, 2000). Across the preschool years, children are increasingly able to converse and cooperate with each other, thereby lending complexity to their play (Hartup, 1983; Parten, 1932; Smilansky, 1968).

Gowen (1995) proposes the following stages in the development of symbolic play: prepretense (approximating pretense with no confirmation), pretend self (apparent pretense behavior toward self), pretend other (pretense behavior directed toward other), substitution (use of objects in a creative, imaginative, or pretense way) *(3 year olds)*, imaginary (pretends that an object, substance, person, or animal is present) *(older preschool child)*, active agent (animation of toy that represents a being), sequence/no story (repetition of a single pretense act/scheme with multiple receivers), sequence/story (use of more than one related scheme), planning (pretend play preceded by evidence of planning). During the third year, for most children, an important change occurs: the child engages in pretend activities while adopting the role of another person – a person with whom the child is intimately familiar (e.g., child's mother or primary caregiver) (Johnson, Christie, & Yawkey, 1999, p. 71).

Parten (1939) devised a method for classifying the types of play engaged in by infants, toddlers, and preschoolers, which included the following categories: unoccupied play (less than 2 years old), solitary play (2 to 2½ years), onlooker play (2½ to 3 years), parallel play (2½ to 3½ years), associative play (3½ to 4½ years), and cooperative play (4½ years). This model suggests that children develop sequentially through these stages, and some preliminary research seemed to support this claim (Barnes, 1971; Parten, 1932; Smith, 1978- references from Sigelman & Shaffer, 1991, p.476). As a result, it was believed that parallel play is common for younger children and fades away over time in favor of cooperative play; more recent research indicates, however, that parallel play is not discontinued over the course of development but rather remains a quantitatively dominant mode of social interaction for 4- to 5-year-old children (e.g., Rubin, Watson, & Jambor, 1978).

In a fine-grained analysis of children's parallel play, researchers have found that children engage in more mature forms of parallel play across these early years than previously thought, moving from parallel-engaged play (conducting similar activities with little awareness of others) to more parallel-aware play (where eye contact with and mutual awareness of others is displayed) (Howes, 1980; Howes & Matheson, 1992). These studies also demonstrate that children become increasingly capable of going beyond parallel-aware play, to simple social play such as talking, giving, and sharing, to engaging in more complex forms of cooperative-social play that involve the enactment of organized, constructive, or socio-dramatic play with complementary roles and communication (Howes & Matheson, 1992; Rubin, Bukowski, & Parker, 1998). Child participation in socio-dramatic play demonstrates increasingly sophisticated abilities to share symbolic meanings through social pretense across the early childhood years (e.g., Howes, 1980; Howes & Matheson, 1992; Johnson & Ershler, 1981; Rubin et al., 1978). We should also note that research has incorporated information about individual, cultural, and contextual variability influencing the development of play (Smith, 1978; Tamis-LeMonda & Bornstein, 1991).

Researchers have documented that toddlers and preschool-age children demonstrate curiosity and take initiative (National Research Council, 1999; Goswami, 1995). The research on "theories of mind" indicates that children intentionally learn and:

> "they can develop theories of what it means to learn and to understand that affect how they function in situations that require effortful learning (Bereiter and Scardamalia, 1989). The more they understand what the learning process requires—that it is not simply a matter of knowing or not knowing, of performing well or of failing to perform—the more directed they will be toward the learning goal (Dweck, 1989; Dweck and Elliott, 1983; Dweck and Leggett, 1988). Researchers have documented cases of preschool children using strategies to remember (Wellman et al., 1975; DeLoache et al., 1985), to count (Siegler, 1988) and to solve problems" (National Research Council, 1999).

Engagement and persistence at the preschool level, and pursuit of understanding and task persistence at the school-age level are measured on the DRDP-R. Young children do many things simply because they want to do them. This relates to "intrinsic motivation" (e.g., selecting a toy to play with or a shirt to wear). The child makes her *own* choice and achieves satisfaction from both the act of choosing and from the action following that choice. Children also engage in some activities because adults tell them to, or in an effort to please another party. These activities are "extrinsically motivated." When a child is extrinsically motivated, the reward comes from outside. Intrinsically motivated activity is more rewarding in and of itself, so that children tend to learn more from this sort of activity, and they retain that learning better. Intrinsically motivated children are more involved in their own learning and development.

A number of behavioral characteristics are indicators of high motivation. Persistence is the ability to stay with a task for a reasonably long period of time. The more motivated children are, the longer they can stay involved in an activity. Children learn persistence when they are successful at a challenging task. Choice of challenge is another characteristic of motivation. Children who experience success in meeting one challenge will become motivated, welcoming another. When children successfully complete such a task, children gain a high level of satisfaction. Children who have not experienced early success tend to pick something that is very easy to ensure instant success, but being aware that the task offered little challenge, they usually feel a very low level of satisfaction.

The amount of dependency on adults is another indicator of motivation. Children with strong intrinsic motivation do not need an adult constantly watching and helping with activities. Children who have a lower level of motivation or are extrinsically motivated need constant attention from adults. The last indicator of motivational level is emotion. Children who are clearly motivated will have a positive display of emotion. They are satisfied with their work and show more enjoyment in the activity. Children without appropriate motivation will be less willing to engage in proposed activities or initiate them on their own.

One way to tie all these different elements together is to approach motivation from a constructivist learning perspective, where knowledge is seen as actively built up by the individual (Watson, 2000). Stemming from this perspective, one could assume that, to build knowledge, young children must be highly involved in the learning process to ensure that effective learning takes place. Laevers (1993) defined involvement as “a quality of human activity, characterized not only by a high level of motivation, but also by concentration and persistence, intense perceptions and experience of meaning, a strong flow of energy and a high degree of satisfaction” (p. 61). This view recognizes the importance of the intrinsic motivation for young children’s learning and educational achievement, and this view is widely stressed by others such as Ames (1992), Deci and Ryan (1980, 1985), Dweck (1999), Dweck and Leggett (1988), and Heyman and Dweck (1998). Children armed with this internal drive become what Dweck (1986) referred to as “mastery” learners—that is, learners who are challenge seeking, who persist in the face of difficulty, and who enjoy “exerting effort in the pursuit of task mastery” (p. 1040). As a consequence, this model supports the idea that fostering a positive disposition toward learning (i.e., developing an environment in which children are fully motivated and actively absorbed in the learning process) is as important as developing young children’s knowledge and skill acquisition (Katz, 1993, 1995, 1999).

2.1.5 *Domain: Mathematical Development (MATH)*

The Mathematical Development domain is represented by one Indicator of the same name. The Math indicator is summarized as: Children show interest in real-life mathematical concepts. The mathematics domain on the DRDP-R is comprised of measures that allow children to demonstrate their level of competence with everyday mathematical concepts. Subjects considered part of mathematical development for infants and toddlers include a beginning understanding of numbers, space and size, time, and classification and matching. Preschool children should have a more sophisticated understanding of these areas, plus begin to understand measurement and patterning. School-Age children should have a more sophisticated understanding of these areas than preschool children.

Mathematics understanding is addressed because early mathematics skills are shown to predict future academic success (Duncan et al., 2007) and because children have the interest and ability to engage in significant mathematical reasoning and learning (Sarama & Clements, 2004). Mathematics is a foundational academic subject that preschool children will learn formally during their school-age years. Research suggests that children start developing number sense in early infancy (Clements, 2004; Feigenson, Dehaene, & Spelke, 2004), and much of what preschool children know about numbers is closely related to and depends on their understanding and mastery of counting (Kilpatrick, Swafford, & Findel, 2001).

Children's understanding of numbers is initially qualitative, as they gain an understanding of numbers with small quantities, using subitizing (i.e., visually knowing 'how many' are in a set without actually counting them (Clements, 2004; Fuson, 1988, 1992a). The three major basic building blocks for counting are learning of 1) the sequence of number words, 2) one-to-one correspondence, and 3) cardinality (knowing that the last number assigned to the last object counted gives the total number in the set) (Becker, 1989; Clements, 2004; Fuson, 1988, 1992a, 1992b; Hiebert, 1997; National Research Council, 2001; Sophian, 1988). Research shows that very young children (ages up to 3) may be able to handle small quantities first (groups of 2 to 3), and as they grow older, they are more likely to be able to manage larger sets. Cardinality is typically developed between the ages of 3 and 4 years (Fuson, 1988). Children's early experience with number operations and with counting are a key factors in children's development of number sense (Clements, 2004; Hiebert et al, 1997; Kilpatrick, Swafford, & Findell, 2001; NCTM, 2000). Children as young as 3 years are able to understand simple visual number patterns that involve number operations such as, "two fingers and two fingers make four" (Fuson, 1988, 1992a). Young children initially understand a quantity as an aggregate of single units (Fuson, 1988; 1992a, 1992b; Carpenter and Moser, 1988; Hiebert et. al, 1997; Geary, 1994). Along with counting, children are exposed to a variety of geometric principles including those concerning basic shapes.

The developmental trajectory for the composition of geometric figures indicates that children begin by being able to use shapes individually to represent objects. This progresses to outlining with shapes, and eventually being able to combine shapes without needing an outline and make shape units (i.e., smaller shapes that make a larger shape which is itself a part of a larger picture) (Clements, 2004; Clements & Sarama, 2000).

The conception of time for children goes from the sequencing of events to thinking of time in terms of hours or minutes. Piaget, Inhelder, & Szeminka (1960) found that young children can understand duration, elapsed time, or succession of events in relation to their daily routine or their own age. Greenes (1999) recommends that time explorations for young children should focus on comparisons of amounts of time required to complete various tasks and the terminology necessary to express those comparisons (i.e., more time, less time, the same amount of time). She also recommends that preschoolers learn to describe duration (i.e., long time, short time) and temporal sequencing (i.e., what happened first, second, last, before or after). The DRDP-R addresses the concept of time in the ways research suggests.

The term "classification" focuses on the child's ability to categorize, compare, match, and sort objects into groups according to a common attribute. It is a building block for data analysis and problem solving in mathematical competence (Ginsburg, Inoue, & Seo, 1999). It is defined as the systematic arrangement of objects into groups according to established criteria, and involves sorting, grouping, and categorizing. Classification is at the heart of identifying what is *invariant* across groups of mathematical objects or mathematical processes. Clements (2004) suggests that analyzing, comparing, and classifying objects helps create new knowledge of objects and their relationships.

Measurement is a mathematical process that involves assigning numbers to a set of continuous quantities (Clements & Stephen, 2004). It is a number that indicates a comparison between the attribute of the object being measured and the same attribute of a given unit of measure. For example, the expectation is that three-year-olds are ready to measure in non-standard units, by connecting number and quantities in everyday situations, while four-year-olds begin to make progress in reasoning about measuring quantities with less dependence on perceptual cues (Clements, 2004b, Clements & Stephen, 2004). The research base regarding children's development of measurement concepts is richer regarding children's ability to make general comparisons in measurements rather than the development of children's ability to use specific measurement procedures (Clements & Stephen, 2004). To understand the concept of measurement, children must be able to decide on the attribute of objects to measure, select the units to measure the attribute, and use measuring skills and tools to compare the units (Clements, 2004; Van de Walle, 2001). Children should understand the different units that are assigned to physical quantities such as length, height, weight, volume, and nonphysical quantities such as time, and temperature (Smith, 2001). Generally, children first learn to use words that represent quantities or magnitude of a certain attribute—such as longer, taller, shorter, the same length, holds less, holds the same amount (Greenes, 1999). Then, they begin to demonstrate an ability to compare two objects directly and recognize equality or inequality. Last, children learn to measure, which requires them to connect numbers to attributes of objects, such as length, weight, amount, area, and time (Clements, 2004a; Ginsburg, Inoue, and Seo, 1999). Measurement on the DRDP-R suggests a child shows (increasing) understanding of measurable properties such as length, weight, and capacity and begins to quantify those properties.

While the concept of measurement emphasizes the idea of measurable properties, the patterning indicator emphasizes recognition, reproduction, and creation of patterns of varying complexity. Patterns help children learn to find order, cohesion, and predictability in seemingly disorganized situations. Some researchers contend that mathematics itself consists of the study of patterns (Devlin, 1988; Steen, 1990). The recognition and analysis of patterns clearly provide a foundation for the development of algebraic thinking (Clements, 2004). Patterns involve replication, completion, prediction, extension, and description or generalization (Greenes, 1999). Young children gradually develop a concept of patterns that includes recognizing a pattern, describing a pattern, extending a pattern, and finally creating a pattern. To understand a pattern, children should identify similarities and differences among elements of a pattern, note the number of elements in the repeatable group, identify when the first group of elements begins to replicate itself, and make predictions about the order of elements based on given information. The developmental trajectory of patterns has been characterized as developing from 3 year old children's ability to identify repeating patterns, to 4 year old children's ability to engage in pattern duplication and pattern extension (Klein and Starkey, 2004). The perception of the initial unit plays a fundamental role in both the duplication and extension of patterns. Developers of the DRDP-R considered these mathematics themes and consulted with experts and practitioners to develop the mathematics measures on the DRDP-R.

2.1.6 *Domain: Motor Development (MOT)*

The Motor Development domain is represented by the Motor Skills Indicator. This indicator is summarized as: Children demonstrate an increased proficiency in motor skills. On the Infant/Toddler DRDP-R motor development is measured in terms of gross motor, fine motor, balance, and eye-hand coordination. On the Preschool DRDP-R the same areas are addressed except for eye-hand coordination, while the School-Age DRDP-R measures gross motor skills in terms of movement and coordination, and fine motor skills in terms of dexterity.

There is reluctance in the literature to assign ages to specific motor skills because these skills require extended practice over time and are accomplished within broad age ranges, depending on factors such as exposure, practice, heredity, nutrition, and access to health care. Gross motor skills involve moving the whole body and major parts of the body such as arms and legs. They include locomotor skills, such as walking and stretching. Children develop various gross motor skills, while moving in and exploring their environment, therefore, program standards require a safe and supportive environment for this purpose. When children can coordinate their movements, they are ready to learn more advanced gross motor skills, such as riding tricycles; doing somersaults; and catching, throwing, and kicking a ball. One important aspect of gross motor skills development is maintaining balance. Through their exploration, children learn to balance themselves while coordinating other movements. Generally, the direction of muscle development is from head to toe, and the sequence of development begins with muscles closest to the center of the body and progresses to those in the extremities. Thus, children refine their gross motor movements before they can control fine motor skills (McDevitt & Ormrod, 2002; Sanders, 2002).

Fine motor skills involve use of the small muscles found in individual body parts, like hands and feet. Children use their fine motor skills to grasp, hold, and manipulate small objects and tools. As they gain eye-hand coordination, they learn to direct the movements of their fingers, hands, and wrists to perform more complex tasks. With access to appropriate materials and activities, children can practice and refine their fine motor skills during a variety of activities and while performing self-help routines (Battelle, 1988).

As a part of motor development children grow a sense of balance. As they age, they show ability to balance themselves in increasingly complex settings. They also increase in sophistication with regard to fine-motor skills. These skills can be observed through children’s work with small-sized objects (Lerner and Hultsch, 1983).

The concepts covered under the motor development domain are taken from standard program and Head Start practices, and are aligned with the Physical Education Framework (NASPE, 1995; Sanders, 2002; Salkind, 2002). General feedback about these measures (from practitioners and experts) was that they are highly observable and appropriate for programs.

2.2 Summary

Research presented in this chapter was used to develop and/or refine the measures on the DRDP-R. The measures were grouped by the six indicator sets or domains and discussed at this level. The six domains are (1) Self and Social Development, (2) Self Regulation and Self Care, (3) Language and Literacy, (4) Cognitive Development, (5) Mathematical Development, and (6) Motor Development. An extensive literature review was conducted to develop the measures so that they would be representative of the domains, and be based on leading child development research for the different age groups. Researchers in early-childhood education and developmental psychology and seasoned practitioners in Infant/Toddler, Preschool, and School-Age care programs were an integral part of the development of the DRDP-R. Measurement researchers reviewed the materials as well, to investigate and align measures and levels to be consistent with expectations.

3 DRDP-R 2005 Calibration Study

3.1 Calibration Study Design

In the spring of 2005, the DRDP-R assessment system was tested across the state of California. This section describes the rationale for the design of the study, which we called the "Calibration Study." The following sections describe the methodology used and the data collected.

The objective of the Calibration Study was to calibrate the three DRDP-Rs together and to provide data to enable the investigation of the reliability, validity and fairness of the instruments. In order to calibrate the instruments, the study was designed to collect rating data on a large sample of children. The data provided valuable evidence for scaling the developmental levels of each indicator, within, and across age-groups. A representative sample of children across age groups, gender and ethnicity was gathered so that the calibrated instruments could be used to make valid inferences about the population. Moreover, the representative sample was important to the evaluation of the sensitivity of the instrument to ethnicity and gender.

A secondary goal of the study was to examine the differences and similarities between ratings made by raters who have primary contact with the child (e.g., teachers, caregivers), versus raters who have only secondary contact with the child (e.g., teacher aids, administrators). This aspect of raters' agreement is important because primary caregivers may choose to delegate some of the assessment responsibilities to the secondary contact persons. We need to understand the effects of this as it relates to deploying the DRDP-R. Our expectations for the Calibration Study are to have only moderate agreement between primary and secondary raters. See Chapter 7 on inter-rater effects.

Because of the developmental perspective of the DRDP-R, there is a need to address the articulation between age-contiguous instruments (e.g., IT and PS, PS and SA). Many CDD programs serve children whose ages collectively span two or more instruments; the continuity of our developmental measurement indicators from one instrument to the next has to be clear and obvious to teachers. A confirmation that this requirement is satisfied would be to find empirically that, in the main, a teacher marks the developmental level a child has reached on a measure at consistent points on contiguous instruments. Establishing this continuity is also essential in constructing scales for our indicators that span the entire DRDP-R age range. Establishing this linkage structure across instruments is important evidence to support the calibration of the DRDP-R assessment system. See Chapter 7 for an account of the Secondary Study.

The Calibration Study, then, has four goals, which are discussed at length in subsequent chapters:

- Calibrate the three DRDP-Rs.
- Study the developmental linkage structure across the three DRDP-Rs.
- Examine the sensitivity of the DRDP-Rs to ethnicity and gender.
- Evaluate the inter-rater reliability of the DRDP-R.

To accommodate these goals, the Calibration Study included three conditions under which teachers rated children on the DRDP-R.

Condition 1 (Primary rater)**:** children were evaluated once by their primary caregiver, using the instrument appropriate for their age group. Condition 1 provided evidence of how well the DRDP-R measures capture the development of infant/toddlers on the indicators of the Desired Results (DR).

Condition 2 (Double raters)**:** each child was evaluated twice, once by the primary caregiver and once by a secondary caregiver such as a teacher aid or an administrator. Both raters used the appropriate age-group instrument. The difficulty with condition 2 was to find two caregivers who both knew a child well enough to bring sufficient experience with the child to their observations.

Condition 3 (Double instruments)**:** each child was evaluated twice by the primary caregiver, once using the appropriate age-group instrument, and once using the instrument of the next age group. For example, a child from the IT group was evaluated using the IT and the PS instruments. The ages of children rated in this condition were around the transition points between programs (and hence between instruments), around 3 years old and around 5 years old. Condition 3 provided evidence of the association between age-contiguous instruments. This information helped evaluate the instruments' validity.

Note that Conditions 2 and 3 contain data sets that augment Condition 1– a child rated by his primary teacher, using an age-appropriate DRDP-R. Therefore, the main sample used for calibrating the instruments contained all "primary and age-appropriate" ratings from Conditions 1, 2 and 3.

3.2 Training Sessions and Data Collection

Data for the Calibration Study was collected by teachers from 140 state-funded center-based programs, preschools and school-age programs, sampled from throughout California[11]. More than 700 staff personnel were trained in observing children and completing the DRDP-R in the day-long seminars we conducted in 9 sites: Fresno, Los Angeles, Orange County, Pomona,

[11] Programs included in the sample for the DRDP-R studies were selected from a roster of all programs served by CDD by region within California and SES within region.

Riverside, Sacramento, San Francisco, San Mateo, and Santa Barbara. The training seminars were carried out by WestEd, California Institute on Human Services (CIHS) and Berkeley Evaluation and Assessment Research (BEAR) personnel and supervised by Child Development Division (CDD), Mathematica Policy Research (MPR) and American Institute for Research (AIR). Following the training seminars, teachers returned to their programs to observe and evaluate a sample of the children under their care (each teacher rated between one and six children – the median number of rated children per teacher was three). The DRDP-Rs were completed only for children who attended the program consistently for at least 10 hours per week in the previous month. In addition, only children that were under the rater's direct care for at least 30 days were rated. This way, we ensured that observers had ample time to know the child before completing his or her DRDP-R. The complete set of instructions for completing the DRDP is at the second and third page of each instrument. The DRDP-R instruments can be found at http://www.cde.ca.gov/sp/cd/ci/drdpforms.asp. Following the training sessions, the teachers had about 2-4 months to observe the child, complete the DRDP-Rs and return them to the study center. The teachers were compensated for their time and effort. Information about ongoing CDE-sponsored training for DRDP-R is available through these links: www.wested.org/desiredresults/training/resources.htm ; www.desiredresults.us ; www.cpin.us .

In addition to data collected using the DRDP-Rs meeting the three conditions, qualitative data was collected from caregivers and site administrators using interviews and surveys. Additional data of this kind was obtained for participants' feedback on the training, and the ease of use and clarity of the instruments. The information obtained from these questionnaires was used to inform the development team on the strengths and weaknesses of the instruments. The teachers were also asked to use the documentation section on the DRDP-Rs to provide evidence for their ratings. This information was also used in the further revisions and refinements of the instruments. For instance, mock examples used to exemplify developmental levels (under the initial versions of the instrument) were replaced with real ones.

3.3 Data

The data were comprised of a total of 2355 acceptable DRDP-Rs. Primary caregivers completed 1821 DRDP-Rs, 271 were completed by secondary raters (Condition 2), and an additional 263 were completed using the age-consecutive instrument (Condition 3). For the primary caregivers, there were 506 Infant/Toddler assessments completed by the primary rater; 610 Preschool assessments completed; and 705 School-age assessments completed. The analysis discussed in this chapter is based on the 1821 children who were rated by the primary raters. The numbers of completed DRDP-Rs are shown in Table 3.1.

Table 3.1 – Number of DRDP-Rs in the Calibration Study Data Set

	IT	PS	SA	Total
Rated by primary	506	610	705	1821
Rated also by secondary	145	84	42	271
Rated with age-consecutive instruments	90	122	51	263
Total number of DRDP-Rs				2355

145 DRDPs were not used in the analysis because they failed to meet one or more of the criteria specified for inclusion. DRDP-Rs were excluded from further analysis if a child was in care less than 10 hours per week (Hours), if a child's age was out of the target range (Age), or if a child was rated only by the secondary rater (Secondary) and not by the primary rater. The frequencies of reasons for exclusion in the study by each age group are given in Table 3.2.

Table 3.2 – Frequency of Reasons for Discarding DRDP-Rs from the Calibration Study Data Set

Reason	IT	PS	SA	Total
Hours	39	31	46	116
Age	0	7	7	14
Secondary rater	2	11	2	15
Total	41	49	55	145

Overall, there were 49% males and 51% females in the sample of 1821 children considered in the main analysis. Figure 3.1 shows the distribution of ages of the children rated with each age-specific DRDP-R. Within each program, children were rated using the appropriate instrument.

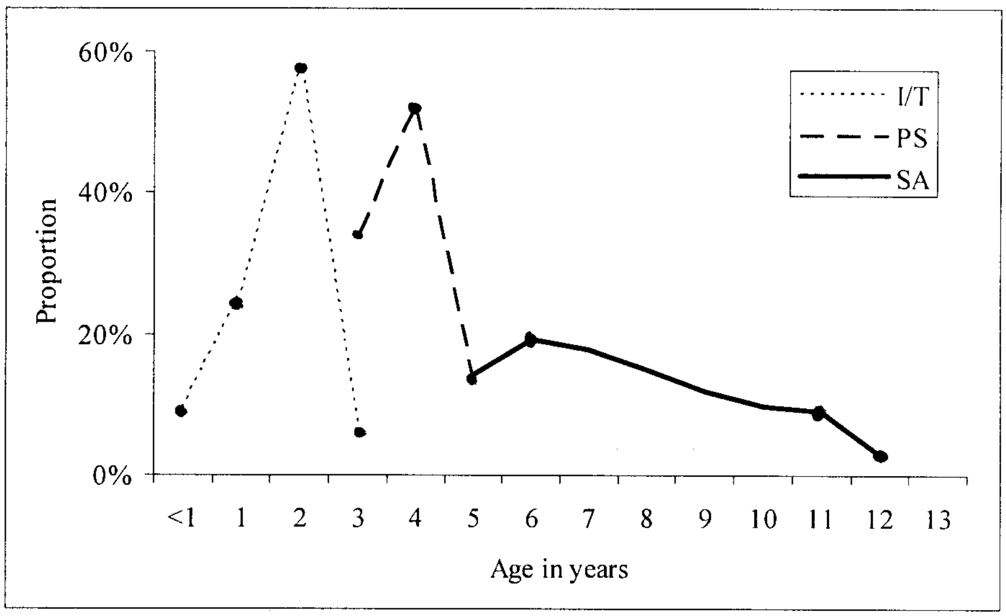


Figure 3.1—Age distribution of children within the three DRDP-Rs.

For 70% of the sample, English is spoken at home, and for 41% of the children, Spanish is spoken at home. Children's ethnicity was reasonably representative of the California population at this age – 57% Hispanic-American, 18% European-American, 10% African-

American, 4% Asian-American and 11% of other ethnicities compared to the 2008-2009 California school-age enrollment statistics available on line from the California Department of Education's data and statistics site. Table 3.3 lists the distribution within our sample of gender, ethnicity, language spoken at home, and number of weekly hours with caregiver. Results are presented separately for each age group and over all groups.

Table 3.3 – Demographic Distribution of the Calibration Sample (N=1821)

		IT	PS	SA	Overall
Gender	Female	51%	51%	50%	51%
Ethnicity	African-American	14%	11%	7%	10%
	Asian-American	4%	4%	3%	4%
	European-American	18%	12%	22%	18%
	Hispanic- American	50%	59%	59%	57%
	Other	14%	12%	8%	11%
	Missing	0%	1%	1%	1%
Language Spoken at Home	English	59%	50%	54%	54%
	Spanish	19%	28%	30%	26%
	Other	3%	5%	2%	3%
	Bilingual[a]	20%	17%	14%	16%
Number of weekly hours with child	10 or less[b]	0%	6%	1%	2%
	11-20	10%	84%	69%	58%
	21-30	13%	1%	24%	14%
	31-40	59%	6%	5%	20%
	40+	18%	1%	1%	6%

[a]Child speaks English and Spanish or English and another language. [b]Most DRDP-Rs that indicated 10 or less hours were discarded from the data set. A few DRDP-Rs were kept because they were part of a pair in Condition 2 or 3.

4 Calibration of the DRDP-R

4.1 Introduction

The main goal of the Calibration Study was to gather evidence about how well the DRDP-R measures the six indicators of Desired Results (DR). An essential step in this process is the calibration of the DRDP-R instruments, which will establish scales for the three instruments within each measured domain.

The Introduction to the Calibration Study provides the theoretical background for the calibration process. The Method section describes the technical details of the calibration. The Results section provides information about children's ability estimates, the measures' difficulty estimates, fit statistics and DRDP-R's reliability. The Conclusion section consists of an argument about the validity of the DRDP-R instruments in assessing the DR indicators, based on the calibration results.

4.1.1 *Measurement Model*

Assessments are designed to measure abilities, proficiencies or traits that cannot be directly observed. Every assessment is an interaction between a person and an item. Both items and persons are considered to be located on latent variables to be estimated from the data. The analysis of test data using Item Response Theory (IRT) is an attempt to jointly estimate items' characteristics and person abilities (introductions to IRT can be found in Hambleton, Swaminathan, & Rogers 1991; and Embretson & Reise 2000).

To analyze the data from the Calibration Study, we fitted a Multidimensional Partial Credit Model (MD-PCM) (RCML; Adams & Wilson, 1996), which is an extension of the Rasch Simple Logistic Model. This model allows us to separately estimate multiple latent abilities. This formulation estimates the difficulty of attaining each developmental level separately for each measure. The model was fitted using *ConQuest* (Wu, Adams, & Wilson, 1998). The Method section describes the technical details of the measurement model used to calibrate the DRDP-R.

4.2 Method

4.2.1 *Calibration Analysis*

ConQuest (Wu, Adams & Wilson, 1998) is a computer program for item response modeling based on the Random Coefficients Multinomial Logit Model (RCML), using the Rasch family of IRT models (Adams & Wilson, 1996). Each DRDP-R measure is assumed to follow a Partial Credit (PC) model for polytomously scored item response categories (Wright & Masters, 1982). That is, the developmental levels within a given theme are seen as analogous to response categories in an open-ended test item where the various responses might receive varying degrees of credit. Thus, if 5 responses were written so that only one is completely wrong, three others are partially right to increasing degrees, and only the remaining one is completely correct, then the scoring scheme applied might be 0, 1, 2, 3, 4. This is called polytomous scoring.

The analogy to DRDP-R developmental levels is straightforward. If the observer marks a child as having reached only the lowest developmental level on a particular theme, that response is scored 1; a mark at the next level up is scored 2, and so on. This scoring scheme is reflected in the **B** matrix in Equation 1. Note that DRDP-R allows an observer to record that a child is not yet at even the lowest developmental level on the instrument. That response is associated with a zero score[12].

[12] Because the “not yet” response category cannot be used for the IT population, an arbitrary low θ value was used to anchor the scaling procedure in this group.

Of course, there is an important difference between giving a child a set of number scores for his responses to some polytomous items on a math test, and giving a child a set of number scores that corresponds to the developmental-level observations a teacher has made for the child on the measures within the DRDP-R SELF/SOC domain. One way to think of this difference is in terms of how one would interpret the θ value estimated for the child, using Equation 2. If the θ value comes from using scores on the math items as one of the latent dimensions in the θ vector in Equation 1, then it would be an estimate of his math proficiency at the time he took the test. If, however, the θ value comes from using scores corresponding to the developmental levels in, say, the DRDP-R SELF/SOC domain as one of the latent dimensions in the θ vector, then it would be an estimate of how far along the developmental continuum (as defined for overall Socio-Emotional development), the child had reached at the time the DRDP was completed for him. As applied to DRDP-R, the θ vector in Equation 2 has an element for each one of the DRDP-R dimensions (domains). So the set of DRDP-R observations for a child is the basis for a θ value for that child on each of the dimensions defined in the θ vector.

We now describe the model assumed to hold for each individual DRDP-R dimension within RCML—i.e., the Partial Credit (PC) model. The PC model is commonly expressed according to Equation 2 below:

$$P\left(X_{nij} = s \mid \delta, \theta\right) = \frac{\exp \sum_{j=0}^{s} (\theta_n - \delta_{ij})}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (\theta_n - \delta_{ij})} \qquad (1)$$

Here, θ_n is the proficiency (or location) of the n^{th} child on the dimension defined for a single DRDP-R indicator. Each δ_{ij} is an item parameter corresponding to the difficulty of attaining the j^{th} developmental level (compared to the $j\text{-}1^{th}$ developmental level) on measure i.

The specific analysis used in this study linearly decomposes the δ_{ij} to an average item difficulty and individual step difficulties. More information about the measurement model used in this analysis can be found in the ConQuest manual (Wu, Adams, & Wilson, 1998).

When calibrating the instruments, the PC formulation of RCML model was used on a sub-sample of the DRDP-Rs collected in the Calibration Study. The general RCML formulation for the probability of a response vector, **x**, is:

$$P(\mathbf{x};\xi \mid \theta) = \frac{\exp\left[\mathbf{x}'(\mathbf{b}\theta - \mathbf{A}\xi)\right]}{\sum_{\mathbf{z}\in\Omega}\exp\left[\mathbf{z}'(\mathbf{b}\theta - \mathbf{A}\xi)\right]} \tag{2}$$

The θ term in Equation 2 is a vector of θ values, each indicating a person's location along a single dimension. Applied to DRDP-R, every indicator set is represented by a single θ vector. In IRT analysis, the symbol θ is often used to represent a person's ability or proficiency level as a location along some latent variable, construct, or dimension. For the DRDP-R analysis, θ represents children's scores on an indicator set. The equation defines the probability of attaining a certain response pattern given known item parameters and conditioned on θ.

In Equation 2, **ξ** is the vector of item parameters which represent the difficulty of a getting a set of ratings on the DRDP-R measures. Both z and x represent a response vector which is the set of ratings a single child received on his or her DRDP-R. The denominator in Equation 2 is a normalizing value used to calculate the probability of observing response vector x out of all possible response vectors, represented by Ω. The Scoring Function, **b**, is used to construct the θ component of the probability equations while the Design Matrix, **A**, is used to construct the **ξ** component. There would be a column in **A** for each item parameter. **A** and **b** will always have the same number of rows, with one row for each response category.

The next section describes the calibration sample data in more detail.

4.2.2 *Calibration Sample*

The data used for calibrating the instruments consisted of observations on 1821 children, across the three age groups. A sub-sample of the children was rated in condition 3 using two age-consecutive instruments. Thus there are more responses than children in the sample. That is, the total number of DRDP-R responses used in this analysis was 2084, which included 263 children who were rated twice.

Figure 4.1 shows a schematic presentation of the data structure we used for the calibration analysis. Rows represent children' responses, and columns represents the IT, PS, and SA measures. Empty spaces represent structurally missing data.

<table>
<tr><th>Infant/Toddler</th><th>Preschool Age</th><th>School Age</th></tr>
<tr><td>Group 1</td><td></td><td></td></tr>
<tr><td></td><td>Group 2</td><td></td></tr>
<tr><td></td><td></td><td>Group 3</td></tr>
<tr><td colspan="2">Group 4</td><td></td></tr>
<tr><td></td><td colspan="2">Group 5</td></tr>
</table>

Figure 4.1—Schematic presentation of the data structure used for calibration analysis.

Table 4.1 shows the sample characteristics of the 1821 children divided into groups. Group 1, 2, and 3 were administered solely the DRDP-R for infant/toddler (IT), preschool age (PS), and school age (SA), respectively. Group 4 was administered both IT and PS instruments and was comprised of three-year-olds. Group 5 was administered both PS and SA instruments and was comprised of five-year-olds. Table 4.2 shows the number of missing measures in the 2084 age-specific DRDP-R booklets.

Table 4.1—Sample Characteristics for DRDP-R Calibration ($N_{children}$ = 1821)

	Group 1 – IT (*n* = 416)	Group 2 – PS (*n* = 488)	Group 3 – SA (*n* = 654)	Group 4 – IT+PS (*n* = 144)	Group 5 – PS+SA (*n* = 119)
Female (%)	50.0	50.6	50.9	55.6	47.1
Age in months					
Mean	23.71	51.32	100.19	37.26	62.31
SD	(8.72)	(6.70)	(23.16)	(5.07)	(9.10)
Range	1-36	32-65	62-156	30-51	49-104

Note. *n* represents the number of children in each group.

Table 4.2—Number of Missing Measures in DRDP-R Calibration Sample ($N_{DRDP\text{-}R}$ = 2084)

Instrument	*Mean*	*Median*	*SD*	*Min*	*Max*
IT (*n* = 560)	.43	0	1.18	0	19
PS (*n* = 751)	.62	0	1.91	0	26
SA (*n* = 773)	2.05	1	1.53	1	17

Note. *n* represents the number of DRDP-R responses.

4.2.3 *Estimation Procedure*

We conducted six unidimensional partial credit analyses with the data set described above (including all five Groups of data), one analysis for each indicator set. We analyzed all measures on a given indicator set, such as SELF/SOC, from the three instruments together to see how developmental levels from these instruments are spread out along the developmental continuum. We also analyzed each age group separately for each of the indicator sets. All models were estimated using the *ConQuest* program using a marginal maximum likelihood procedure. The mean person location was set to zero for each model to be identified.

We obtained six sets of item and person parameter estimates. We used the Weighted Likelihood Estimates (WLE) as a scale score (Wu, Adams, & Wilson, 1998). Results involving children's abilities are based on WLE estimates of Groups 1, 2 and 3 only. We excluded those of Groups 4 and 5 because these data were collected for the two special studies, and the estimates were derived from more than one instrument (i.e., both IT and PS instruments for Group 4 and both PS and SA instruments for Group 5).

A unidimensional measurement model was used to examine the data instead of a multidimensional model because using this model allowed for an examination of the dimensions separately.

4.3 Results

4.3.1 *Wright Map*

Figures 4.2 to 4.7 show Wright maps (Wright, 1977) of the six indicator sets. A Wright map helps illustrate how children and DRDP-R measures are jointly distributed along the ability continuum on the logit[13] scale. The ability histograms for three age groups, represented by horizontal lines of Xs on the left hand side of each ability scale, illustrate distributions of the estimated latent ability estimates (WLEs). An item threshold is indicated in a format of "p.q" on the right hand side of each ability scale. "p" represents a measure number, and "q" represents the developmental level. Thus, if the highest number of "q" is 4, this means that this measure has 5 response levels. The item threshold is defined as the point on the ability continuum where the probability of achieving at least the indicated level of performance on the measure is .50 (Wu, Adams, & Wilson, 1998). Within each measure, the thresholds are ordered with respect to their logit values.

Interpretation of the Wright map can be done jointly for children and measures. It can be said that a child has a probability of .50 to be rated at least at level *k* if his or her ability value is the same as that level's difficulty value. If the child's ability value is higher (lower) than the

[13] Technically, the logit is the log of the odds ratio. In this case, the logit is the log of the ratio between probabilities of correct and incorrect response. The logit scale is commonly used in psychometric research and can be easily rescaled to any other score range without loss of generality.

level's difficulty value, he or she are more (less) likely to be rated at that level. For example, the 10.1 in Figure 3.2 represents the threshold for Level *Exploring* in the *Identity of Self* measure of the preschool instrument. The person distributions on the SELF/SOC continuum, relative to this item threshold, suggest that almost all preschoolers and school age children are rated above this level on this measure, and by contrast, several infants and toddlers have not reached this level yet.

All Wright maps in Figures 4.2 to 4.7 show that item thresholds of three age groups as a whole cover the full range of the children's ability distributions. A set of item thresholds is more difficult as age increases: SA measures are generally more difficult than PS ones, which are generally more difficult than IT ones. Also, each set of item thresholds appear to spread over the corresponding age group's ability distribution, except for SELF/SOC and REG/SH of the IT group. For both indicators, the higher end of IT children is not covered by IT item thresholds, suggesting that IT measures do not range quite high enough to measure or differentiate much among children's abilities that are high on SELF/SOC. It appears that PS and SA measures cover the corresponding ability distributions very well. The Wright map of MOT provides a different picture (Figure 4.7). First, there is more overlap in ability distribution among the three age groups than the other indicators, and there is less differentiation within SA children than the other two age groups. Also, a substantial number of children are located on the highest end of the scale for all age groups, suggesting that MOT measures may be somewhat easy so that they do not measure or differentiate much among children's abilities that are high on MOT.

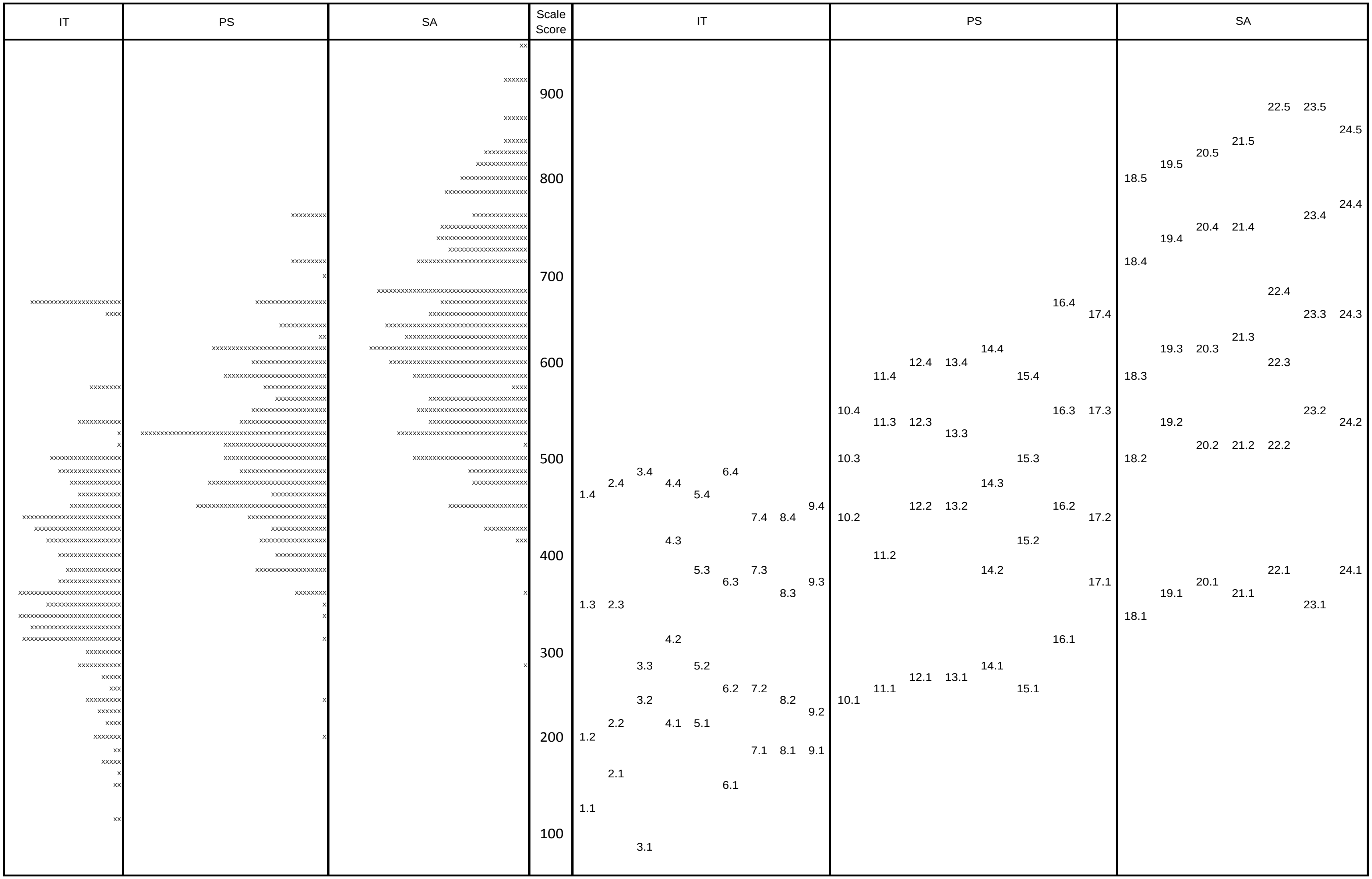


Figure 4.2—Wright Map for Self Concept and Social Interpersonal Skills.

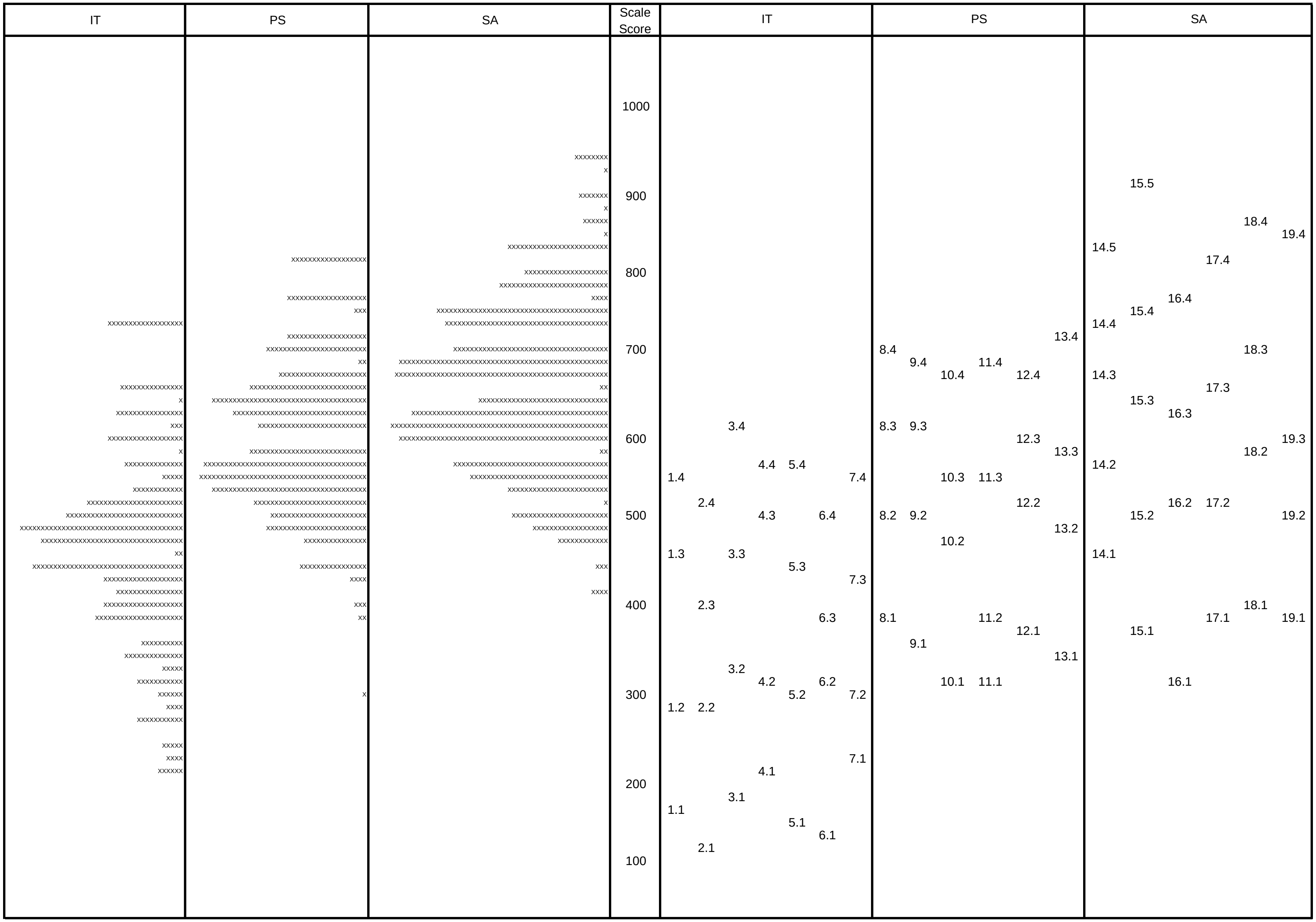


Figure 4.3—Wright Map for Self Regulation and Safety and Health.

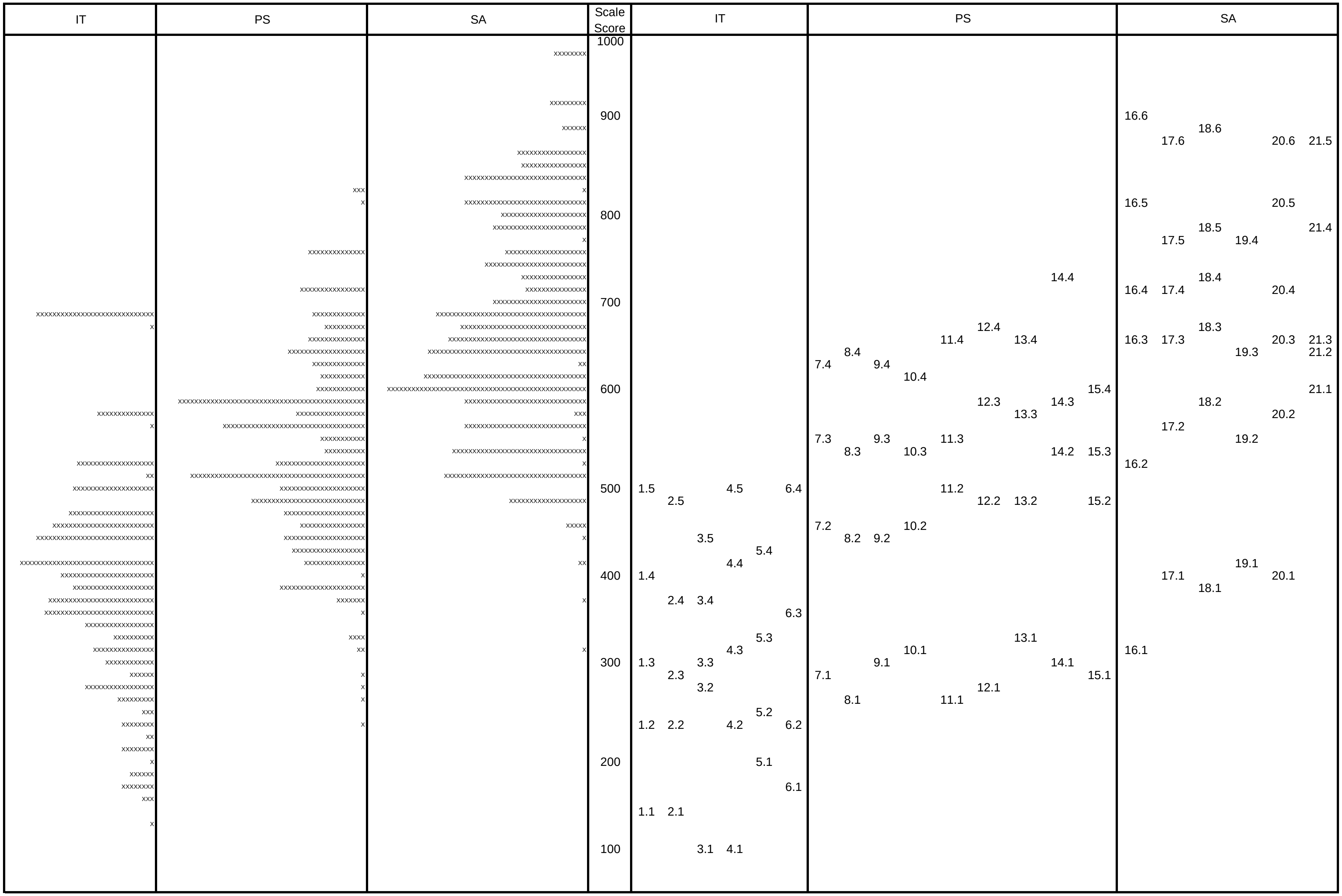


Figure 4.4—Wright Map for Language and Literacy.

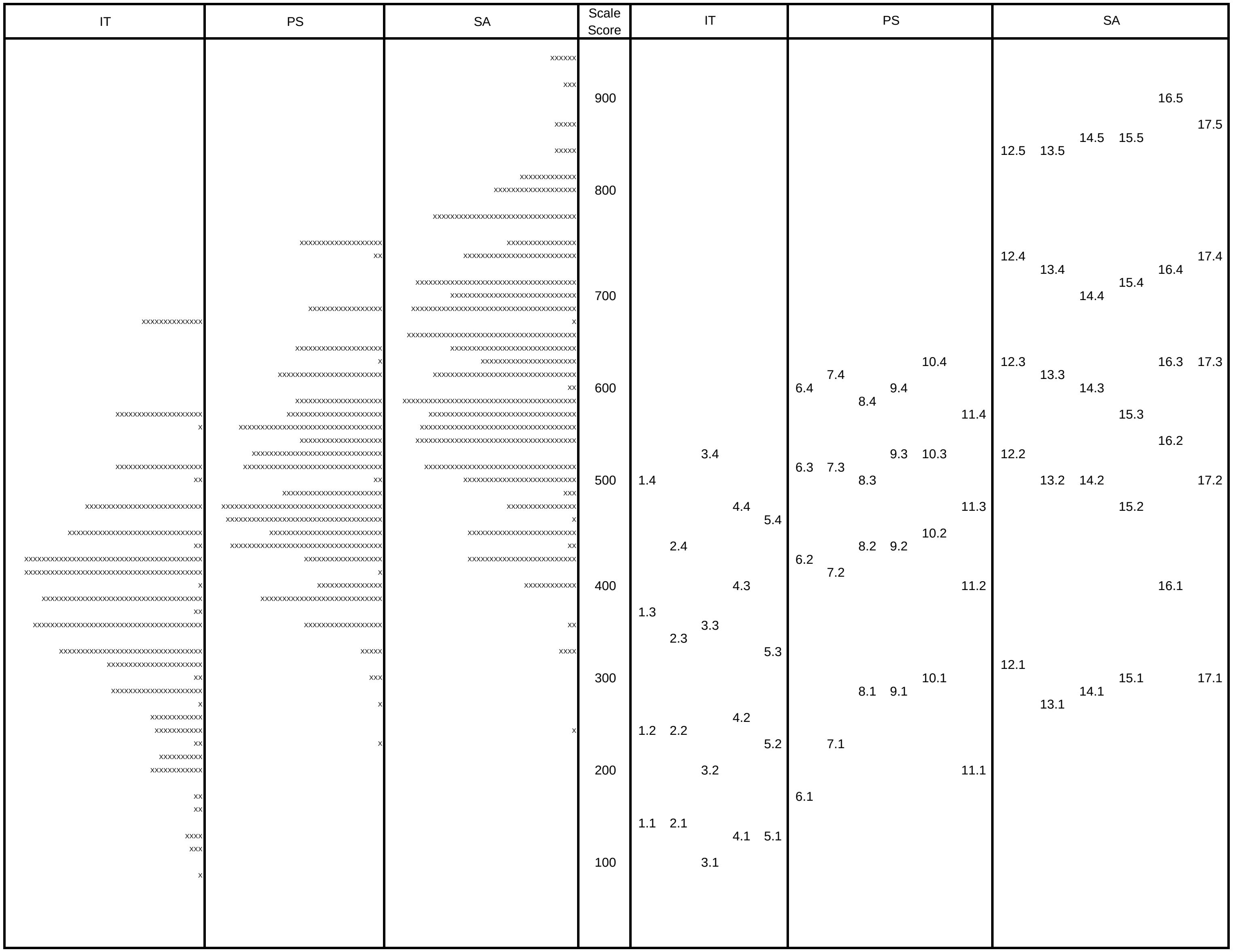


Figure 4.5—Wright Map for Learning and Cognitive Competence.

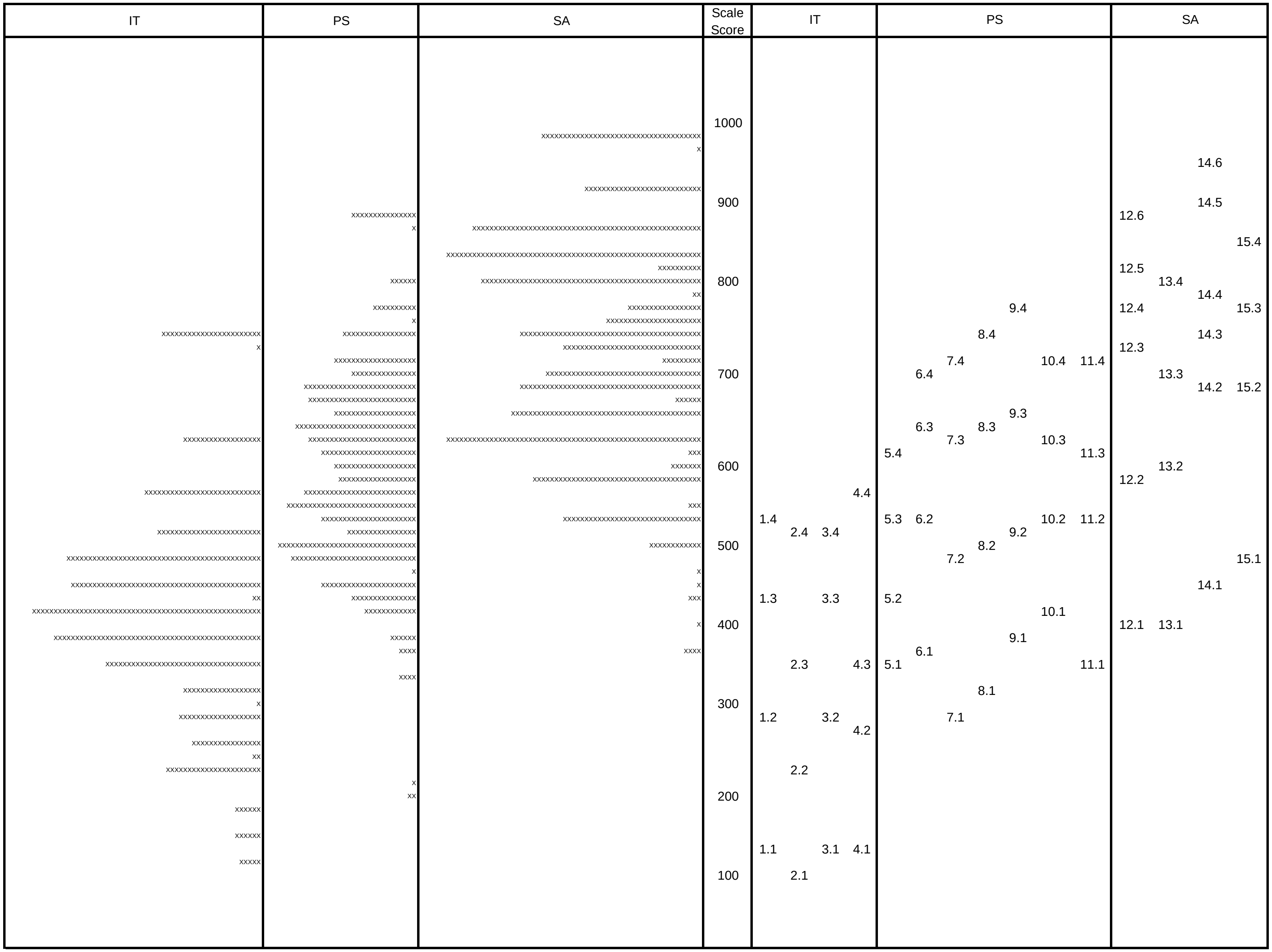


Figure 4.6—Wright Map for Math.

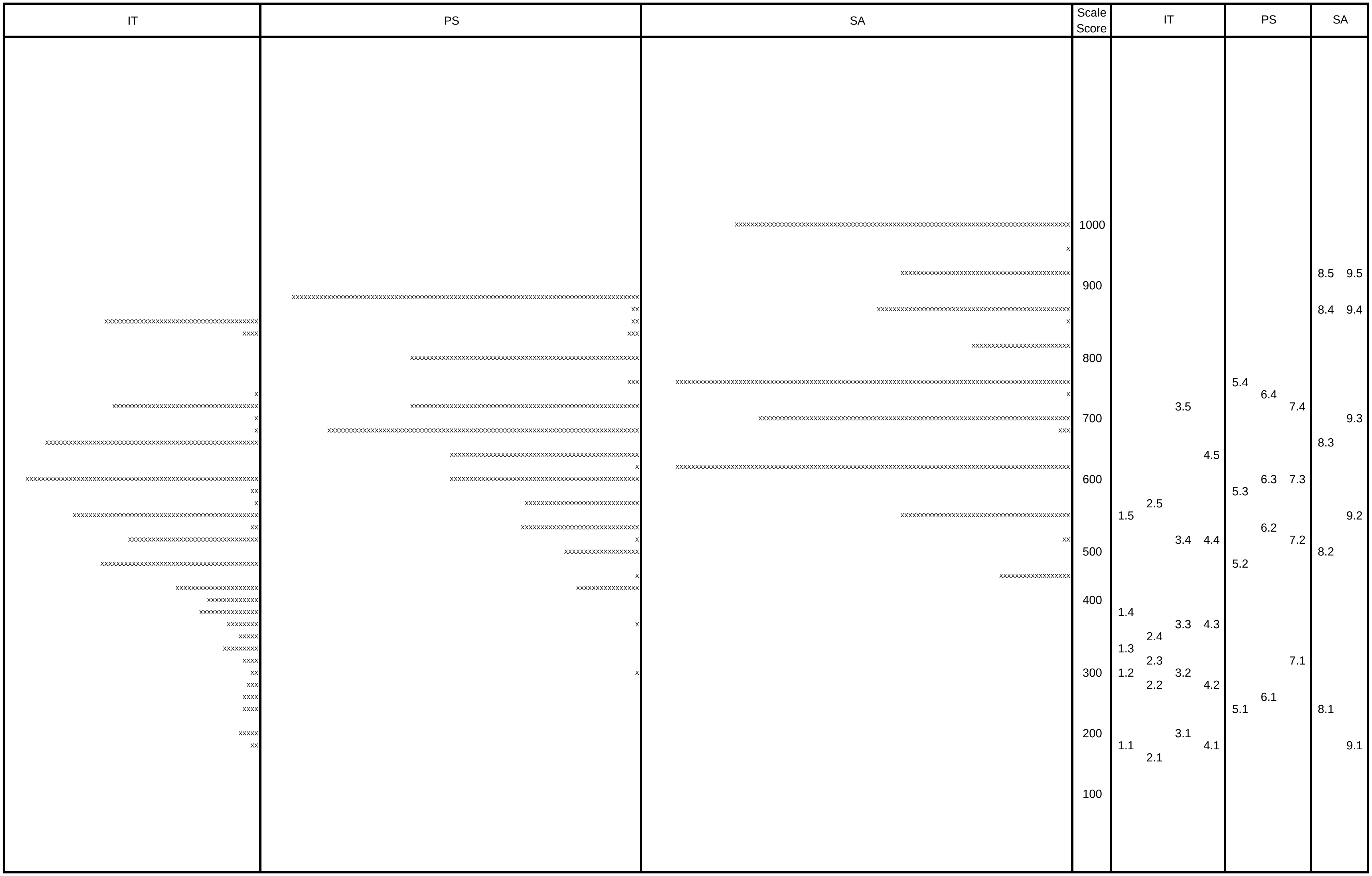


Figure 4.7—Wright Map for Motor Skills.

4.3.2 *Item Fit*

We inspected how each measure fits the model. As fit statistics, we used the weighted mean square (WMS) and the corresponding t-statistic (Wright & Masters, 1982; Wu, 1997). For each measure, the fit statistic compares the variability in the observers' ratings with that expected by the model, given the distribution of ability scores. When the observed variability is about the same as the expected, the WMS and the t-statistic are around 1 and 0, respectively. WMS values larger than 1 indicate more variability in the ratings than expected, suggesting that with the given measure, the teachers' ratings vary in a way that is somewhat inconsistent with model expectation. WMS values smaller than 1 indicate less variability than expected, meaning that with this measure, observers rate children in a way that is most likely conditionally dependent in some way. Generally, WMS between 0.75 and 1.33 and t-statistic between -1.96 and +1.96 indicate reasonable item fit (Adams & Khoo, 1996). We identified misfit measures when both statistics of a given measure were outside of the range cited above (Wilson, 2005).

Table 4.3 presents the number of measures identified as acceptable or problematic. The fit statistics of almost all measures fell within the acceptable range. Six measures from the IT instrument showed less variance in the ratings than expected: *Identity of Self and Connection to Others*, *Self Expression, Interactions with Adults, Responsiveness to Language, Communication of Needs, Feelings, and Interests, and Reciprocal Communication*. One measure from the SA instrument, *Exercise and fitness*, indicated more variance than expected. More consistency than the model expects is less of a problem than greater variability. Now, one measure with WMS out of range high (implying inconsistency) is no more than would be expected by chance in a set of items of this size hence we consider the total set to be within an acceptable range. More detailed information about item and step estimates and their fit statistics are available in Appendix E.

Table 4.3—Distribution of Item Fit Indices

Indicator	Age Group	Possible Conditional Dependence	Fit	Excess variability
SELF/SOC	IT	3	6	0
	PS	0	8	0
	SA	0	7	0
REG/SH	IT	0	7	0
	PS	0	6	0
	SA	0	5	1
LANG/LIT	IT	3	3	0
	PS	0	9	0
	SA	0	6	0
LANG/LIT	IT	0	5	0
	PS	0	6	0
	SA	0	6	0
MATH	IT	0	4	0
	PS	0	7	0
	SA	0	4	0
MOT	IT	0	4	0
	PS	0	3	0
	SA	0	2	0
Total		6	98	1

4.3.3 *Person Fit*

Similarly to item fit, we inspected how each child's response pattern fits the model. We employed analogous fit statistics as those used for measures (i.e., WMS and t-statistic). Person fit statistics measure the variability of observers' ratings as compared to their expected ratings based on the children's estimated ability and the estimated measures' thresholds. Interpretations of person fit are similar to those of item fit: Values above 1 (WMS) and 0 (t-statistic) mean more variability in the ratings than expected, suggesting that with a given child's estimated ability, the observer's rating pattern is somewhat more random. In contrast, values below 1 and 0 indicate less variability than expected, meaning that the observer's rating pattern is somewhat more consistent with the modeled pattern and hence does not allow for some random fluctuations expected by the model (and hence, could indicate conditional dependence). We used the same criteria as item fit to evaluate person fit.

Table 4.4 presents the percentages of children's estimated performance identified as acceptable or problematic. The fit statistics of almost all groups fell within the acceptable range (i.e., we expect no more than 5% either below or above this criterion). Relative to other indicators, school-aged children's performance in SELF/SOC show larger fluctuations (13%), as well as LANG/LIT and MATH (both 9%).

Table 4.4—Distribution of Person Fit By Percentage

Indicator	Group	Smaller	Fit	Larger
SELF/SOC	IT	3	97	1
	PS	3	96	1
	SA	0	87	13
REG/SH	IT	1	97	1
	PS	1	97	2
	SA	0	97	3
LANG/LIT	IT	0	99	0
	PS	5	92	4
	SA	1	96	3
LANG/LIT	IT	0	100	0
	PS	0	98	2
	SA	0	91	9
MATH	IT	0	99	1
	PS	2	97	1
	SA	0	91	9
MOT	IT	0	99	1
	PS	0	100	0
	SA	0	100	0

Note. Percentages out of age-group samples. Percentages may not add up to 100 because of rounding of third digit.

As can be seen on Table 4.4, the performance of school-aged children showed greater variation than either the infant-toddler or preschoolers data. These results suggest the results for the school-aged instrument in SELF/SOC, LANG/LIT, and MATH should be treated with caution.

4.3.4 *Reliability*

Table 4.5 presents the reliability of the six indicators. The EAP reliability as an internal consistency index which is the ratio between the ability variance based on Expected A Posteriori (EAP) values and the estimated variance of the ability latent distribution (Mislevy, 1984). Overall, the instruments show very high internal consistency. The lowest reliability, although still within acceptable range (.85), was found for MOT. Note that MOT has only 3 measures per child whereas other indicator sets average about 6 measures per child.

Table 4.5—Reliability of Six Indicators

	Number of measures	MML reliability
SELF/SOC	24	.99
REG/SH	19	.94
LANG/LIT	21	.99
LRN/COG	17	.99
MATH	15	.99
MOT	9	.85

4.3.5 *Standard Error*

The reliability of an assessment can also be evaluated through an examination of individual children's standard error. The standard error provides an index of the precision of measurement (i.e., the extent to which measurement is without error), at any given point on the development scale. Using standard errors, confidence intervals around an estimate can be created. For the DRDP-R, the standard error was estimated using the Expected A Posteriori measure (EAP). EAP estimation derives from Bayesian statistical principles. The term "a posteriori" derives from the Bayesian concept of a posterior probability. In this context it refers to a posterior probability distribution of latent trait scores—specifically, the predicted distribution of scores for a given case considering (a) the response pattern of that case, and (b) the estimated model parameters. The term "expected" derives from the concept of an expected value. Thus an "expected a posteriori" estimate refers to the expected value of the posterior probability distribution of latent trait scores for a given case. Generally, EAP estimates performed well in the middle range of the ability distributions, but not as well for extremely low and high ability distributions. Generally, we found that standard errors for EAP estimates were smaller in the middle range than at the tails of ability distributions. Because student abilities in the middle range of ability distribution matched the distribution of the items, more information was collected there, and, hence more precise abilities were estimated.

For the purpose of the DRDP-R, the aim is that each age-specific instrument be most sensitive at an ability range appropriate for that age group, with some overlap between instruments. This is because we expect the appropriate age-specific instrument to be administered to a child. Figures 4.8 to 4.13 show test standard error plots. Each figure contains three plots per indicator, one for each age-level instrument. Low standard error values mean that ability estimates at a given logit level are more accurate. Note that the scale of these distributions may differ for each age group.

SELF/SOC— Figure 4.8 shows that for IT, the ability estimates between -7 and -1 appear to be the most accurate in terms of the standard errors (range from 0.43 to 0.96). For PS, there seems to be a more distinct extreme of standard errors about -0.5 on the ability logit scale. It showed that the ability estimates between -2 and +2 appear to be the most accurate, and the standard errors range between 0.45 and 1.21. For SA, the ability estimates between 0 and 5 look to be the most accurate, and the standard errors are between 0.52 and 1.36.

REG/SH—Figure 4.9 shows that for IT, the ability estimates between -6 and -1 appear to be the most accurate, and that the amount of the standard error ranges from 0.52 to 1.09. For PS, there seems to be a more distinct extreme around 0. The ability estimates between -2 and +2 appear to be the most accurate, and the standard error is between 0.50 and 1.19. For SA, the ability estimates between -1 and 3 appear to be the most accurate, and the standard error ranges from 0.56 to 1.32.

LANG/LIT—Figure 4.10 shows that for IT, the ability estimates between -7 and -2 appear to be the most accurate, and that the standard error ranges from 0.46 to 0.93. For PS, there seems to be a more distinct extreme around 0. The ability estimates between -2 and +2 appear to be the most accurate, and the standard error ranges between 0.41 and 1.22. For SA, there seems to be a more distinct extreme around 3. The ability estimates between 1 and 5 look to be the most accurate, and the standard error is between 0.44 and 1.26.

LRN/COG—Figure 4.11 shows that for IT, the ability estimates between -6 and -2 appear to be the most accurate, and that the standard error ranges from 0.59 to 1.10. For PS, there seems to be a more distinct extreme around 0. The ability estimates between -2 and +2 appear to be the most accurate, and the standard error ranges between 0.50 and 1.24. For SA, the ability estimates between 0 and 5 appear to be the most accurate, and the standard error is between 0.55 and 1.41.

MATH—Figure 4.12 shows that for IT, the information plot looks a little flatter than the other plots. The ability estimates between -5 and -2 appear the most accurate, and the standard error is between 0.67 and 1.08. For PS, there seems a more distinct peak around 0. The ability estimates between -2 and +2 appear the most accurate, and the standard error ranges between 0.48 and 1.17. For SA, there seems to be a more distinct peak around 3. The ability estimates between 2 and 4 look the most accurate, and the standard error is between 0.54 and 1.31.

MOT—Figure 4.13 shows that unlike the other indicators, there are clear extremes for each age-group plot. For IT, the extreme appears around -5, and the standard error is between 0.46 and 1.01. For PS, there appears to be an extreme around -1, and the standard error is between 0.68 and 1.27. For SA, the extreme appears to be around 3, and the standard error is between 0.86 and 1.64. As they are close to both ends, the estimates are less accurate.

Overall, the standard error plots suggest that, as desired, each age-specific DRDP-R is suitably sensitive for the age range it was designed for.

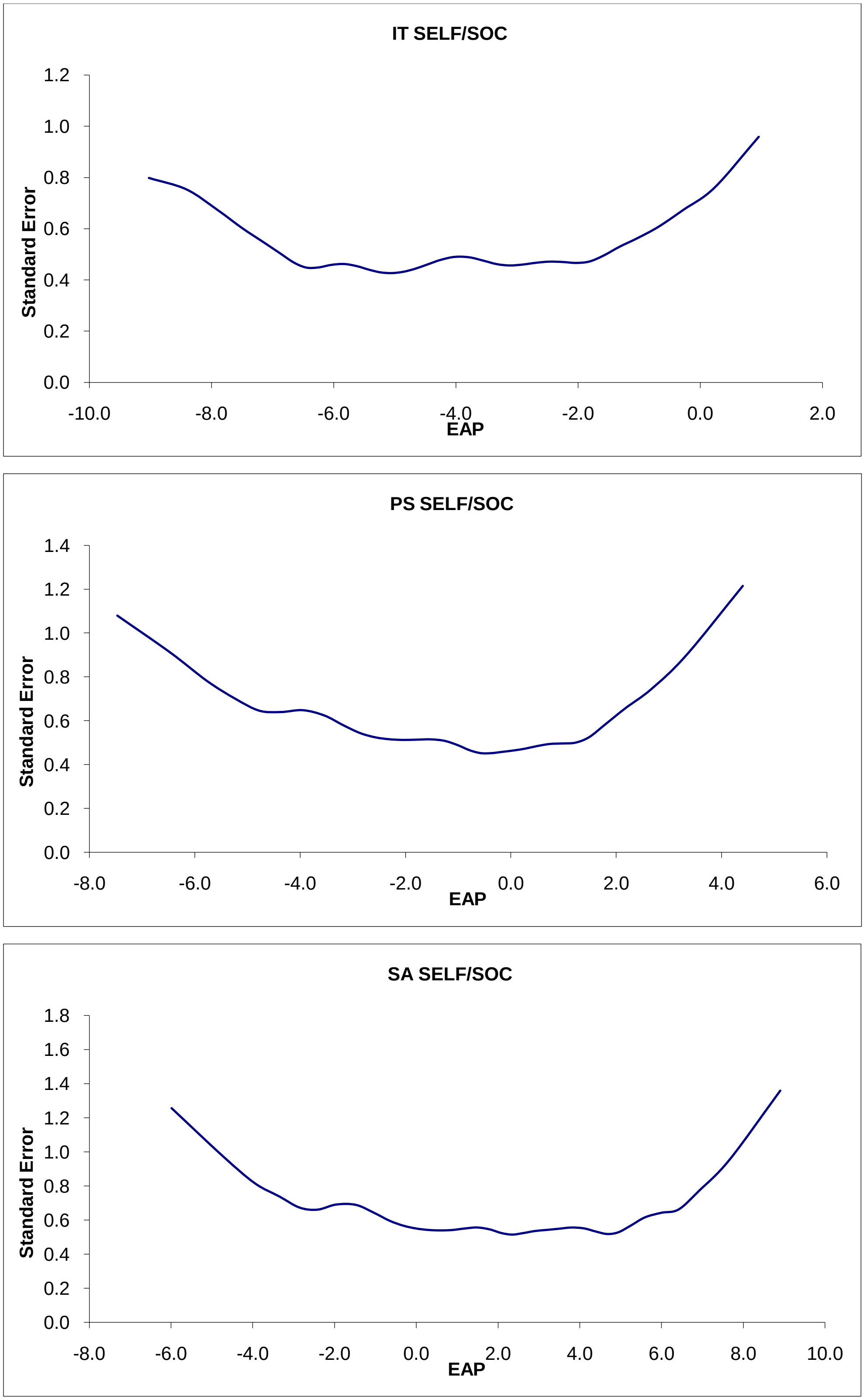


Figure 4.8—SEM plots for Self Concept and Social Interpersonal Skills.

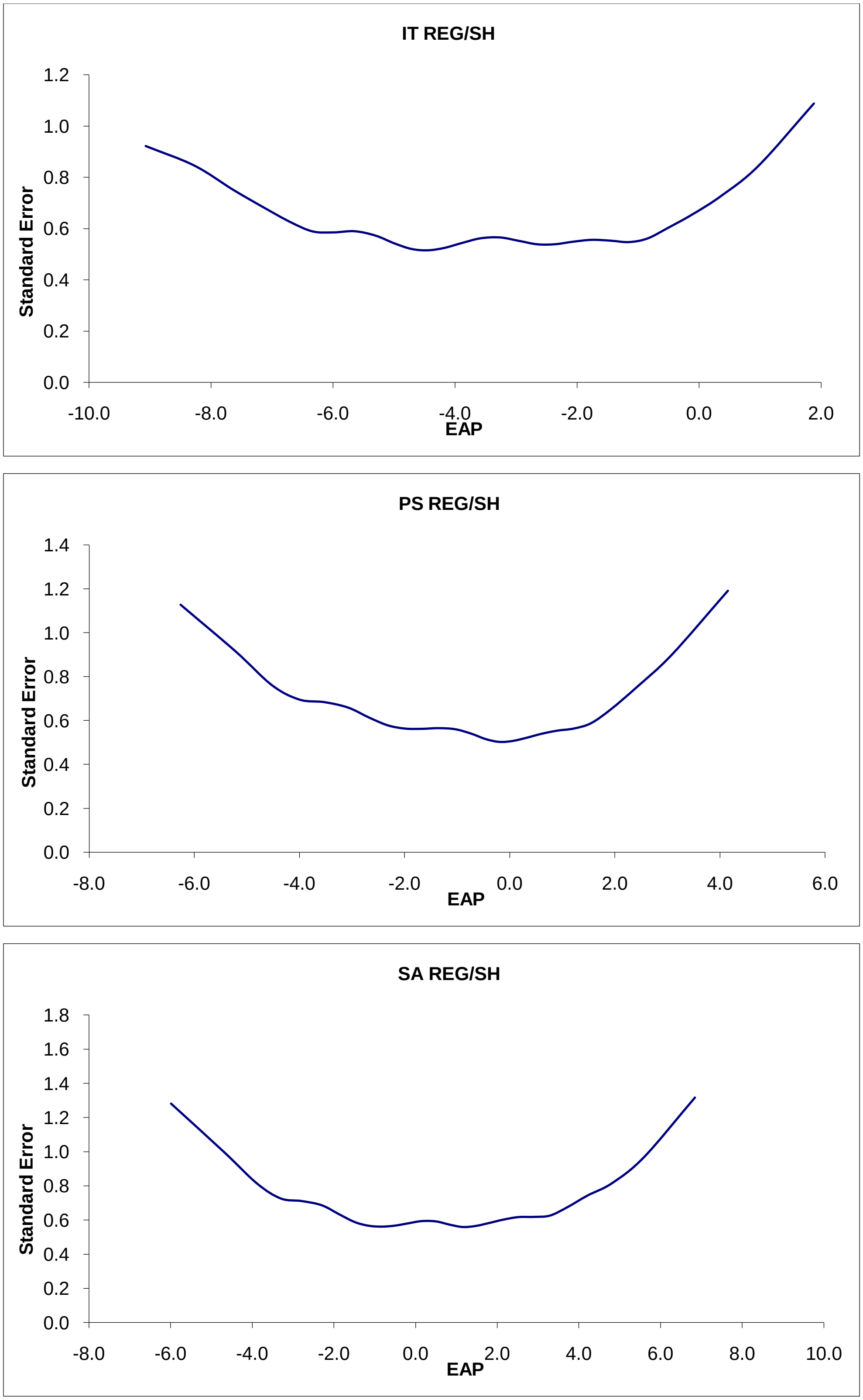


Figure 4.9—SEM plots for Self Regulation and Safety and Health.

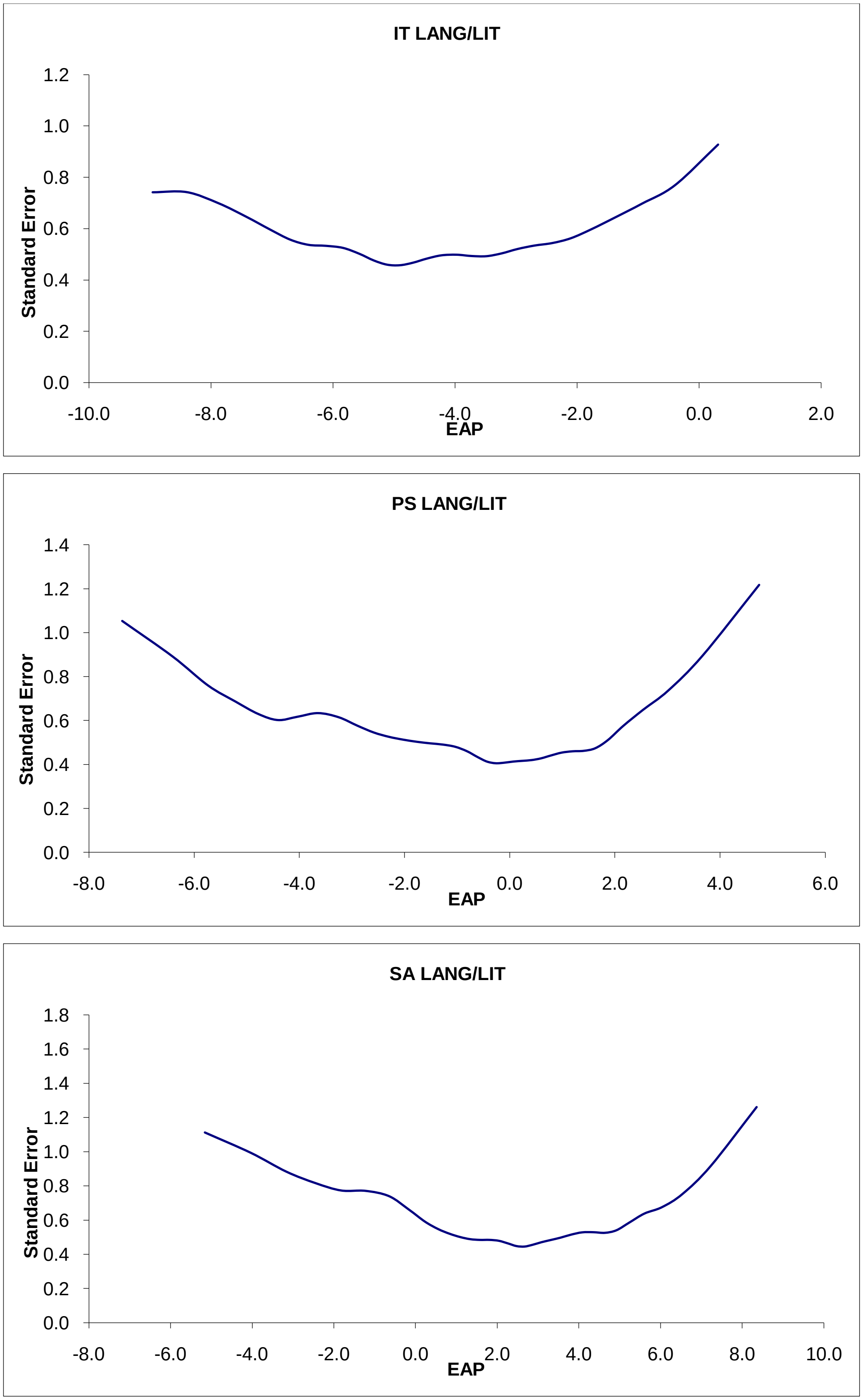


Figure 4.10—SEM plots for Language and Literacy.

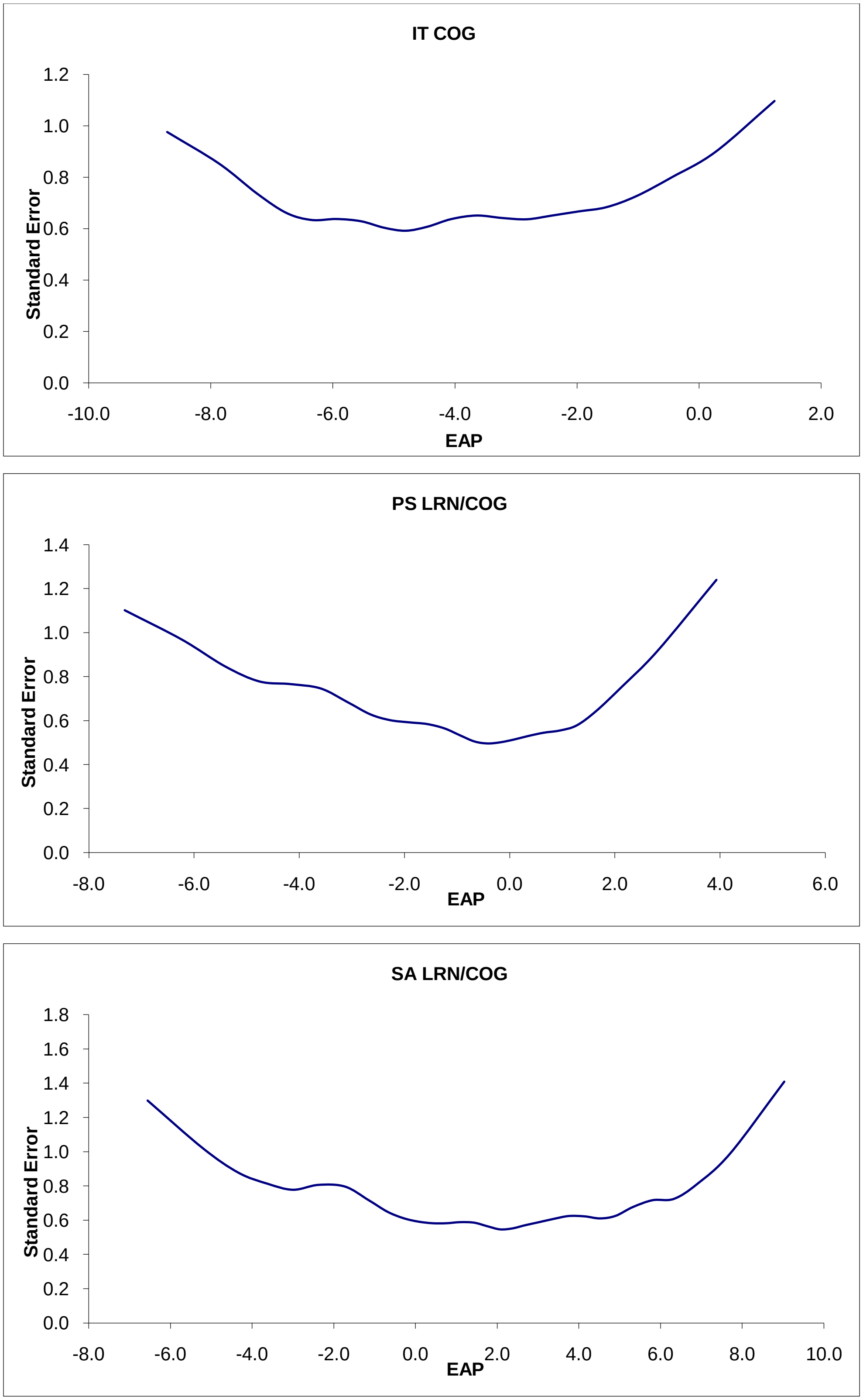


Figure 4.11—SEM plots for Learning and Cognitive Competence.

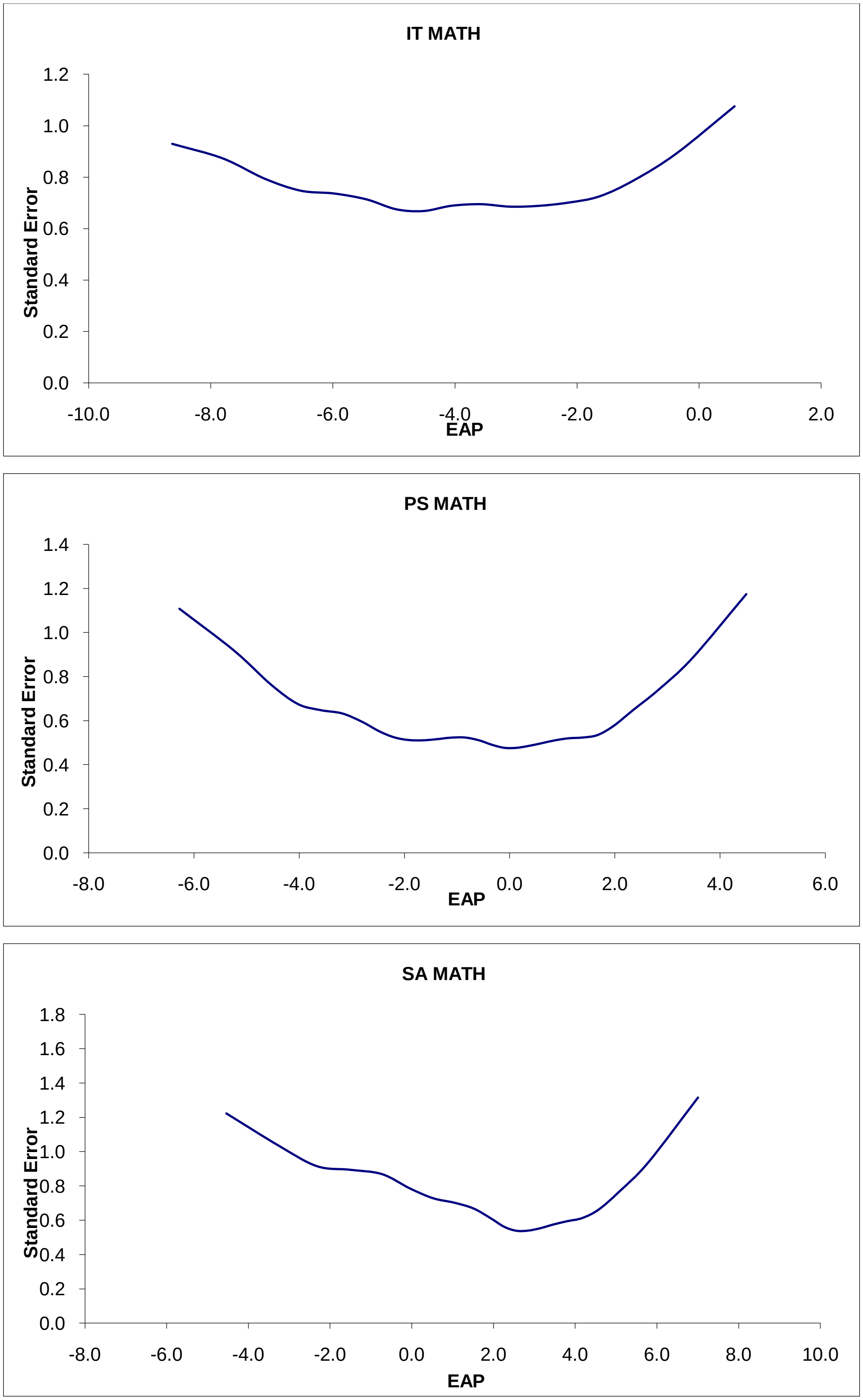


Figure 4.12—SEM plots for Mathematics.

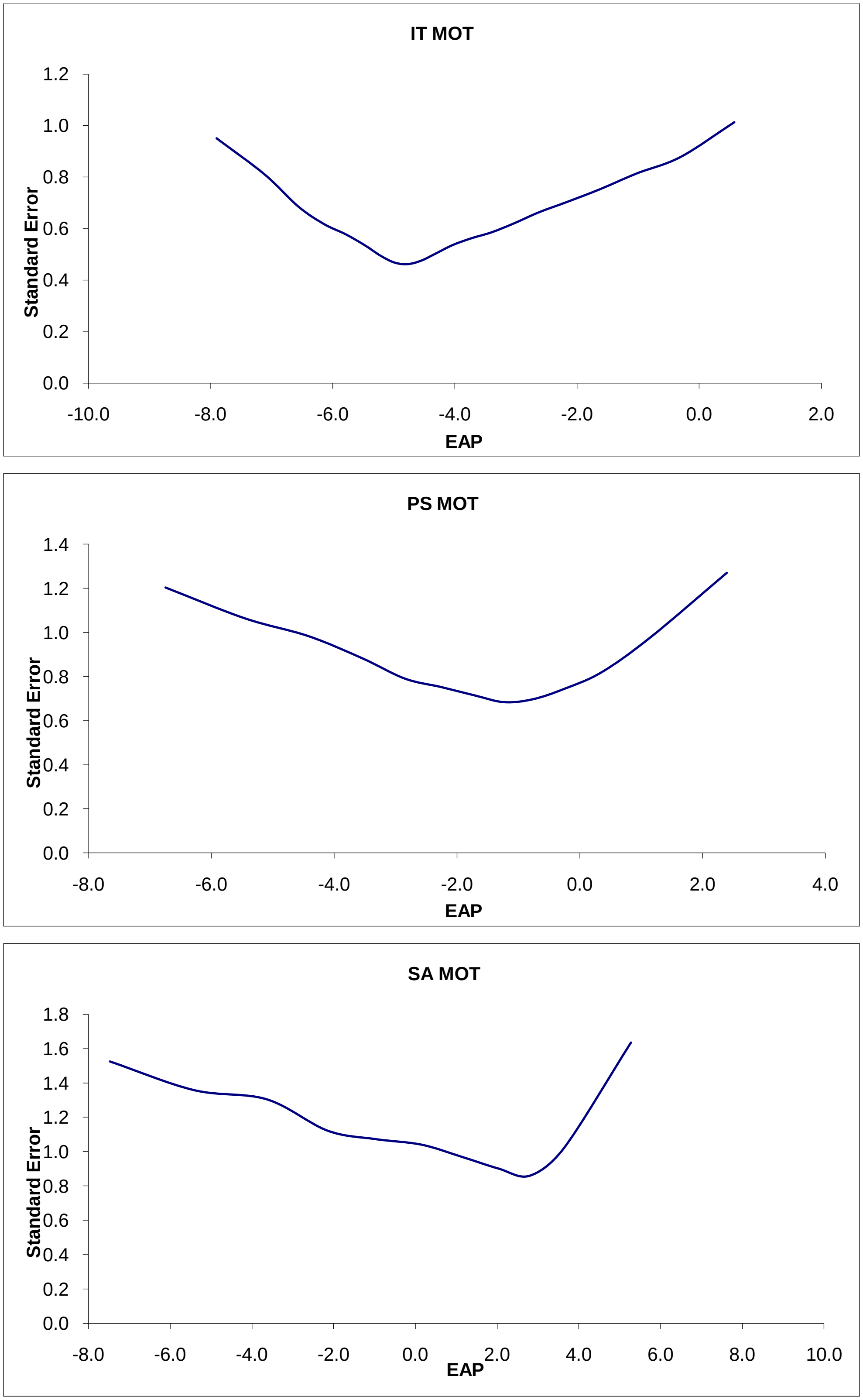


Figure 4.13—SEM plots for Motor Skills.

4.4 Conclusion

The main conclusion from the results discussed in this chapter is that the DRDP-R assessment system was reasonably well calibrated using the Calibration Study sample. The instruments are shown to have desirable technical properties that support the objective of validity evidence. The calibration analysis produced a usefully reliable set of estimates for developmental levels' difficulties which appropriately span the underlying developmental dimension across all three age groups. Moreover, the instruments may be used to obtain reliable estimates of children's abilities on six indicators of DRs so long as the standard procedures are followed. Consequently, we can proceed to examine validity evidence.

5 Internal Structures Validity Evidence for the DRDP-R

5.1 Introduction

According to the *Standards of Educational and Psychological Testing* (APA, AERA, NCME,1999, p.13), "analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based." To collect evidence based on the internal structure of an assessment, researchers "must first ensure that there is an intention of internal structure" (Wilson, 2005, p. 157). The BEAR Assessment System (Wilson & Sloane, 2000) provides a roadmap to follow to ensure adherence to an intentional approach to DRDP-R development. The steps of the BEAR Assessment System were followed to collect evidence regarding DRDP-R's internal structure.

Studies of the internal structure of DRDP-R were carried out to examine student ability distributions, internal correlations among domains, and evidence for differential item function (DIF). Differential item functioning "occurs when different subgroups of examinees with similar overall ability, or similar status on an appropriate criterion, have, on average, systematically different responses to a particular item" (p. 13). The analysis was completed using *ConQuest* (Wu, Adams, & Wilson, 1998).

5.1.1 *Ability Distributions*

As children grew older, we expect their location on the DRDP-R dimensions to increase. As the figures (5.1 – 5.6) show, the mode for each age range tends to increase as expected. These are essentially the same distribution shown in the form of Xs on the Wright maps in Figures 4.2 – 4.7. For example, consider the three distribution curves in Figure 5.1. As expected, below ability of about -2, most children are Infant/Toddlers; in the range of -2 to 2, most children are Preschoolers; and above 2, they are School-Age children. A similar pattern is seen in SELF/SOC, LANG/LIT and MATH. A similar but somewhat narrower pattern is seen for REG/SH, LRN/COG and MOT. Overall, the distribution shapes are all reasonably shaped, although some show multi-modalities (e.g., LANG/LIT for PS, or MOT for all the age groups). In particular, some distributions (mostly IT) show the existence of a "bubble" of children at the upper end of the group's distribution. For example, in Figure 5.1, a group of Infant/Toddlers are centered around the ability value of 3, separate from the rest of that age group.

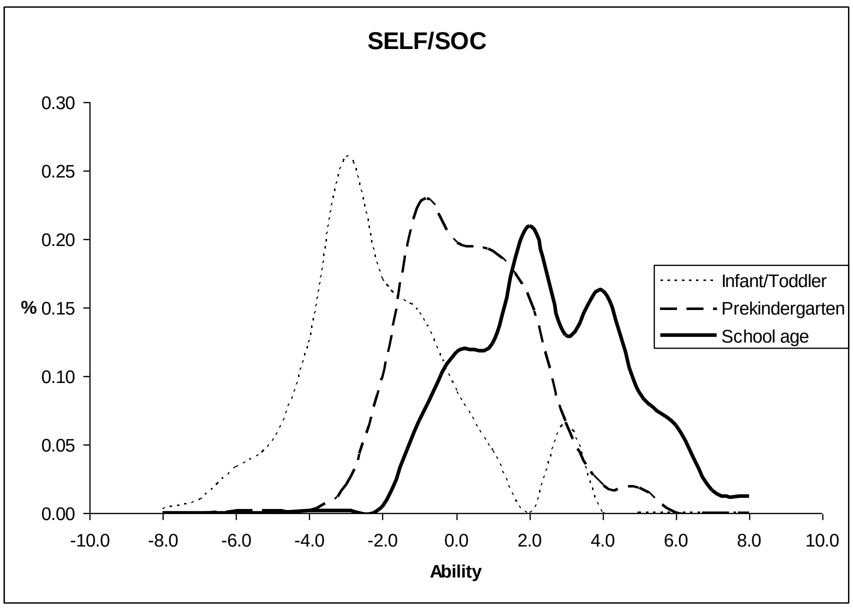


Figure 5.1—Ability distributions for Self Concept and Social Interpersonal Skills.

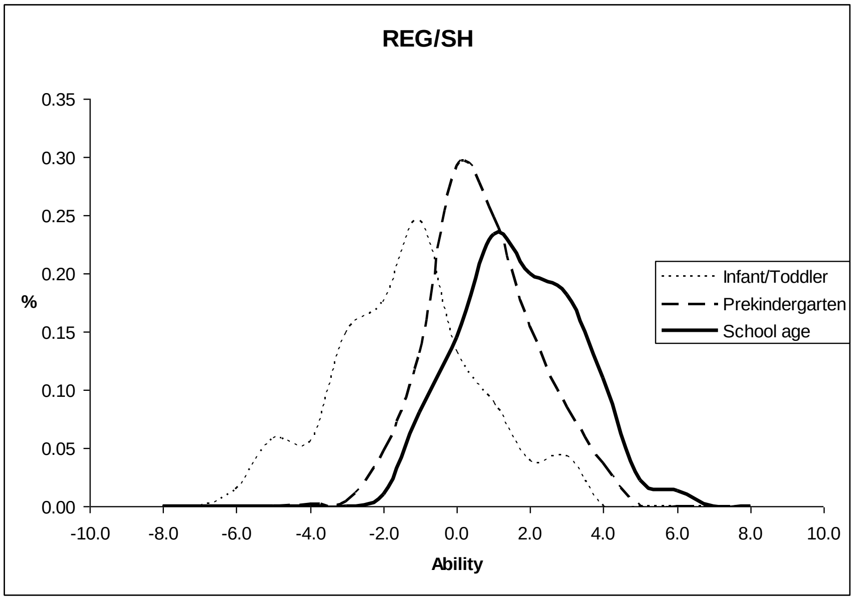


Figure 5.2—Ability distributions for Self Regulation and Safety and Health.

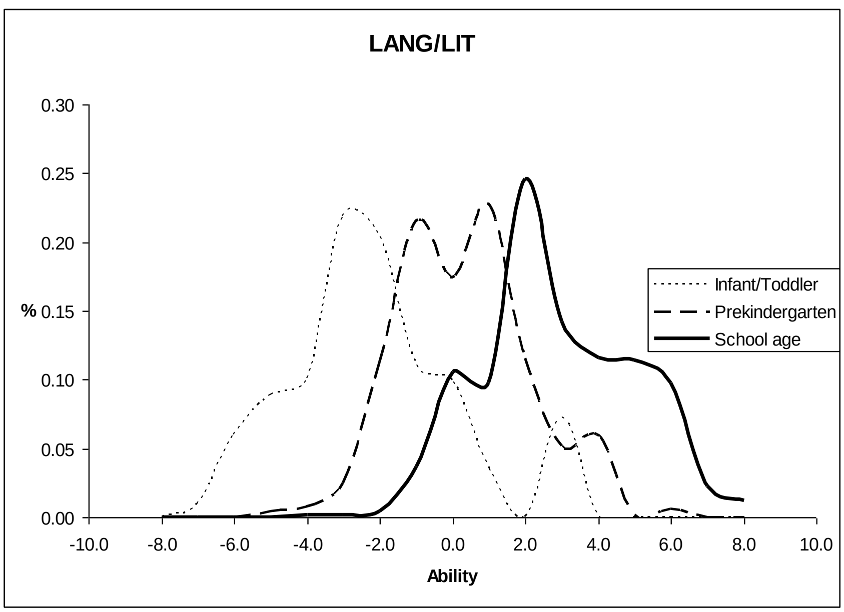


Figure 5.3—Ability distributions for Language and Literacy.

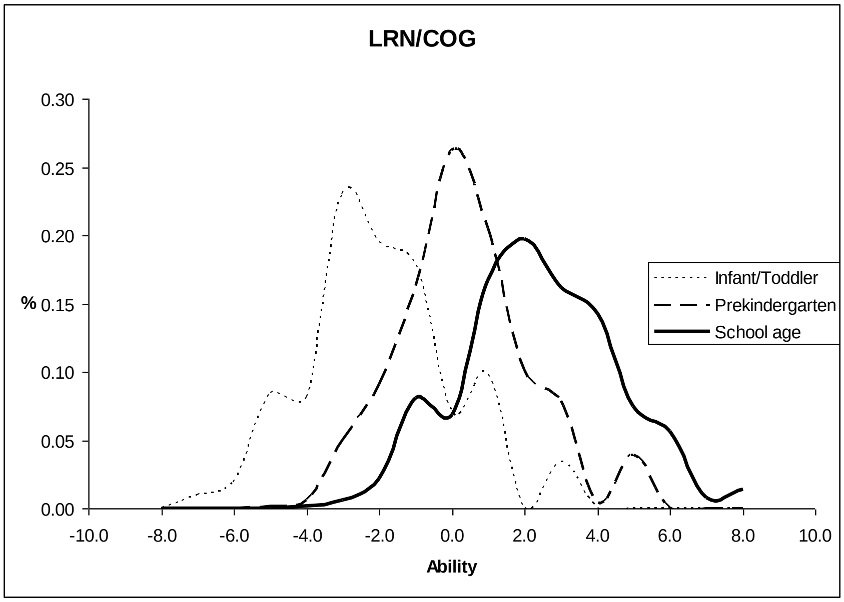


Figure 5.4—Ability distributions for Learning and Cognitive Competence.

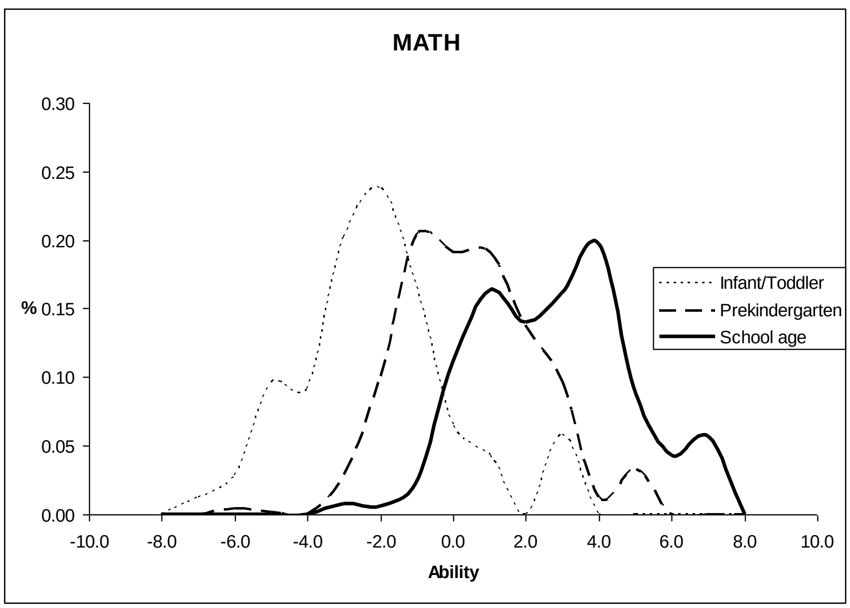


Figure 5.5—Ability distributions for Math.

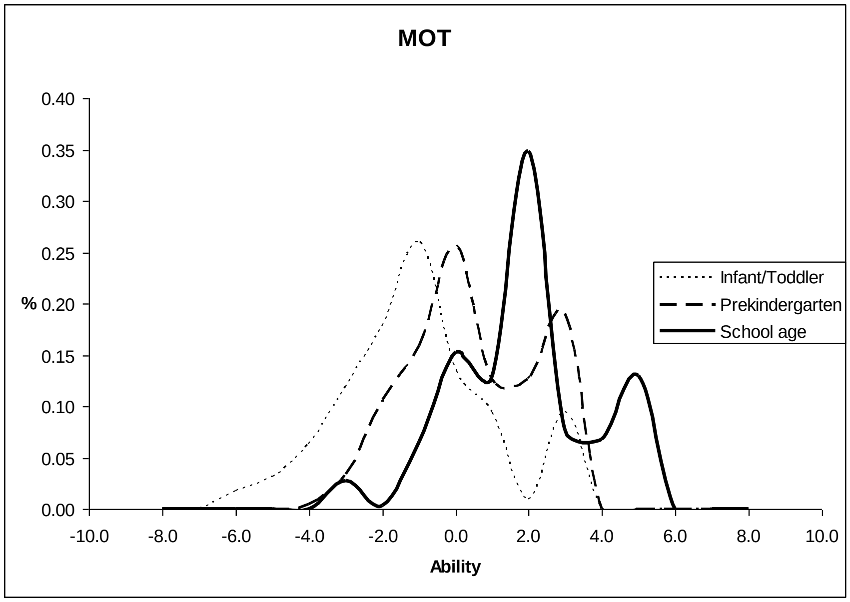


Figure 5.6—Ability distributions for Motor Skills.

5.1.2 *Correlation Information Among the Six DRDP-R Domains*

Table 5.1 shows the correlations between the six DRDP-R indicator sets. Overall, the correlations between the indicator estimates are moderate to high with the highest being between the two most academic of dimensions – Language and Literacy (LANG/LIT) and Mathematics (MATH) with an r of .90. The lowest correlation identified is that between Language and Literacy (LANG/LIT) and Self Regulation/Self Care (REG/SH) where r = .66. However, several other correlations were close. For example, Language/Literacy (LANG/LIT) and Motor (MOT) have an r of .68, and the correlation of Self Regulation/Self Care (REG/SH) with Mathematics (MATH) is .69. The reliabilities vary as anticipated, with dimensions that have the largest number of items having the higher reliabilities.

Table 5.1—Correlations Among the Six DRDP-R Domains

	LANG/LIT	MATH	LRN/COG	SELF/SOC	REG/SH	MOT
LANG/ LIT	1					
MATH	.90	1				
LRN/COG	.85	.83	1			
SELF/SOC	.83	.80	.83	1		
REG/SH	.66	.69	.70	.72	1	
MOT	.68	.70	.71	.68	.71	1

Note. Attenuated correlations. Disattenuated correlations are shown in the Appendix G.

5.1.3 *Comparison with Correlations Reported for the Measures*

The Programme of International Student Assessment (PISA) is a highly respected large-scale international study of academic achievement for 15-year-old students. Table 5.2 shows the 2003 PISA correlations in the four cognitive domains, math, reading, science, and problem solving. They fall in the range of r = .77 to r = .89. By age 15, student skills and abilities in these separate domains have had time to differentiate one from the other much more than the domains could be expected to differentiate at preschool age. Yet, the inter-correlations among corresponding DRDP-R scales (Language/Literacy, Learning/Cognition, and Math) are still quite similar to PISA results. Hence, the correlations among DRDP-R domain scores fit within an expected range.

Table 5.2—Correlations Among 2003 PISA Study Measures in Four Cognitive Domains

	Reading	Science	Problem solving
Mathematics	.77	.82	.89
Reading		.82	.82
Science			.78

5.1.4 *Differential Item Functioning (DIF)*

We examined differential item functioning (DIF) of gender (male vs. female) and ethnicity (African-American vs. European-American; Hispanic-American vs. European-American) for six indicators (SELF/SOC, REG/SH, LANG/LIT, LRN/COG, MATH, and MOT) on all three instruments (IT, PS, and SA). An item (i.e., DRDP-R measure) is flagged to exhibit DIF if it performs differently across the groups of interest (i.e., gender or ethnicity) for children who possess the same ability for a specific indicator (i.e., DRDP-R indicator). In other words, DIF is identified if the item obtained different item difficulties for two interest groups conditioned on the same ability for this indicator. For instance, an item is flagged to exhibit gender DIF if, on average, it is more difficult for females of a certain ability than for males of the same ability.

The software program *ConQuest* is used to perform the DIF analyses. Two partial credit models (Masters, 1982) were applied to examine DIF on the difference between item difficulties of items across groups. For each indicator on each instrument, we compared two models (Masters, 1982) with or without DIF parameters using the likelihood ratio test with degree of freedom equal to the difference of estimated parameters between two models:

No-DIF model: $\log\text{it}\{P(x_{ni} \mid \theta_n, g)\} = \theta_n - (\beta_i + \delta_{ij} - \Delta G')$

DIF model: $\log\text{it}\{P(x_{ni} \mid \theta_n, g)\} = \theta_n - (\beta_i + \delta_{ij} - \Delta G' + \gamma_i G)$

Here, x_{ni} is the score of person *n* for item *i*; θ_n is a scale proficiency (ability) parameter for person *n*; β_i is the item difficulty parameter for item *i*; δ_{ij} is the step difficulty of level *j* on item *i*; γ_i is the DIF parameter for item *i*; G' is a group indicator variable contrasting the reference group, R, and the focal group, F, by taking the values of 1 and -1, respectively; g indicates either the reference group or focal group, and G = 1 if g = R (reference group), G = 0 if g = F (focal group). If a DIF model showed better fit at the .05 significance level, we further examined by items whether the absolute values of the DIF parameters were significantly larger than zero using the Bonferroni criterion (Moore & McCabe, 1998).

The test is carried out by calculating a test statistic (the absolute value of the ratio between the DIF parameter and its standard error), and comparing it to the relevant criterion value. If the test statistic exceeds the criterion, the item is considered to have statistically significant DIF. Then, effect size, two times of DIF parameter, was checked to determine whether the identified DIF has a practical impact (Longford, Holland, & Thayer, 1993). For any item that exhibited statistically significant DIF, we checked magnitude of effect size for a DIF item as: (a) negligible if its absolute value is smaller than .426; (b) intermediate if its absolute value is .426 or more and smaller than .638; and (c) large if its absolute value is .638 or more (Paek, 2002). Only DIF items with intermediate or large effect size were reported here.

5.2 Results

5.2.1 *Infant Toddler Instrument*

5.2.1.1 Gender DIF

The results of the likelihood tests revealed that the model with gender DIF parameters showed a significantly better fit for the COG indicator (that is, only the subset of COG measures in the LRN/COG set for infant toddler instrument). However, in the follow up item-wise inspection, no measure exhibited statistically significant DIF. Table 5.3 shows the DIF parameters and their Standard Errors (SE) of males obtained for the COG DIF analyses. The last column shows the test statistics needed to test the significance of the DIF parameter. If the absolute value of test statistic exceeds the Bonferroni criterion (Moore & McCabe, 1998), the measure has exhibited significant DIF. In this case, all the measures are below the criterion, so none shows DIF.

Table 5.3—DIF Parameters for Males in the DRDP-R Infant Toddler Instrument (*n = 506*)

Measure	DIF Parameter	SE	Test Statistic
COG			
Memory	0.08	0.08	0.95
Cause and Effect	0.10	0.08	1.24
Problem Solving	-0.23	0.09	-2.53
Symbolic Play	0.09	0.08	1.19
Curiosity	-0.11	0.08	-1.33

*Bonferroni Criterion for COG indicator is 2.59.

5.2.1.2 African-American vs. European-American DIF

The results of the likelihood tests revealed that the model with DIF parameters showed a significantly better fit for all indicators. Therefore, we can conclude that there is no DIF on ethnicity of African-American and European-American.

5.2.1.3 Hispanic-American vs. European-American DIF

The results of the likelihood tests revealed that the model with DIF parameters showed a significantly better fit for all indicators. Therefore, we can conclude that there is no DIF on ethnicity of Hispanic-American and European-American.

5.2.2 *Preschool Instrument*

5.2.2.1 Gender DIF

The results of the likelihood tests revealed that the model with DIF parameters showed a significantly better fit for the LRN/COG and MOT indicators. Only the *Socio-Dramatic Play* measure exhibited statistically significant gender DIF in the item-wise inspection. Since the effect size of *Socio-Dramatic Play* measure is .50 (= .25 * 2), it should be labeled as an intermediate DIF on gender. Because the DIF parameter of *Socio-Dramatic Play* measure is positive, it favors females with an intermediate DIF effect size. Table 5.4 shows the DIF parameters and SE of males obtained for the LRN/COG and MOT DIF analyses.

Table 5.4—DIF Parameters for Males in the DRDP-R Preschool Instrument (*n = 609*)

Measure	DIF Parameter	SE	Test Statistic
LRN/COG			
Curiosity and Initiative	-0.12	0.07	-1.77
Engagement and Persistence	0.02	0.07	0.35
Memory and Knowledge	0.00	0.06	-0.03
Cause and Effect	0.04	0.06	0.60
Engages in Problem Solving	-0.14	0.06	-2.17
Socio-Dramatic Play	0.25	0.07	3.73*
MOT			
Gross Motor Movement	-0.13	0.07	-1.74
Balance	-0.04	0.07	-0.50
Fine Motor Skills	0.16	0.07	2.28

*Bonferroni Criteria for LRN/COG and MOT indicators are 2.65 and 2.41, respectively.

5.2.2.2 African-American vs. European-American DIF

The results of the likelihood tests revealed that the model with DIF parameters showed a significantly better fit for the REG/SH and LANG/LIT indicators. However, item-wise inspection within each of the indicators indicated that no DIF item values were significantly larger than zero. Table 5.5 shows the DIF parameters and SE of the African-American group obtained for the REG/SH and LANG/LIT DIF analyses.

Table 5.5—DIF Parameters for African-American in the DRDP-R Preschool Instrument (*n = 144*)

Measure	DIF Parameter	SE	Test Statistic
REG/SH			
Impulse Control	0.20	0.14	1.42
Taking Turns	0.21	0.13	1.66
Shared Use of Space and Materials	0.07	0.14	0.46
Personal Care Routines	-0.37	0.16	-2.35
Personal Safety	-0.01	0.13	-0.11
Understanding Healthy Lifestyle	-0.31	0.14	-2.33
LANG/LIT			
Comprehends Meaning	0.34	0.15	2.29
Follows Increasingly Complex Instructions	-0.07	0.15	-0.50
Expresses Self Through Language	0.14	0.15	0.93
Uses Language in Conversation	0.08	0.14	0.59
Interest in Literacy	-0.25	0.14	-1.80
Concepts of Print	0.19	0.14	1.37
Letter and Word Knowledge	-0.05	0.14	-0.38
Phonological Awareness	-0.36	0.15	-2.43
Emerging Writing	0.01	0.14	0.09

*Bonferroni Criteria for REG/SH and LANG/LIT indicators are 2.68 and 2.82, respectively.

5.2.2.3 Hispanic-American vs. European-American DIF

The results of the likelihood tests revealed that the DIF model showed a significantly better fit for the SELF/SOC, LANG/LIT, LRN/COG and MATH indicators. No measure had significant DIF value in the item-wise inspection for SELF/SOC indicator. For the LANG/LIT indicator, *Comprehends Meaning*, *Uses Language in Conversation* and *Emerging Writing* measures exhibited statistically significant ethnic DIF in the item-wise inspection. The effect sizes of these three measures (0.57, 0.58 and 0.62, respectively) were intermediate. Because DIF parameters of *Comprehends Meaning* and *Uses Language in Conversation* measures are positive, these two measures favor European-American with an intermediate DIF effect size. Because the DIF parameter of *Emerging Writing* measure is negative, it favors Hispanic-

American with an intermediate DIF effect size. For the LRN/COG indicator, *Memory and Knowledge*, *Cause and Effect*, and *Engages in Problem Solving* measures exhibited statistically significant ethnic DIF in the item-wise inspection. The effect sizes of the first two measures (0.49 and 0.49, respectively) were intermediate, and the effect size of last one (0.42) was negligible. Because the DIF parameters of *Memory and Knowledge* and *Cause and Effect* measures are positive, these two measures favor European-American with an intermediate DIF effect size. For the MATH indicator, *Patterning* and *Time* measures exhibited statistically significant ethnic DIF in the item-wise inspection. The effect sizes of these two measures (.52 and .61, respectively) were intermediate. The *Patterning* measure has a negative DIF parameter which means that it favors Hispanic-American with an intermediate DIF effect size. The *Time* measure has a positive DIF parameter which means it favors European-American with an intermediate DIF effect size. Table 5.6 shows the DIF parameters and SE of Hispanic-American obtained for the SELF/SOC, LANG/LIT, LRN/COG and MATH DIF analyses.

Table 5.6—DIF Parameters for Hispanic-American in the DRDP-R Preschool Instrument (*n = 436*)

Measure	DIF Parameter	SE	Test Statistic
SELF/SOC			
Identity of Self	0.18	0.08	2.40
Recognition of Own Skills and Accomplishments	0.16	0.08	2.01
Expressions of Empathy	0.09	0.07	1.26
Building Cooperative Relationships With Adults	0.13	0.08	1.64
Building Cooperative Play With Other Children	-0.08	0.08	-0.98
Developing Friendships	-0.16	0.08	-1.99
Conflict Negotiation	-0.10	0.08	-1.26
Awareness of Diversity in Self and Others	-0.21	0.08	-2.75
LANG/LIT			
Comprehends Meaning	0.29	0.08	3.68*
Follows Increasingly Complex Instructions	-0.09	0.08	-1.17
Expresses Self Through Language	0.20	0.08	2.60
Uses Language In Conversation	0.29	0.07	3.92*
Interest in Literacy	-0.11	0.08	-1.43
Concepts of Print	0.11	0.08	1.35
Letter and Word Knowledge	-0.16	0.08	-2.04
Phonological Awareness	-0.21	0.08	-2.61
Emerging Writing	-0.31	0.07	-4.16*
LRN/COG			
Curiosity and Initiative	-0.04	0.08	-0.52
Engagement and Persistence	-0.09	0.08	-1.12
Memory And Knowledge	0.25	0.07	3.38*
Cause And Effect	0.25	0.07	3.36*
Engages in Problem Solving	-0.21	0.07	-2.85
Socio-Dramatic Play	-0.16	0.08	-2.08
MATH			
Number Sense: Understands Quantity and Counting	-0.16	0.08	-2.06
Number Sense: Math Operations	0.00	0.08	0.04
Shapes	0.05	0.08	0.62
Classification	-0.02	0.08	-0.27
Measurement	0.22	0.09	2.57
Patterning	-0.26	0.07	-3.51
Time	0.31	0.08	3.78*

*Bonferroni Criteria for SELF/SOC, LANG/LIT, LRN/COG and MATH indicators are 2.76, 2.80, 2.66 and 2.71, respectively.

5.2.3 *School Age Instrument*

5.2.3.1 Gender DIF

The results of the likelihood tests revealed that the model with DIF parameters showed a significantly better fit for the SELF/SOC and REG/SH indicators. Only the *Safety* and *Exercise and Fitness* measures exhibited statistically significant DIF in the item-wise inspection. The effect sizes for these two measures are 0.42 and 1.04, respectively. Hence, they should be labeled as negligible and large DIF, respectively. Because the DIF parameter of the *Exercise and Fitness* measure is negative, it represents that this measure favors males with a large DIF effect size, whereas *Safety* favors females with a negligible effect size. Table 5.7 shows the DIF parameters and SE of males obtained for the SELF/SOC and REG/SH DIF analyses.

Table 5.7—DIF Parameters for Males in DRDP-R School Age Instrument (*n = 701*)

Measure	DIF Parameter	SE	Test Statistic
SELF/SOC			
Identity of Self and Connection to Others	-0.15	0.06	-2.32
Self-Esteem	-0.15	0.06	-2.42
Empathy	0.15	0.06	2.40
Interactions with Adults	0.03	0.06	0.44
Friendship	0.05	0.06	0.77
Conflict Negotiation	-0.02	0.06	-0.30
Awareness of Diversity: Appreciation of Differences and Similarities	0.09	0.06	1.44
REG/SH			
Impulse Control	0.09	0.05	1.79
Follows Rules	-0.02	0.06	-0.38
Personal Care Routines	0.07	0.06	1.19
Safety	0.21	0.06	3.43*
Understanding Healthy Lifestyle	0.04	0.06	0.72
Exercise and Fitness	-0.52	0.06	-8.09*

*Bonferroni Criteria for SELF/SOC and REG/SH indicators are 2.70 and 2.65, respectively.

5.2.3.2 African-American vs. European-American DIF

The results of the likelihood tests revealed that the model without DIF parameters did not show a significantly better fit for the indicators. Therefore, we can conclude that there is no evidence for DIF on ethnicity of African-American and European-American.

5.2.3.3 Hispanic-American vs. European-American DIF

The results of the likelihood tests revealed that the model without DIF parameters did not show a significantly better fit for the indicators. Therefore, we can conclude that there is no evidence for DIF on ethnicity of Hispanic-American and European-American.

5.2.4 *Other Types of Validity Evidence*

In addition to evidence based on test content and internal structures, evidence based on response processes, relation to other variables, and consequences are important aspects of validity. These forms of evidence will be gathered at a later date.

5.3 Conclusion

As anticipated, the ability distributions generally moved forward for each age range; Figures (5.1 – 5.6) show the mode for each age range tended to increase. Furthermore, the DRDP-R domains are correlated. Correlations at these levels among such domains are not atypical. For example, the domains on the PISA (a test of 15 year olds academic achievement) are also correlated.

The main conclusion from the DIF analysis is that for the most part, DRDP-R measures do not show any great bias towards certain demographics. Even when statistically significant DIF was detected at the indicator level, actual DIF was found only in a handful of measures out of the whole instrument. Overall, there are gender DIF for the COG indicator in the infant toddler instrument, gender DIF for the LRN/COG and MOT indicators, African-American/European-American DIF for the REG/SH and LANG/LIT indicators, and Hispanic-

American/European-American DIF for the SELF/SOC, LANG/LIT, LRN/COG and MATH indicators in the preschool instrument, and gender DIF for the SELF/SOC and REG/SH indicators in the school age instrument. After item-wise inspection, only *Socio-Dramatic Play* of LRN/COG indicator for gender DIF and *Comprehends Meaning*, *Uses Language in Conversation*, and *Emerging Writing* of LANG/LIT indicator, *Memory and Knowledge* and *Cause and Effect* of LRN/COG indicator, and *Patterning* and *Time* of MATH indicator for Hispanic-American/European-American DIF in the preschool instrument, and *Exercise and fitness* of REG/SH indicator for gender DIF in the school age instrument were found to exhibit at least intermediate DIF.

6 Special Study I: Articulation of DRDP-R Across Age Groups

6.1 Introduction

This chapter describes the process of developing and validating the DRDP-R as a set of age-consecutive developmental assessments. We refer to this process as the articulation of the DRDP-R instruments.

The DRDP-R assessment system was designed with the goal of creating linked instruments to increase sensitivity to the multiple developmental paths children take through their first 12 years of life. Well-linked instruments can accommodate the wide range of typical developmental trajectories that young children typically exhibit. The main factor that affects the continuity between instruments is the linkage structure between measures that measure the same constructs across different age instruments.

We will illustrate how measures function as links between the three DRDP-Rs with a hypothetical example. Suppose we had one measure that assesses children between birth and 10 years of age, and also suppose the measure has 10 developmental levels, as in Figure 6.1 below.

\

0-10 years									
1	2	3	4	5	6	7	8	9	10

Figure 6.1—Levels of multi-year developmental measure.

With so many levels, it becomes cumbersome to display all ten of them for every child. Practitioners need only see the relevant levels for the age group under their care – and the levels they are most likely to rate. If we were to break this measure into three sections by age group, we would get a set of three instruments that assess the same construct, but differ in their target physical age, as in Figure 6.2.

IT: 0-3 years			PS: 3-5 years		SA: 5-10 years				
1	2	3	4	5	6	7	8	9	10

Figure 6.2—Levels of three non-overlapping age-specific measures.

In Figure 6.2, there is no overlap between the three age-specific instruments. This means that a child who is at the Preschool physical age, but developmentally is still in the last level in Infant/Toddler, could not be properly assessed using the PS instrument alone. There would be no level to describe this child's typical behavior.

One way to resolve this problem is to design instruments with some overlap between them, as in Figure 6.3. The expansion of levels in each instrument creates a link between them. Linkage in this paper means that consecutive instruments are connected to one another by having the same or similar developmental levels.

IT: 0-3 years	1	2	3	4						
PS: 3-5 years			3	4	5	6				
SA: 5-10 years					5	6	7	8	9	10

Figure 6.3—Levels of three overlapping age-specific measures.

In Figure 6.3, each age-specific instrument is expanded in the appropriate direction(s) to better assess all likely levels of development for an individual child of each age range, as shown in Table 6.1.

Table 6.1—Hypothetical Linking Structure of Levels of a Developmental Assessment Instrument (as in Figure 6.3)

Chronological age	Developmental age IT	PS	SA	Total levels in instrument
IT	1-3	4		4
PS	3	4-5	6	4
SA		5	6-10	6

Note.

a) An additional level (4) is added at the top of the IT measure, to accommodate toddlers who develop relatively quickly.
b) An additional level (3) is added to the beginning of the PS measure, to accommodate children who begin preschool based on physical age, but developmentally may be more like an older toddler.
c) Level (6) is added to the top of the PS measure, to accommodate children who develop more quickly in the preschool years.
d) Level (5) is added to the beginning of the SA measure, to accommodate children who begin Kindergarten based on physical age, but developmentally may be more like an older preschooler.

Linking instruments in this way can increase the sensitivity of assessments to relatively high-achieving ("early bloomers") and low-achieving children ("late bloomers"). Therefore, linkage between successive instruments increases their validity as a set, since the bridged measures allow the instrument to better assess child development on the intended construct across time**.** The process of creating a linkage between the DRDP-R measures is somewhat more complex than the simple example in Figures 6.1 to 6.3 above. We refer to the methodology of designing the linkage structure for the DRDP-R as "articulation."

6.1.1 *Articulation of the DRDP-R Instruments*

We define articulation as the extent to which individual measures on the DRDP-R instruments can be linked across age ranges. Linkage in this sense means similarity of behaviors described by measures from consecutive DRDP-Rs. We considered articulation with respect to several principles: developmental constructs (indicators and measures) should follow a coherent continuum from birth through 12 years.

- The developmental levels for bridged measures are linked appropriately at the intersection points between consecutive DRDP-R instruments.
- There are appropriate basal/ceiling cut-points for measures in each age range.
- The linkage structure is based on theories and knowledge of development.
- The expected linkage between measures should be supported by empirical evidence.

While developing the instruments, we found that there were discontinuities in the way theorists with expertise at different age levels conceived the developmental constructs, levels of development, and the observable behaviors that signify these levels. A concerted effort was made to bridge these differences in order to maximize continuity. By identifying common ground among age-level experts with regard to the representation of research-based developmental trajectories (and associated developmental constructs), and by acknowledging where discontinuities were justified by the research literature, we developed a continuum of indicators that reflects the achievement of Desired Results (DR) for the birth through 12 years developmental period.

Instrument developers also ensured the appropriate articulation so that practitioners would be able to track children's development within and across age groups, with smooth transition points where one instrument "ends" and another "begins." The process of building and articulating developmental constructs and levels relied on the most current evidence-based research and theories of child development, and included broad input from experts in developmental theory, experts in psychometric theory, and practitioners working with children at each of the age groups. The resulting profiles are grounded in developmental theory and research, highlighting discrete levels of progress along a continuum within key domains that can be observed and tracked by teachers and caregivers.

6.1.2 *Articulation as an Iterative Process*

Each of the teams of age-level experts was asked to use the list of DRs and Indicators in order to delineate key constructs that should be measured within each indicator. Thus, for example, each team was asked to consider the development of the *Self Concept* indicator (SELF) by delineating the major developmental constructs contributing to this indicator. Then for each of the constructs, age-specific teams created a developmental measure that describes the sequence of developmental levels that would typically occur within their age range.

After each age-specific team independently developed their conceptual framework of measures and developmental levels for the 10 indicators, conceptual "maps" were created to examine the articulation (or lack thereof) of all developmental constructs across the age groups (i.e., between IT and PS and between PS and SA). Developmental dimensions within the indicator for *Literacy,* for example, included measures such as *Visual Representation* and *Motivation and Interest* at the IT level, *Letter Knowledge*, *Word Knowledge*, *Phonemic Awareness*, *Story comprehension*, *Interest in Reading*, *Concepts of Print*, and *Emerging Writing* at the PS level, and *Word/Letter Structure and Attack, Comprehension*, *Comprehension Strategies*, *Connections of Stories to Personal Life*, *Writing Application* and *Writing Reviews/Edits* for SA children and youth. Using these dimensions as a starting point, the articulation team looked for logical connections between age groups, such as the developmental progression from *Visual Representation* in infancy to *Letter Knowledge* in preschool, to *Word/Letter Structure and Attack in School-Age.*

However, given that the primary developmental indicator may change for each age group, there were also discontinuities between particular constructs that were either more important for one age group than another or, in some cases, even nonexistent before or after the transition points between one age group and the next. Such discontinuities were evaluated from the perspectives of both (and in some cases, all three) age groups to determine (1) if a theoretically sound (i.e., developmentally appropriate and research-based) case could be made for the discontinuity of a construct from one age group to another, or (2) if one or more of the age groups needed to reconsider how to create a "bridge" between the age groups that could be defended on the basis of existing research and theories of child development. For the *Literacy* constructs listed above, general constructs of *Word Level*, *Comprehension*, *Interest* and *Writing* were used to organize all of the developmental constructs involved in the development of literacy. Where any discontinuities existed across the age groups (such as the absence of *Writing* in the IT conceptual map), input was sought from the developmental experts from one or more age groups to determine if the omission was indeed justified.

Thus, the initial work of the articulation team was to propose conceptual "bridges" between age groups where that had been deemed appropriate, and to identify points where such conceptual bridges may not have been obvious, leading to deliberations about the appropriateness of conceptual discontinuity across age groups. For example, physical development is much more differentiated in infancy as babies and toddlers gain increasing control over fine and gross motor skills, yet by school age, motor development consists mostly of increasing refinement of coordination and dexterity rather than continued differentiation of discrete motor skills. Therefore, the DRDP-R measures multiple aspects of physical development in the birth-to-three age range (e.g., *Gross Motor, Balance , Fine Motor,* and *Eye-Hand Coordination*), which articulate with increasingly fewer constructs for physical development by the time one is measuring school-age children (e.g., *Fine Motor* and *Gross Motor*). Likewise, the DRDP-R measures relatively fewer constructs related to early literacy for the birth-to-three age range (namely, *Interest in Literacy* and *Recognition of Symbols*) as compared to the numerous literacy-related measures in the DRDP-Rs for preschool- and school-age children, in which the expansion and development of multiple types of literacy skills are more prominent.

In addition, articulation of the three age-related versions of the DRDP-R was examined from the perspective of developmental levels linked within measures. Thus, the "highest" levels of development included on the IT instrument for a given measure had to link with the "lowest" level of development of the linked measure on the PS instrument. Similarly, the "highest" PS levels had to link with the "lowest" SA levels. Using this linkage structure, the research team ensured that the developmental progression for a particular construct was consistent across age groups.

6.1.3 *Articulation Challenges*

The main factor to consider in articulation is whether there should be a link between two measures based on the common nature of their underlying construct. The process of articulation begins by asking the following question: *Are these measures conceptually similar and is a conceptual link supported by the developmental literature?* Occasionally, there was not a simple answer to this question, because measures may have been revised or discarded in the process of instrument development. In some cases, the developmental literature suggests a different emphasis or direction for different age groups, thus measures that assess the same construct have different orientation in each DRDP-R. Specifically, a number of challenges surfaced during the process of examining articulation of developmental levels across constructs in adjacent instruments. We discuss here the key challenges we faced and how each was addressed.

6.1.4 *Variegated Nature of Development*

How do we represent development as a continuum from birth through 12 years? Given that the structure of the DRDP-R relies on the assumption that typical patterns of development occur in an ordered fashion, as represented by sequences of levels for each measure, it was a challenge to accommodate aspects of development that do not necessarily follow such a pattern. For example, while Piagetian theory defines developmental stages, as in the transition from concrete to abstract thinking, school-age children who are making this transition may exhibit aspects of both concrete and abstract thinking, and this transition period may persist over many months (Ausubel & Ausubel, 1966).

The DRDP-R accommodates the variegated nature of development by focusing on those major milestones that do occur sequentially. In this way, the emphasis of the DRDP-R is on those aspects of development that are known to occur sequentially – such as the sequence of physical development that the majority of infants and toddlers demonstrate, from rolling over to crawling, followed by walking and then running. Thus, by focusing on major milestones of development, and defining "mastery" of a developmental level as that behavior which the child demonstrates repeatedly and on a regular basis, the DRDP-R accommodates children's typical and sometimes non-linear developmental pathways. In addition, since children are rated on discrete developmental constructs at the level that most closely matches their current level of functioning, the instrument reflects children's higher or lower levels of development across a wide range of measures. Thus, the developmental profile of a preschooler who may have more advanced gross motor skills but less advanced fine motor skills will reflect these differential levels of mastery appropriately.

6.1.5 *Discontinuities Among Theoretical Conceptualizations of Development Within and Across Age Ranges*

Because theoretical concepts of development vary both within and across age groups, there was constant tension between maintaining fidelity to developmental theories and accommodating variations among them. In most cases, an attempt was made to first accommodate the overriding theory or theories chosen by developmental experts as being the best representation of a specific stage on a construct, for a given age group. When developmental experts across age groups differed with regard to the conceptual framework to use, a series of "bridging" discussions ensued. The goal of these discussions was to reach consensus on the most appropriate theoretical representation of developmental levels for each of the three adjoining DRDP-R instruments, while allowing for reasonable overlap between instruments. The intent was to ensure that the DRDP-R for each age group accurately captured the developmental progress typical of the majority of children within each age range, while also allowing for enough of a link (or overlap) between the highest level(s) of the "younger" DRDP-R with the lowest levels of the next "older" DRDP-R instrument. This articulation process produced a linkage structure between the DRDP-R instruments, based on their content analysis. The following section describes this structure in more detail.

6.1.6 *Content Linkage Structure of DRDP-R Measures*

The development team identified where levels link on measures across the DRDP-R instruments. The established linkage serves as the a priori basis of empirical investigation, because it determines which measures will be scaled together in a cross-age psychometric analysis. Overall there are 105 measures in the three instruments (35 measures in IT, 39 in PS and 31 in SA). Appendix C contains a list of measure names and descriptions within all DRDP-R indicators. In the following discussion we will refer to instrument-specific measures as "measures" (as we have before), and the set of measures that link together as "linked measures." The reader is encouraged to use the description of measures in Appendix C to understand what exactly is being assessed by each measure.

Not all the measures span across all instruments, meaning there are structurally missing links between the DRDP-Rs. When the instruments were constructed, decisions were made about the appropriateness and relevance of each measure in each age group. In some cases, an IT measure was discontinued in PS because it is not as important developmentally in preschool years as it is in toddlerhood. Similarly, a measure may have first appeared in the PS DRDP-R because that is the age level at which the skills typically emerge. For example, in Figure 6.4, the PS LIT measure *Emerging Writing* does not link to any IT measure because precursor writing behaviors in infancy are complicated to assess. We identified 33 linked measures assessed by 88 instrument specific measures (84% of all the DRDP-R measures). Two-thirds of the linked measures connect all three instruments and one-third link only between two consecutive instruments. Most linked measures had one or two of their levels linked, and the typical linkage pattern had two levels linked on both instruments. Overall, there were 54 individually linked levels between IT and PS, and 55 linked levels between PS and SA. In the remainder of this section we explain how we present our content linkage structure and describe the types of linkage we identified in between measures. At the end of the section, we describe the features of the unlinked measures (16% of all the DRDP-R measures).

	Expressions of Empathy **PS**	**Empathy** **SA**
		Considering the Needs of My Community
		Shows understanding of feelings and experiences through words or actions for people who live in his or her community (may not know them)
		Considering Other Perspectives
		Shows how someone else might feel in a certain (hypothetical) situation
	Integrating	**Understanding Someone Else**
	Uses words or actions to demonstrate concern for what others are feeling	Shows awareness of feelings of others with appropriate words or actions
Empathy	**Building**	**Focusing on Me**

IT		
	Accurately labels own feelings, as well as those of others	Demonstrates awareness of own feelings
Developing Ideas	**Developing**	
Offers comfort to someone showing distress	Offers simple assistance when he or she thinks it is needed- even if not really needed	
Discovering Ideas	**Exploring**	
Shows concern for others' feelings	Shows awareness when others are unhappy or upset	
Acting with Purpose		
Changes behavior based on others' expressions of emotions		
Expanding Responses		
Shows awareness of others		
Responding with Reflexes		
Responds to others with reflexes		

Measure definitions:
IT: Empathy – Child shows awareness of other's feelings and responds to expressions of feelings by others
PS: Expressions of empathy – Child shows awareness of other's feelings and responds to expressions of feelings by others in ways that are appropriate to the other person's needs
SA: Empathy – Child shows awareness of others' feelings and experiences and responds appropriately through words or actions

Figure 6.4—Linkage structure of Empathy measures (SOC) – example of strong linkage.

Table 6.2 shows the content linkage structure of the DRDP-R, based on the age-specific teams' analyses. For brevity, level names and descriptors are not shown (these can be found in Appendix C). Within each indicator, we identified constructs that appear across more than one instrument. Each row in Table 6.2 contains a link between two or more measures that assess the same construct. An empty cell indicates no link on that construct. Values from 1-6 represent which levels link between the measures on each pair of instruments. For example, the first construct, *Identity of Self* (SELF indicator), has "4/5" under IT and "1" under PS, meaning the last two levels in the IT measure (levels 4 and 5) link to the first level in the PS measure (level 1). The link between PS and SA is represented as “2/3,4” and “1,2”, respectively. This means that levels 2 and 3 in the PS measure are linked to level 1 in the SA measure and level 4 in PS is linked to level 2 in SA.

Table 6.2—Linkage Structure of the DRDP-R Based on Content Analysis for the 2005 Calibration Study

Indicator	Linked Measures	Content Linkage IT→	PS	PS→	SA
Self-concept (SELF)	Identity of Self	4/5	1	2/3,4	1,2
	Self Esteem	4,5	1,2	4	1
Social Interpersonal Skills (SOC)	Empathy	4,5	1,2	3,4	1,2
	Relationships w/ Familiar Adults	4,5	1,2	-	-
	Interactions with Peers	3,4,5	1,2,3	-	-
	Friendship	3,4/5	1,2	4	2
	Conflict Negotiation	-	-	3,4	1,2
	Awareness of Diversity	5	1/2	2,3,4	1,2,3
Self Regulation (REG)	Impulse Control	-	-	2,3/4	1,2
Safety and Health (SH)	Personal Care Routines	4,5	1,2	1,2,3, 4	1,2,3, 4
	Safety	4,5	1,2	1,2	1/2,3
Language (LANG)	Comprehends Meaning	4/5,6	1,2	1/2,3, 4	1,2,3
	Responsiveness to Language	4/5/6	1	1,2,3/ 4	1,2,3
	Expression of Oral Language	5/6	1	3,4	1,2
	Uses Language in Conversation	5,6	2,3	-	-
Literacy (LIT)	Interest in Literacy	4,5	1,2	3,4	1,2
	Letter and Word Knowledge	5	1	4	1/2
	Writing	-	-	3,4	1,2
Learning (LRN)	Curiosity and Initiative	3,4/5	1,2	3/4	1
	Engagement and Persistence	3,4,5	1,2,3	2/3/4	1/2
Cognitive Competence (COG)	Memory	4,5	1,2	3,4	1,2
	Cause and Effect	3,4,5	1,2,3	3/4	1/2/3
	Problem Solving	4/5	1	4	1/2
	Socio Dramatic Play	-	-	4	1
Math (MATH)	Number Sense	5	1	4	1
	Math operations	-	-	4	2
	Shapes	5	1	4	1
	Classification	3/4,5	1,2	-	-
	Measurement	-	-	2,3,4	1,2,3
	Time	5	1	4	1
Motor Skills (MOT)	Gross Motor	6	1	3,4	1,2
	Balance	6	1	-	-
	Fine Motor	6	1/2	3,4	1,2

Note.

1. Values represent linked levels between pairs of instruments.
2. Commas separate linked levels. For example 4,5 under IT and 1,2 under PS means that level 4 in PS links with level 1 in IT, and level 5 in PS links with level 2 in IT .
3. A slash between levels means they share a link to the other instrument. For example, 4/5 links to 1.

Table 6.3 summarizes the information in Table 6.2 by showing the number of constructs that link across all instruments and those that link only between two instruments for each indicator.

Table 6.3—Summary of the Theoretical Linkage Structure for the DRDP-R Calibration Study 2005

Indicator	Constructs	IT→PS→SA Links	IT→PS Links	PS→SA Links
SELF	2	2	-	-
SOC	6	3	2	1
REG	1	-	-	1
SH	2	2	-	-
LANG	4	3	1	-
LIT	3	2	-	1
LRN	2	2	-	-
COG	4	3	-	1
MATH	6	3	1	2
MOT	3	2	1	-
DRDP-R Total	*33*	*22*	*5*	*6*

Most measures linked to other measures within the same indicator across different age-groups. Three IT measures were identified as having conceptual links to measures outside their indicator. Specifically, *Curiosity* (COG indicator) and *Attention Maintenance* (IT-REG indicator) both link to the *Learning* constructs in Table 6.2: *Curiosity and Initiative*, and *Engagement and Persistence*, respectively. The *Learning* indicator is absent from the IT instruments, but its constructs are assessed by IT measures, nevertheless. Additionally, the IT SELF measure *Awareness of Diversity* links to the *Awareness of Diversity* measures of the

SOC indicator. In IT, awareness is perceived mainly with respect to self, while in preschool and school age it is mainly perceived with respect to others.

Not all links are the same. In the first typical situation, and the most common, measures assess the same aspects of development, meaning they are conceptually similar, and use similar language to describe behaviors. For example, the three measures for the SOC construct, *Empathy,* are aligned based on content links, as shown in Figure 6.4. Levels 4 and 5 in IT link to levels 1 and 2 in PS, and levels 3 and 4 in PS link to levels 1 and 2 in SA (we will represent this as IT4,5→PS1,2; PS3,4→SA1,2). The descriptors describe very similar behaviors, thus creating a strong linkage between the measures. The development of these measures and the linkage between them is based on research about children's development of empathy, including, but not limited to, Bergman and Wilson (1984), Hoffman (2000) and Eisenberg, Spinrad and Sadovsky (2006).

In the second typical situation, measures link conceptually, but the description of behaviors in the measure may not show the link in a straightforward way, meaning there is a partial linkage between the measures. For example, the skills assessed by several IT measures may develop into one general ability that is assessed by the PS measures. In the LANG indicator, for instance, *Responsiveness to Language* in IT links to *Follows Increasingly Complex Instructions* in PS. Both are assessing responsiveness to language, but the emphasis in PS is on following instructions because it is very accessible for preschool teachers to assess in their program's setting. In such cases, the DRDP-R examples may shed light on the nature of the link between the measures. There are many kinds of partial links, such as when levels are linked out of sequence (e.g., a link is missing on one level, although adjacent levels do link), or when the link bridges across levels (i.e., it is not a one-to-one relation). For example, consider a hypothetical linked measure on two successive instruments A and B. Partial links can occur when the first level in B links to the second-to-last level in A, rather than to the last level. Other examples occur when more than one level in instrument A links to only one level in instrument B, or when one level in A links to many levels in B. Examples of partial linkages found in the DRDP-R are provided in Figures 6.5 and 6.6.

Reciprocal Communication IT	Uses Language in Conversation PS
	Integrating
	Has extended conversations that include discussions of emotions, ideas, and information obtained from the other person
Connecting Ideas	**Building**
Engages in simple conversations with caregiver that involve several ideas	Has long back and forth conversations about real or imaginary experiences
Developing Ideas	**Developing**
Introduces one or two simple ideas in back-and-forth communication with caregiver	Carries out back and forth conversations
Discovering Ideas	**Exploring**
Engages in back- and-forth communication with caregiver using familiar single words	Uses language for three or more purposes, such as requesting, refusing, describing, and answering questions
Acting with Purpose	
Engages in back- and-forth communication with caregiver using vocalizations, gestures, or facial expressions	
Expanding Responses	
Responds to caregiver's voice or facial expressions during interaction	
Responding with Reflexes	
Responds to sounds with reflexes	

Measure definitions:
IT: Reciprocal Communication — Child engages in back-and-forth communication or conversation
PS: Uses language in conversation – Child engages in back-and-forth communication or conversations following the appropriate social use of language

Figure 6.5—Linkage structure of Uses Language in Conversation measures (LANG) – example of partial linkage.

Decoding

SA

Using Skills Effectively
Figures out new, multisyllable words by using strategies
Building on Skills
Sounds out written multisyllable words that he or she knows
Beginning Literacy Skills
Identifies simple/familiar words and sentences
Pre-Literacy Skills
Recognizes and names at least half of the printed alphabet letters.

Letter and Word Knowledge

PS

Integrating
Knows most of the letters by sight and by name, and recognizes some familiar whole written words
Building
Knows 10 or more letters by sight and by name, and understands that letters make up words and have corresponding sounds
Developing
Knows some letters by sight and by name, or recognizes own name in print
Exploring
Recognizes simple symbols (numbers, letters, logos) in the environment

Recognition of Symbols

IT

Developing Ideas
Shows understanding that a series of pictures represents a story, and recognizes simple symbols
Discovering Ideas
Shows understanding that pictures represent people and things
Acting with Purpose
Attends to things that caregiver points to, shows, or talks about
Expanding Responses
Responds to movements, patterns, gestures, and facial expressions
Responding with Reflexes
Responds with reflexes to people and things

Measures definitions:
IT: Recognition of Symbols – Child shows awareness that symbols and pictures represent people, objects, and actions
PS: Letter and word knowledge – Child shows awareness of symbols, letters, and words in the environment, and their relationship to sounds
SA: Decoding (Word Recognition) – Child shows increasing recognition and understanding of letters and words

Figure 6.6—Linkage structure of Letter & Word measures (LIT) – example of partial linkage.

Figure 6.5 provides the linkage structure of the IT and PS LANG measures of *Uses Language in Conversation*. The development of these measures is based on the literature about social aspects of language development, including, but not limited to, Girolametto and Weitzman (2002) and Tamis-LeMonda, Bornstein and Baumwell (2001). The measures concentrate on the development of language as a tool for communicative purposes, specifically as part of a social interaction with peers and adults. Due to the large variability in the development of language skills, the first level in PS is relatively low; it describes behaviors that can be expected in toddlerhood. Figure 6.5 is an example of a partial link because the "Exploring" level in PS is linked to both the "Acting with Purpose" and "Discovering Ideas" levels in IT, which describe emerging uses of language for basic purposes. The two highest levels in IT link with the two intermediate levels in PS ("Building" and "Developing"). Although the descriptors of these use different words, the intended meaning is similar. For instance, the examples in the "Connecting Ideas" level in IT describe conversations involving pretend play, similar to the examples given in the "Building" level in PS.

Figure 6.6 shows the linkage structure between the age-group specific LIT measures of *Letter and Word Knowledge*. The ability to recognize letters is a basic step in the process of learning to read and write. The development of these measures is based on the literature about social aspects of language development, including, but not limited to, Durrell (1980), Ehri (1995), Treiman and Broderick (1998), DeLoache (1996), Whitehurst and Lonigan (1998) and International Reading Association and NAEYC (1998). The IT and PS measures have a strong linkage between the last IT level and the first PS level as they both describe simple symbol recognition. The last PS level has a partial linkage with the first two SA levels. These levels describe both the ability to recognize most letters of the alphabet and a few familiar words.

Out of the 105 DRDP-R measures, 17 measures did not form a link across age consecutive instruments. Some measures did not form a link because they assess age-specific aspects of development that are less relevant or less important in other age groups. For example, the IT REG measures, *Self Comforting*, *Seek Other's Help to Regulate Self* and *Responsiveness to Other's Support* are observable behaviors that are important to assess in IT programs settings, but less important in older children. These measures existed only in the IT instrument. The other age-specific measures that were unlinked are the IT SELF measure, *Self Expression*, and MOT measure, *Eye-Hand Coordination*, the PS REG measures, *Taking Turns* and *Shared Use of Space and Materials*, the LIT measures, *Concepts of Print*, *Phonological Awareness*, and the MATH measure *Patterning*, the SA REG measure, *Follows Rules* and the SH measure, *Exercise and fitness*. Other measures existed in more than one instrument, but emphasize age-specific aspects of development that could not be linked across instruments. For example, the IT COG measure, *Symbolic Play*, does not form a link with the PS COG measure, *Socio-Dramatic Play*, because the emphasis in PS is on play that involves social interactions and language skills. Other unlinked measures in this category are the IT REG measure, *Impulse Control*, the PS SH measure, *Understanding Healthy Lifestyle*, the SA SOC measure, *Interactions with Adults* and SH measure *Understanding Healthy Lifestyle*.

Based on the content linkage structure, we expect that if a child were rated on a linked level in the child's age-group instrument, then the teacher would likely rate him or her on the matching linked level in the adjacent instrument. These rating tendencies should be reflected in the structure of psychometric difficulties of the developmental levels. For measures that conceptually link and where practitioners were able to observe the described behaviors, the probability of rating at linked levels, for a given child, should be similar across measures. If the measures from both instruments are scaled together, then the alignment of level difficulties would show evidence of the content linkage structure between measures. Therefore, the psychometric analysis can help validate the linkage structure established from the content analysis.

6.1.7 *The DRDP-R Empirical Linkage Analysis*

In educational measurement, the term “linking” generally refers to various approaches for making comparisons between results from different assessments. The literature distinguishes between different types of linking processes such as *equating*, *calibration*, *statistical moderation* and *projection* (Linn, 1993; Mislevy, 1992; National Research Council, 1999). The general goal of these linking approaches is to find a set of transformations between students’ scores on two related tests. Thus, students’ scores on different tests would be represented with the same scale so the scores can be compared. These transformations are very useful for making measurement-related decisions. For example, linking can be used to convert a score on a test the student has taken to a score on a test he or she has not taken. Similarly, when a new form of a standardized test is created, the scores on the new form are scaled so they can be compared to those obtained from previous forms of the test.

Strong forms of linking, such as equating, require that the linked tests measure the same construct with about the same difficulty and reliability. Linking scores on tests that measure the same construct but differ in difficulty or reliability is referred to as *calibration*. In this sense, the DRDP-R setting allows for a calibration approach, as the instruments target increasingly higher levels of development on the same 10 indicators. One form of calibration is *vertical scaling* where scores from tests that target increasing difficulties (e.g., different grade level), are calibrated to have the same scale (Kolen, 2004). Vertical scaling poses considerable challenge for psychometrics as the content of the tests tends to change across grades, which is a challenge to the assumption of a common construct. The conceptual and content continuity of the DRDP-R instruments with respect to the constructs assessed by measures and clusters of measures (indicators), was addressed by the articulation team work and is the topic of the current study.

A common study design that supports linking of two tests is to have a group of children rated on both tests so that scores can be scaled together. For example, if a group of preschoolers is rated on both IT and PS DRDP-R instruments, their ratings could be calibrated together in the same analysis. Different linking methods can be used to relate the scores on the two instruments, so that scores for children in infant-toddler programs are on the same developmental scale as the scores for preschoolers. Moreover, the measure difficulties of the younger-age test are ordered on the same scale as the older-age difficulties, thus allowing researchers to see if theoretical links between measures are supported by the data. For example, since each DRDP-R targets more complex and more difficult behaviors than the younger-age DRDP-R, we would expect to find the level difficulties of the SA instrument to be higher than the PS level difficulties, which will be higher than IT. The 2005 Calibration Study included just such a “double-instrument” condition for a sample of children at the transition ages between groups (Condition 3 sample). We elaborate on the calibration study and its analysis in the method section.

One way to achieve linking between two tests is to have a common set of measures appear on both tests, allowing the researcher to use the difficulty estimates of these measures to calibrate student scores (i.e., put test scores on a common scale). To keep the DRDP-R manageable for practitioners, whole measures generally are not repeated in age-consecutive instruments. However, parts of measures are repeated, since the DRDP-Rs were designed with some overlap of levels at the intersection points between instruments, as described in previous sections. Special Study I considers only those linked measures. However, scores in the DRDP-R framework are based on the full set of measures, which include the un-linked measures as well. Moreover, the current function of the DRDP-R in California does not require comparison of child development across different age groups, only within each age-group. Therefore, linking analysis in its traditional form (i.e., linking scores) is not relevant for this study.

The validity of inferences about scores depends, among other factors, on the extent to which measures assess the desired underlying construct. When assessments are vertically scaled, a content framework is needed to ensure that the same construct is measured by both assessments. The DRDP-R articulation process constitutes such a framework. Moreover, it is crucial to study the relationships between measures that constitute linkages between the three DRDP-Rs, before any cross-age score linking or inferences can be made. Our empirical linkage analysis aims to find alignment between difficulties of levels that measure the same construct-specific developmental stage on age-consecutive instruments. We need to find out how these alignments relate to our articulation process. Specifically, our analysis aims to replicate the linkage structure represented in Table 6.2 using the empirical alignment of psychometric level difficulties.

6.2 Method

The Condition 3 sample of the DRDP-R Calibration Study 2005 was collected to investigate the nature of linkages between the instruments. Children whose chronological age was close to the limits of their age group were rated in a “Double-Instrument” condition. The younger sample included ratings for 144 children, all of whom were approximately 3 years old (from the IT and PS samples). The older sample included ratings for 119 children, all of whom were approximately 5 years old (from the PS and SA samples). Each child was assessed by the same rater, using the relevant two instruments (e.g., preschool teachers used the IT and PS instruments to rate the three-year-olds, and the PS and SA instruments to rate the 5-year-olds).

For the psychometric analysis, the data were augmented with 488 children from the PS sample, to achieve better estimates of model parameters. Overall, the analysis included 751 children between the ages of 3 and 5. The structure of the data set used in this analysis is presented in Table 6.4.

Table 6.4—Number of Rated DRDP-Rs Used in the Analysis

	Rated on instrument		
Age range	IT DRDP-R	PS DRDP-R	SA DRDP-R
Approx. 3 years old	144	144	
3 to 5 years old		488	
Approx. 5 years old		119	119
Total	144	751	119

The double instrument data set allows us to analyze ratings for an individual child on pairs of DRDP-Rs together. The result is that we can compare measures of the same construct from two age-consecutive instruments. To achieve this, we fitted a Random Coefficients Multinomial Logit Model (RCML; Adams, Wilson, & Wang, 1997), which is an extension of the Rasch family of measure response models. Rasch type models enable estimation of children's abilities and item difficulties on a common scale. Within this formulation, a Partial Credit model (Masters, 1982) was employed, so the difficulty of attaining each developmental level can be estimated separately for each measure (as opposed to only one general measure of difficulty). The model was fitted using *ConQuest* (Wu, Adams & Wilson, 1998). Technically, this analysis was similar to the calibration analysis, only the data set was different. The linking analysis included only children who were rated on the PS instrument, and any other instrument that was used to rate them (i.e., IT or SA ratings from Condition 3).

6.3 Results

The first results section describes the rating tendencies of teachers in the Double Instrument condition. The second section describes the information obtained from unidimensional analyses within each indicator. Finally, the third section describes how estimates of level difficulties were used to investigate the linkage structure obtained from the DRDP-R content analysis described in the introduction.

6.3.1 *Double-Instrument Ratings*

In this section we briefly discuss how raters agree with themselves in assessing the same child using two different instruments. Every pair of linked measures produced a contingency table across all pairs of raters and children. For example, Figures 6.7 and 6.8 show the percentages of children rated at any combination of levels between the IT and PS instruments (zero frequency cells are empty). The shaded cells of Figure 6.7 show that children were rated as expected in the linked levels of the *Self Concept* measure *Recognition of Ability* (SELF2, Expected Linkage from Table 6.2: 4,5→1,2). The shaded cells of Figure 6.8 show that children were rated somewhat more leniently than expected on the *Math* measure *Shapes* (MATH2, Expected Linkage from Table 6.2: 5→1). Specifically, most children who were rated at level 5 on the IT instrument were rated at levels 1 **or** 2 on the PS instrument.

IT level	PS Level Not Yet	1	2	3	4	IT Total
1						
2				*.01*		*.01*
3	*.01*	*.05*	*.03*			*.09*
4	*.01*	*.11*	*.17*	*.01*		*.30*
5	*.01*	*.13*	*.40*	*.03*	*.01*	*.60*
PS Total	*.04*	*.30*	*.60*	*.04*	*.01*	*1.00*

*N= 141

Figure 6.7—Contingency table of teachers' rating on Self Esteem construct (SELF).

IT Levels	PS Levels Not Yet	1	2	3	4	IT Total
1		*.01*				*.01*
2	*.01*	*.01*				*.02*
3	*.04*	*.05*	*.01*			*.10*
4	*.04*	*.17*	*.10*			*.31*
5	*.03*	*.22*	*.22*	*.06*	*.01*	*.55*
PS Total	*.13*	*.46*	*.33*	*.06*	*.01*	*1.00*

*N=138

Figure 6.8—Contingency table of teachers' rating on Shapes construct (MATH).

The PS and SA instruments have an additional "Not yet" level before the first developmental level. This level is used to rate children who are not yet exhibiting behaviors described in the first level of these instruments. The "Not yet" level ensured that teachers had a level to fit each child. For example, suppose two age-consecutive instruments are linked so the last level on the younger measure is linked to the first level on the older measure. We expect children who were rated below the last level on the younger instrument to be mostly rated at the “Not yet” level in the older instrument, simply because the first level on the older instruments is not supposed to describe the behaviors they exhibit. Such a pattern can be seen in Figure 6.8 as non-zero entries on the “Not yet” column in PS. This may be a result of children who were rated below the linked levels in IT (levels 1-3). Because none of the levels in PS described these behaviors, they had to be rated as “Not yet” in PS. Many DRDP-R measures exhibited such patterns. Notice that in Figure 6.8, some of the children who were rated low on IT were rated at level 1 on PS and some of those who were rated “Not yet” on PS were rated as level 4 or 5 on IT. These inconsistencies are expected because the ratings may be affected by many sources of error (rater bias, training, clarity of instrument, time to complete DRDP-R, etc.)

A similar issue occurred at the other end of the scale. A child might show behaviors more complex than the highest developmental level on his or her age-group instrument (“early bloomer”), but the rater will have to rate that child on that highest level, for lack of a higher option. If that child is also assessed with the next age-group instrument, his or her behaviors may be well described by one of the higher levels in the older age instrument, and not necessarily those that link to the last level in the younger age instrument. Such patterns can be seen in Figure 6.7 and 6.8 as non-zero entries on the level 5 row in IT. For example, in Figure 6.7, we expected children who were rated at level 5 in IT to be rated at the linked level in PS (level 2). However, the figure shows that teachers rated some of these children higher than level 2. This may be because these children exhibited more complex behaviors but were rated at level 5 in IT for lack of a higher level. These children may have been rated higher than the linked levels on PS because those levels better described their behaviors[14]. Again, the contingency tables of most DRDP-R measures resemble such lower-triangular patterns.

6.3.2 *Psychometric Analysis*

The double instrument data set allows us to analyze ratings for an individual child on pairs of DRDPs together. The result is that we can compare items of the same construct from two age-consecutive instruments. To achieve this, we fitted a Random Coefficients Multinomial Logit Model (RCML; Adams, Wilson & Wang, 1997), which is an extension of the Rasch family of item response models. Within this formulation, a Partial Credit model (Masters, 1982) was employed, so the difficulty of attaining each developmental level can be estimated separately for each item (as opposed to only one general item difficulty). The model was fitted using *ConQuest* (Wu, Adams & Wilson, 1998).

The analyses were based on combining items that measure the same constructs in different instruments. Therefore, for each indicator we calibrated only the bridged items from all instruments. For example, there was one analysis of the SELF indicator. We fitted the RCML partial credit model for the bridged SELF items in I/T, PS and SA, using augmented data described in the method section. This analysis was repeated for each indicator except for REG, which had only one set of linked items and therefore did not conform to psychometric requirements. Overall, there were 9 unidimensional analyses; each produced ability estimates for the children and step difficulty estimates for the levels of the items in the sample. The Double Instrument sample served as anchor points for item estimates in each analysis. Because children in this sub-sample were rated on pairs of instruments, the analysis inevitably located abilities and difficulty from all instruments on the same scale.

---

[14] Teachers were instructed not to rate children as "emerging" if they were rated at the highest level because the instruments did not include the description of the next level. Therefore, the "emerging" data were not used in the analyses described in this paper.

Due to the probabilistic nature of the model, it is expected that a child will occasionally be rated higher or lower than his or her true ability. On some items, this rating variance may be problematic. Therefore, we calculated item-fit statistics to measure the extent to which the psychometric model captures the observed variability in the data. The weighted mean square error (WMSE) compares the variability in teachers' ratings to the variability expected by the model, given the distribution of ability scores. Values around 1 indicate that the amount of variability is exactly as expected. Values larger than 1 indicate that there is more variability than expected, meaning that over- or under-rating of children's ability is too common. Values lower than 1 indicates less variability than expected, meaning that teachers rate children in a manner that is too consistent with the model's prediction. Generally, values between 0.75 and 1.33 indicate satisfactory item fit. Table 6.5 shows average and SD of fit statistics for items within each indicator. Overall, the DRDP items showed fit well within the desired range (only six items had WMSE slightly outside the range).

Table 6.5—Distribution of Weighted Mean Square Error (WMSE) for DRDP Items and SD

Indicator	Average WMSE	SD
SELF	0.91	0.11
SOC	1.03	0.12
LANG	0.88	0.15
LRN	1.07	0.09
COG	0.98	0.11
MATH	0.95	0.14
LIT	0.96	0.11
MOT	0.87	0.17
SH	0.94	0.13
Total	0.96	0.14

6.3.3 *Alignment of Developmental Level Difficulties*

The indicator-level psychometric analysis attempts to locate children's abilities and measure-specific developmental levels on a unidimensional continuum across age-groups. Based on response patterns, the analysis determines children's ability relative to the location of level difficulties on all measures of each indicator. For a specific set of linked measures, if level 5 in IT is approximately as difficult to reach as level 1 in PS, then the two instruments are aligned at those levels. The difficulty of reaching each level is best represented by the Thurstonian threshold. For a measure with *K* developmental levels, there are *K-1* thresholds (one between each successive pairs of levels). Threshold *k* is defined as the point on the ability continuum where there is a probability of 0.5 for achieving level *k* or more on the measure (Wu, Adams, & Wilson, 1998). Within each measure, the thresholds are ordered with respect to their logit values, as are the children's abilities. In Figure 6.9, the child's ability is represented by the triangle aligned with the k^{th} threshold. The probability that a teacher would rate that child at any of the levels is represented by the curved line in the figure. The child has a probability of 0.5 to be rated at least at level *k*, if that level's difficulty is equal to his or her ability. The probability that the child will be rated at lower levels (e.g., *k-1, k-2)* is larger than 0.5. The probability that the child will be rated at higher levels (*k+1, k+2)* is smaller than 0.5.

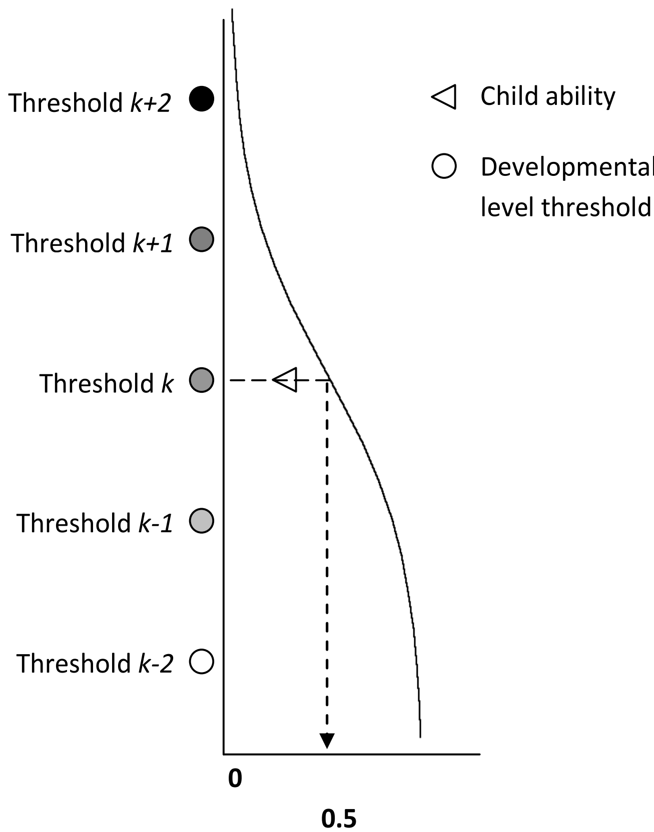


Figure 6.9—Thurstonian thresholds.

The thresholds define the bandwidths of developmental levels within each measure. The k^{th} threshold represents the upper bound of the k^{th} developmental level. For example, in Figure 6.10, the bandwidths of the MATH construct, *Math Operations*, are shown for the PS and SA measures (this construct does not exist in IT). Each "ladder" represents the measure in the corresponding instrument. The steps in the ladder represent the thresholds obtained from fitting the partial credit model. The PS and SA instruments contain the "Not Yet" level that is used for children who have not yet reached the first developmental level on the instrument. Therefore, the first threshold (i.e., the line between levels 0 and 1) indicates where the "Not

Yet" level ends and the first developmental level begins. For the PS instrument this threshold equals -2.9, and -0.48 for SA instrument[15].

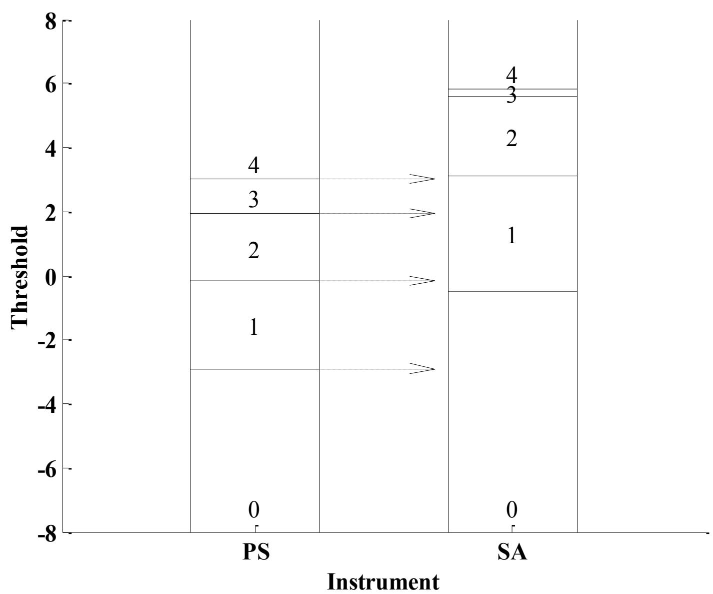


Figure 6.10—Math operations (MATH) levels' bandwidths – complete support of expected linkage.

In Figure 6.10, the last threshold in PS equals 3.02 logits and defines where the third developmental level ends and the fourth begins. Preschoolers whose math ability is higher than 3.02 are most likely to be rated at the fourth level by their teachers. Because thresholds define the upper bounds of their corresponding levels, the bandwidth for the last level has no upper bound. This makes sense because no matter what value of ability the child possesses above the last threshold, he or she will have to be rated at the last level on this measure. In the next few paragraphs we show examples of how levels of linked measures show different degrees of overlap.

[15] For the sake of clarity, all graphs shown in this chapter range between -8 to 8, thus the "Not Yet" level usually seems to span a wide range of logit values

Figure 6.10 juxtaposes the developmental bandwidths of the linked measures from two DRDP-Rs. Because the SA instrument describes more complex performances related to math operations, we would expect the SA levels to be higher (i.e., more difficult) than the PS levels. More specifically, we would expect the two measures to align at the point where our content analysis identified an expected linkage between levels as shown in Table 6.2. In this case, we expected level 4 in PS to link with level 2 in SA (see Table 6.2). Indeed, this expectation is supported by the data, as can be seen in Figure 6.10; level 4 in PS and level 2 in SA begin around the same logit value and therefore their overlap supports the expected linkage. To aid interpretation, we interpret the match in the direction from the younger age instrument to the older age instrument. Figure 6.10 also reveals that levels 2 and 3 in PS overlap with level 1 in SA, and level 1 in PS overlaps with the "Not yet" level in SA. Although these links were not derived from the content analysis of the DRDP-R articulation process, they indicate how teachers tend to interpret these levels.

By visually inspecting the matching of the levels we can determine to what extent the expected linkage structure is supported by the data. For example, in Figure 6.11, the developmental bandwidths for the LIT construct, *Letter and Word Knowledge,* are presented side by side. The expected linkage is between IT5→PS1 and PS4→SA1/2. The link between IT and PS is supported by the data. Level 5 in IT overlaps with about half of level 1 in PS. However, level 4 in PS aligns with level 2 in SA and has almost no overlap with level 1. Therefore, the expected linkage is only partially supported.

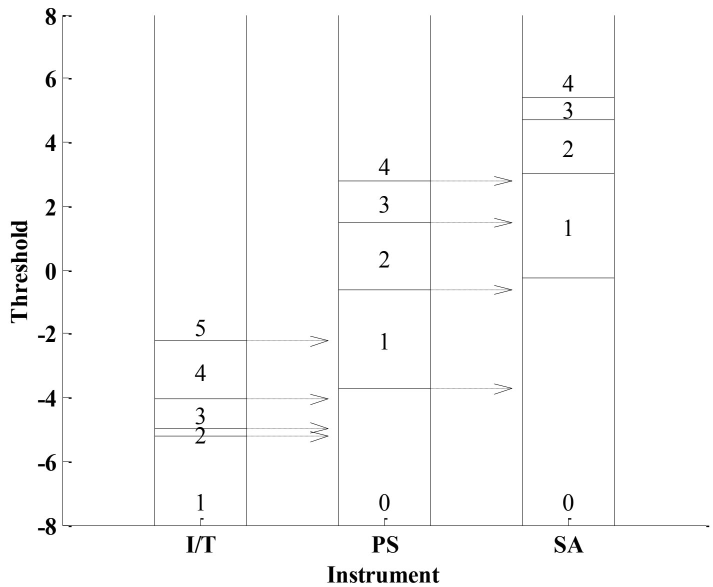


Figure 6.11—Letter and word knowledge (LIT) levels' bandwidths – partial support of expected linkage.

Another type of partial support is shown for the SOC construct, *Conflict Negotiation*, in Figure 6.12. The expected linkage is PS3, 4→SA1, 2. However, level 3 in PS overlaps with both levels 1 and 2 in SA whereas level 4 in PS overlaps with about half of level 2. Although the observed linkage is close to our expectations based on the content analysis, the link is not completely aligned. The measures show more overlap than expected, meaning the older instrument's levels seemed easier than they were supposed to be. In this case, level 3 in PS shows overlap with the expected level in SA and with the level above it. This pattern of partial support was common among the DRDP-R measures. The reverse pattern was also common. Specifically, many measures show partial support by having less overlap than expected, as shown in Figure 6.13, *Uses of Language in Conversation* construct (LANG indicator). The expected linkage was IT5, 6→PS2, 3. However, IT level 5 overlaps with level 1 in PS, and level 6 in IT aligns with level 2 in PS. The measures show less overlap than expected, meaning the older instrument's levels seemed more difficult than they are supposed to be.

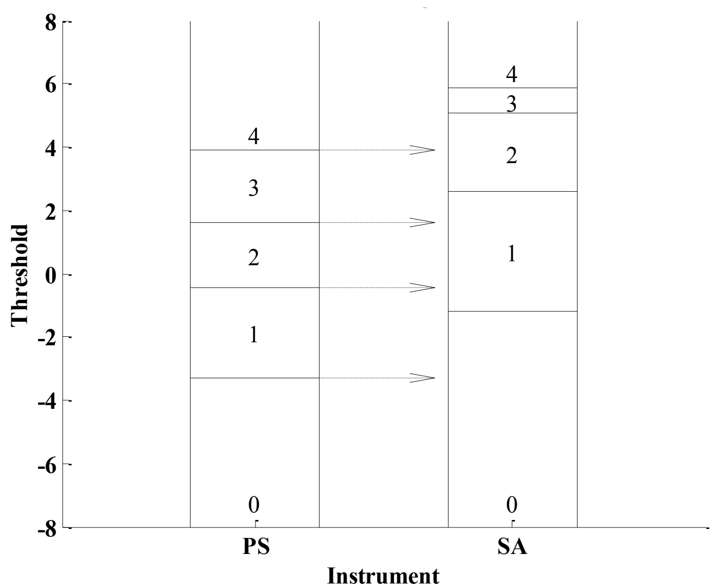


Figure 6.12—Conflict negotiation (SOC) levels' bandwidths – more overlap than expected linkage.

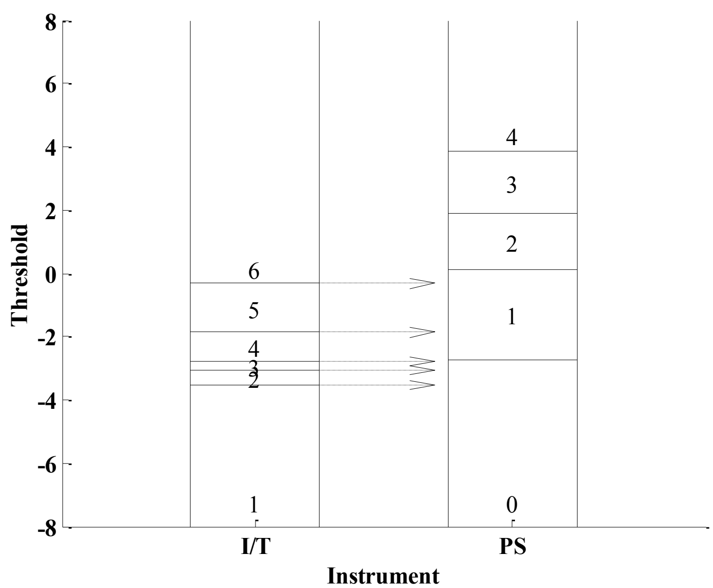


Figure 6.13—Uses of language in conversation (LANG) levels' bandwidths – less overlap than expected linkage.

Figure 6.14 shows how the levels from the *Fine Motor* construct (MOT indicator) align across the three instruments. As detailed in Table 6.2, we expected a link at IT6→PS1 and PS3, 4→SA1, 2. Figure 6.14 shows that level 6 in IT has no overlap with level 1 in PS, meaning the expected link was not supported by the data. However, the link between PS and SA does support our expectations. Many sets of measures had a complex structure across the age-groups, where some of the links behaved as expected, some showed partial support and some showed no support at all. The interested reader can find the graphs of all remaining constructs in Appendix D.

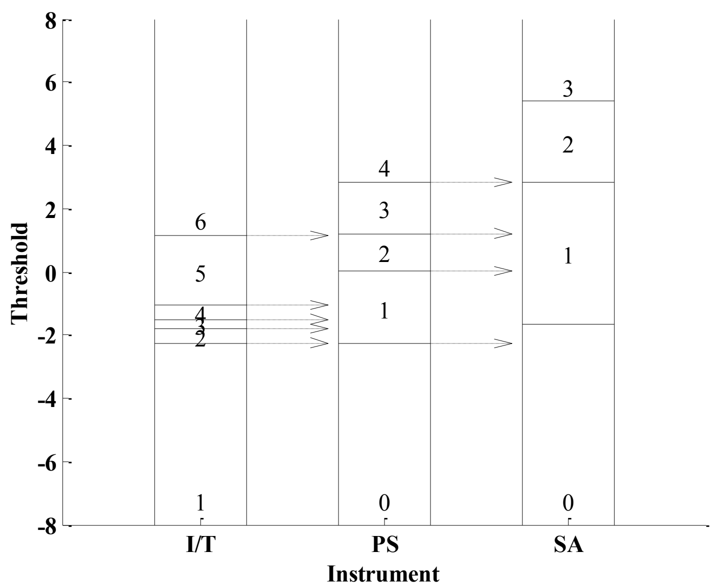


Figure 6.14—Fine motor skills (MOT) levels' bandwidths.

We investigated the observed level overlap of age-consecutive measures within each construct and compared it to the expected linkage as delineated in Table 6.2. If all the bridged levels linked as expected, we labeled those measures as showing "Complete alignment." The measures in Figure 6.12 are an example of complete alignment with expectations. When some levels linked as expected and some levels did not, we labeled the measures as showing "Partial alignment." The PS→SA link in Figure 6.11 shows partial alignment with expectations. When none of the levels linked as expected, we labeled the measures as showing "No alignment" with expectations. The IT→PS link in Figure 6.14 is such an example.

Table 6.6 presents the distribution of alignment types between consecutive DRDP-R pairs within each indicator. For example, the SELF indicator has two measures with links between each pair of instruments, for a total of four links. Table 6.6 shows that the two links between IT and PS had partial alignment with our expectations, one link between PS and SA had complete alignment and another had no alignment. Table 6.6 also shows that 33% of the expected linkage structure between pairs of instruments was well supported by the data. Overall about 87% of the links showed some support to our expectation (by having either complete or partial alignment), and only 13% of the links showed no support at all.

Table 6.6—Distribution of Types of Alignment between Observed and Expected Linkage Structure

	Alignment category					
	Complete		Partial		None	
Indicator	IT→PS	PS→SA	IT→PS	PS→SA	IT→PS	PS→SA
SELF		1	2			1
SOC	2	1	3	3		
LANG	2	1	1	2	1	
LRN		1	2	1		
COG	1	1	2	2		1
MATH	1	3	2	1	1	1
LIT	1	1	1	2		
MOT		1	1	1	2	
SH	1		1	2		
Total	8	10	15	14	4	3
Proportion	0.33		0.54		0.13	

We can also consider the number of aligned links for levels (as opposed to measures, as in the previous paragraph). The partial alignment column in Table 6.6 contains measures whose linked levels showed mixed patterns, meaning some levels were aligned with expectations and some were not. We investigated the number of individual levels within each alignment category. For example, consider the LRN construct, *Curiosity and Initiative*, measures in Figure 6.15. The expected linkage is IT3, 4/5→PS1, 2 and PS3/4→SA1. The figure shows that levels 3 and 5 in IT link as expected to levels 1 and 2 in PS, respectively. However, level 4 in IT links to level 1 rather than level 2 in PS. Similarly, level 3 in PS links to level 1 in SA, but level 4 does not.

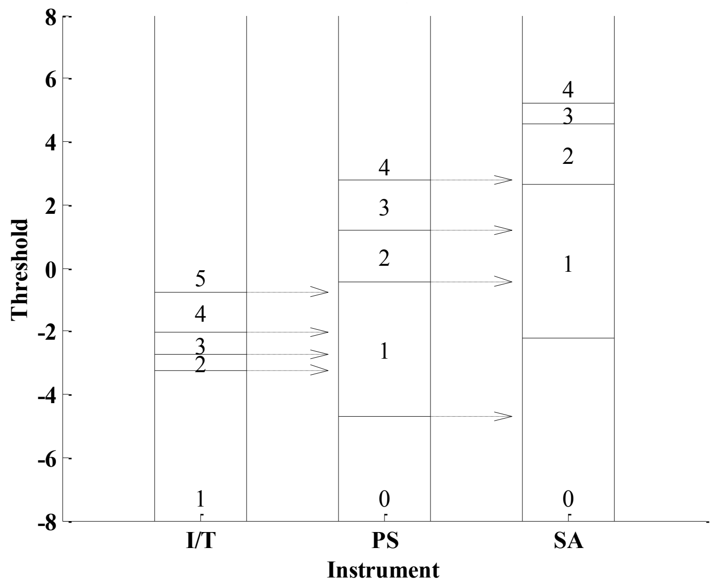


Figure 6.15—Curiosity and initiative (LRN) levels' bandwidths.

Within each pair of instruments, the links show partial alignment with expectations. However, if we look at the alignment pattern of individual levels, we find that 2 levels are aligned with expectations between IT and PS (i.e., levels 3 and 5) and one level is not aligned (level 4). Similarly, one level between PS and SA is aligned, and one is not (levels 3 and 4, respectively). Overall, 50% of individual levels showed complete alignment with expectations, and 27% of the levels showed no alignment. The remaining 23% of the levels showed partial alignment, meaning they overlapped with the expected level and with the level below or above it. The number of cases that overlapped above the expected level was approximately equal to the number of cases that overlapped below the expected level. This implies that the DRDP-R measures showed little or no bias towards being too difficult or too easy across the age groups. Some measures are harder than expected, and others are easier than expected, but there is no specific tendency of DRDP-R measures as a whole to be either too easy or too difficult.

6.4 Conclusion

In this chapter we discussed the importance of articulation of successive developmental assessments to enhance their practical usefulness for measuring the path of children's development. We explained how articulation can be investigated by linking each assessment instrument to developmental levels from adjacent instruments. The DRDP-R was presented as an example of a complex assessment system that was designed to track development across three consecutive age groups. The Double Instrument condition of the DRDP-R Calibration Study 2005 was designed to empirically examine the linkage structure of the DRDP-R. A sample of children was rated by their teachers using two consecutive instruments for each child. Our hypothesis was that if a child was rated at one of the last levels on a younger-age instrument, he or she was likely to be rated at one of the first levels on the older-age instrument, and vice versa. Specifically, within each indicator, we expected these rating tendencies to resemble the linkage structure we identified through the DRDP-R content analysis.

The results indicate that when teachers complete the adjacent-age instrument, in the main they do tend to rate children according to the expected linkages between the instruments. However, in some cases, the expected linkage based on content does not apply empirically. For example, a child who is rated low on one instrument may not have a level to describe his or her behavior on the other instrument. In such a case, teachers tend to rate this child at the “Not yet” or the first level of the adjacent instrument. Moreover, an advanced child may have been rated at the last level of the younger-age instrument, even though it does not describe all of this child’s capabilities. While the expected linkage suggests he or she should be rated at one of the lower levels in the older-age instrument, this child is likely to be rated higher at one of the levels from the older-age instrument that does describe his or her capabilities. The results indicate general support for the DRDP-R linkage structure presented in Table 6.2, but also quite a bit of variation from expectations – probably due to variations in how teachers rate children, and perhaps also due to experimental effects (e.g., the situation of rating a child twice using different, but related, instruments, is quite unusual).

In Figures 6.10 through 6.15 and Appendix D we present results from the psychometric analysis, which shed light on the underlying unidimensional latent variable that aligns measures of the same indicator across successive instruments. The analysis shows that for the majority of the constructs, the DRDP-R linkage structure is either supported or partially supported by the data. In other words, levels that we identified to mean the same thing were estimated to be similarly difficult to attain. In the rest of the cases, the observed link was either lower or higher than the expected link, usually by no more than one level. Overall, measures of the same indicator were generally linked between the last two and first two levels in every pair of adjacent instruments.

The analysis shows that the linkage structure between measures is generally an accurate expectation for teachers’ rating tendencies. When expectations were not met by the data, the link usually shifted, so the first level in the older-age instrument is linked to one level lower or higher than expected in the younger-age instrument (although the DRDP-Rs did not show any specific bias to either direction of misalignment). There could be many reasons why this shift occurs. First, we may have identified the wrong linkage structure. While this might be true in some cases, it is probably not true in others. Specifically, some linked levels are worded in exactly the same way, but still were estimated to have somewhat different difficulties. In this case, it may be the presence of the other categories that has made the difference. Second, the presence of relatively low- or high-achieving children in the sample may have caused many to be rated at the lower two levels of the older-age instrument. This would have a downward or upward "pulling" effect on the thresholds, causing the shifted links. As we have occasionally seen this pattern in other contingency tables of teachers' ratings, we suspect that this could be one of the main reasons for level misalignment. Finally, teachers may have misunderstood the behaviors described in the levels or misunderstood the purpose of the double-instrument study. Such misunderstanding may have caused some teachers to rate children in the same manner (i.e., at the exact same numerical level) on both instruments. These issues will be addressed in future studies of the DRDP-R linkage structure.

The main conclusion from this study is that the three parts of the DRDP-R are linked so that within each age group, children at various levels of development can be readily assessed. In other words, each instrument provides levels of assessment for the typical children in the age group it was designed to assess as well as relatively low-achieving and relatively high-achieving children in this age group. While this is very reassuring for the current version of the DRDP-R, there is a need for further research on the measures that do not link as expected. Moreover, future versions of the DRDP-R will include alignment to state standards, newer research findings and feedback from practitioners. It is crucial that any revisions to one instrument be applied to the linked levels of the other instruments, to the extent possible.

Establishing the theoretical and empirical linkage structure of the three instruments is a central validity check of the DRDP-R assessment system as a whole.

7 Special Study II: Examining the Consistency of Ratings Across Primary and Secondary Raters

7.1 Introduction

Inter-rater reliability is the degree of agreement between raters of the same child. This analysis examines how much measurement error is due to the variability in scoring between raters. An important consideration when conducting this investigation was to ensure the adequacy of the raters’ readiness (e.g., adequate and appropriate training and preparation) to complete the task. Before the Calibration Study in 2005, the CDD provided day-long training sessions for practitioners who participated in the study. The training session provided instruction for teachers on how to observe children and how to complete the DRDP-R on a selected sample of children under their care.

In order to exclude other potential sources of error, such as issues relating to data entry or inclusion/exclusion criteria, a standardized protocol for handling and analyzing the data sets was developed and followed.

This chapter describes the analysis that was conducted to investigate the degree to which pairs of teachers rated children in the same way. We examined the strength of agreement between raters on a single measure, and across measures within an indicator. Most often, the pair of raters consisted of a teacher and an administrator who interacted with or supervised the child. In some cases, another teacher who had the child under their care served as the second rater. These pairs of raters participated in Condition 2 (Double Rater) of the 2005 Calibration Study. To be included in the study, children had to have been under their primary rater’s care for at least two months. Raters were instructed to observe each child for 60 days or more before completing the DRDP-R. The two raters completed the instrument separately for the same child. Data from over 250 pairs of raters were compared and analyzed. The findings from the inter-rater reliability study are presented here.

7.2 Method

7.2.1 *Participants*

The participants data used in this analysis are the sample collected in Condition 2 of the Calibration Study 2005. On the DRDP-R, raters or observers were asked to specify their position title. From the DRDP-Rs received and successfully paired, three categories were developed to classify raters. The “Teacher” category consists of raters who commonly described themselves as teachers, permit teachers, Head Start teachers, master teachers, associate teachers, etc. Individuals in the “Supervisors” category identified themselves as supervisors, directors, site supervisors, program directors, etc. People in the “Other” category identified themselves as child care specialists, enrichment program assistants, and other titles. Rater 1 was the person at the care center who knew the child best. She was the person who spent the most time with the child, interacted most consistently with the child, and felt most comfortable being the primary rater for the child. Rater 1 was almost always in the role of the child’s primary teacher. Rater 2 was the secondary rater for the same child. Rater 2 was usually another teacher at the same care center or in some instances was a site supervisor (or other individual) at the center. Rater 2 was selected because she knew the child almost as well as Rater 1. Table 7.1 presents the distribution of these professions across the age groups.

Table 7.1—Distribution of Raters Profession Type for the Condition 2 Sample

Rater type	Infant/Toddler (N=145 pairs)		Preschool (N=84 pairs)		School-Age (N=42 pairs)	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Teacher	88%	30%	93%	93%	74%	21%
Supervisor	10%	61%	7%	--	12%	45%
Other	2%	8%	--	7%	14%	33%

Regardless of DRDP-R used (e.g., Infant/Toddler, Preschool, or School Age instrument), teachers made up the majority of first raters (or Rater 1). Table 7.1 shows that teachers composed 88% of the IT raters, 93% of the PS raters, and 74% of the SA raters. The percentages for the professional types of Rater 2 varied from the percentages for the professional types of Rater 1. Of the second raters, the majority in IT and SA were supervisors (61% and 40%, respectively) and 93% of the PS raters were teachers.

Overall, there were 271 pairs of raters included in the inter-rater reliability study. There were 145 IT rater pairs, 84 PS pairs, and 42 SA pairs. Pairs were included in the analysis if a child was under the primary rater's care for more than 10 hours per week (but not necessarily under the secondary rater's care for that long). Data were excluded if information on either rater was missing or if the child was under the primary rater's care for 10 or fewer hours per week or the child had been observed for fewer than 60 days.

7.3 Results

The results of the inter-rater reliability analysis are presented separately for each of the three age groups. However, the presentation is in the same order for each group. The first table shows agreement indices at the indicator level. The second table shows agreement indices at the measure level. The contents of the tables will be explained for each group (e.g., Infant/Toddler, Preschool, and School-Age).

7.3.1 *Results for the Infant/Toddler DRDP-R*

Table 7.2 provides inter-rater agreement at the indicator level (e.g., ratings were aggregated across all measures within an indicator). The results shown here are given for the ten individual indicators rather than the six indicator sets. First, the table provides the indicator name and the percent of raw agreement. This agreement index is the proportion of cases along the diagonal (e.g., both raters assigned the child to the same developmental level).

Table 7.2 also provides the percent of emergent agreement in the third column. For example, rater 1 placed a child at the “Exploring” developmental level and noted that the child was emerging into the next (“Developing”) level while rater 2 placed the same child at the “Developing” level. In this case, there is agreement between the raters if the “emerging” rating is taken into consideration.

The fourth column in Table 7.2 shows the percent of adjacent agreement. Adjacent agreement means that one rater rated a child at one developmental level and the other rater rated the same child at the next (or previous) developmental level without either rater indicating that the child was emerging from one level to the next. In addition, the table shows the combination of Raw and Emergent agreement, as well as Raw, Emergent and Adjacent agreement. Another index provided in the table is intraclass correlation (ICC). Intraclass correlation (Shrout & Fleiss, 1979) is the ratio of rating variance to total variance. This correlation compares the covariance of the ratings with the total variance of the data. For calculating this statistic, both children and raters are considered as random factors.

Table 7.2—Infant/Toddler Agreement Indices by Indicator

	Percent Agreement					ICC
Indicator	Raw	Emergent	Adjacent	R+ E	R+E+A	
Self Concept (SELF)	47	28	14	76	89	.71
Social Interpersonal Skills (SOC)	45	22	18	67	84	.64
Self Regulation (REG)	43	24	20	67	87	.65
Language (LANG)	43	25	18	68	86	.76
Cognitive (COG)	44	25	20	68	88	.66
Math (MATH)	41	22	26	62	88	.65
Literacy (LIT)	41	26	18	66	85	.66
Motor Skills (MOT)	47	26	15	73	87	.75
Safety & Health (SH)	48	28	18	75	93	.75

Table 7.3 provides inter-rater agreement information at the measure level. Information is presented for agreements for raw, emergent, and adjacent agreements, their combination, and the intraclass correlation.

Table 7.3—Infant/Toddler Agreement Indices by Measure

Measure	Percent Agreement				ICC
	Raw	Emergent	Adjacent	R + E +A	
Identity of Self and Connection to Others	69	15	13	97	.73
Recognition of Ability	59	12	23	94	.66
Self Expression	75	5	17	97	.73
Awareness of Diversity	56	12	24	92	.69
Empathy	53	13	24	90	.66
Interactions with Adults	65	6	24	94	.68
Relationships with Familiar Adults	52	8	26	87	.62
Interaction with peers	57	13	22	91	.62
Relationships with Familiar Peers	58	8	23	89	.60
Impulse Control	58	14	24	96	.68
Seeking Other's Help to Regulate Self	61	13	18	92	.68
Responsiveness to Other's Support	50	9	36	95	.60
Self Comforting	58	10	26	94	.64
Attention Maintenance	51	9	35	94	.60
Language Comprehension	55	12	26	94	.77
Responsiveness to Language	59	13	23	94	.77
Communication of Needs, Feelings, and Interests	54	13	22	89	.73
Reciprocal Communication	55	12	28	95	.74
Memory	64	10	20	93	.70
Cause and Effect	53	16	25	94	.68
Problem Solving	64	8	25	97	.64
Symbolic Play	55	9	25	89	.61
Curiosity	56	8	28	92	.60
Number	52	8	34	94	.67
Space and Size	52	11	31	93	.58
Time	56	9	29	94	.69
Classification and Matching	55	7	35	97	.63
Interest in Literacy	51	15	27	93	.69
Recognition of Symbols	55	11	26	92	.61
Gross motor	64	18	13	95	.80
Fine Motor	69	13	15	97	.76
Balance	56	10	24	91	.69
Eye-Hand Coordination	52	9	32	93	.69
Personal Care Routines	66	13	18	97	.76

Measure	Percent Agreement				ICC
	Raw	Emergent	Adjacent	R + E +A	
Safety	60	12	23	95	.73

Infant/Toddler agreements (percent Raw + Emerging +Adjacent from Table 7.2) at the indicator level ranged from 84% (Social Interpersonal Skills) to 93% (Safety and Health). Agreements at the measure level (percent Raw + Emerging +Adjacent from Table 7.3) ranged from 87% (*Relationships with Familiar Adults*) to 97% (*Identity of Self and Connection to Others*, *Self Expression*, *Problem Solving*, *Classification and Matching*, *Fine Motor*, and *Personal Care Routines*).

7.3.2 *Results for the Preschool DRDP-R*

Table 7.4 shows inter-rater agreement at the indicator level. The columns have the same meaning as in Table 7.2.

Table 7.4—Preschool Agreement Indices by Indicator

Indicator	Percent Agreement					ICC
	Raw	Emergent	Adjacent	R+ E	R+E+A	
Self Concept (SELF)	51	19	18	69	87	.66
Social Interpersonal Skills (SOC)	40	19	24	59	83	.59
Self Regulation (REG)	41	18	23	59	82	.60
Language (LANG)	47	20	20	67	86	.74
Learning (LRN)	39	23	23	62	85	.62
Cognitive (COG)	43	19	23	62	85	.64
Math (MATH)	44	23	21	67	88	.69
Literacy (LIT)	46	23	20	69	89	.74
Motor Skills (MOT)	42	25	16	67	83	.50
Safety & Health (SH)	39	22	21	61	82	.52

Table 7.5 provides inter-rater agreement information at the measure level. The columns have the same meaning as in Table 7.3. Information is presented for agreements for raw, emergent, and adjacent agreements, their combination, and the intraclass correlation.

Table 7.5—Preschool Agreement Indices by Measure

Measure	Percent Agreement				ICC
	Raw	Emergent	Adjacent	R + E +A	
Identity of self	61	6	21	88	.62
Recognition of own skills and accomplishments	62	10	24	95	.72
Expressions of empathy	45	11	30	86	.55
Building cooperative relationships with adults	49	12	31	93	.70
Building cooperative play with other children	50	12	33	95	.61
Developing friendships	49	6	39	94	.62
Conflict negotiation	52	10	29	90	.47
Awareness of diversity in self and others	52	6	31	89	.47
Impulse Control	45	11	31	87	.48
Taking turns	52	8	33	93	.63
Shared use of space and materials	54	6	29	89	.65
Comprehends meaning	57	11	23	91	.72
Follows increasingly complex instructions	63	8	23	94	.80
Expresses self through language	55	12	25	93	.73
Uses language in conversation	53	7	31	91	.72
Curiosity and initiative	51	13	33	96	.72
Engagement and persistence	47	13	30	90	.51
Memory and knowledge	50	10	30	89	.49
Cause and effect	55	6	28	89	.67
Engages in problem solving	47	11	35	93	.67
Socio-dramatic play	61	9	27	96	.74
Number sense: Understands quantity and counting	61	11	23	94	.68
Number sense: Math operations	49	16	29	94	.70
Shapes	59	15	26	100	.75
Classification	56	11	28	95	.73
Measurement	59	16	16	91	.66
Patterning	54	5	29	89	.67
Time	51	9	27	86	.45
Interest in literacy	63	6	23	92	.69
Concepts of print	50	9	35	94	.64
Letter and word knowledge	63	8	23	94	.76
Phonological awareness	63	9	23	95	.67
Emerging writing	68	5	22	95	.79
Gross motor movement	58	8	23	89	.53
Balance	55	12	23	90	.36
Fine motor skills	60	9	26	94	.59
Personal Care Routines	55	11	31	98	.57

Measure	Percent Agreement				ICC
	Raw	Emergent	Adjacent	R + E +A	
Personal safety	48	11	23	82	.42
Understanding healthy lifestyle	45	13	35	93	.57

Preschool agreements (% R+E+A) at the indicator level ranged from 82% (Self Regulation and Safety and Health) to 89% (Literacy). Agreements at the measure level ranged from 82% (*Personal safety*) to 100% (*Shapes*).

7.3.3 *Results for the School-Age DRDP-R*

Table 7.6 indicates the inter-rater agreement at the indicator level. Information is presented for agreements for raw, emergent, and adjacent agreements, their different combinations, and the intraclass correlation.

Table 7.6—School-Age Agreement Indices by Indicator

Indicator	Percent Agreement					ICC
	Raw	Emergent	Adjacent	R+ E	R+E+A	
Self Concept (SELF)	35	15	29	50	79	.56
Social Interpersonal Skills (SOC)	41	16	22	57	80	.65
Self Regulation (REG)	29	20	24	49	73	.56
Language (LANG)	26	26	35	52	87	.79
Learning (LRN)	35	17	21	51	73	.67
Cognitive (COG)	35	20	28	54	82	.66
Math (MATH)	58	12	15	70	85	.72
Literacy (LIT)	55	17	15	72	88	.66
Motor Skills (MOT)	42	17	23	58	81	.58
Safety & Health (SH)	39	15	26	54	80	.44

Table 7.7 provides inter-rater agreement information at the measure level. Information is presented for agreements for raw, emergent, and adjacent agreements, their combination, and the intraclass correlation.

Table 7.7—School-Age Agreement Indices by Measure

Measure	Percent Agreement				ICC
	Raw	Emergent	Adjacent	R + E +A	
Identity of self and connection to others	33	19	26	79	.53
Self-esteem	36	12	31	79	.59
Empathy	38	14	21	74	.59
Interactions with adults	38	12	31	81	.74
Friendship	48	24	12	83	.66
Conflict negotiation	45	17	17	79	.69
Awareness of diversity: Appreciation of differences and similarities	38	12	31	81	.51
Impulse Control	31	17	21	69	.65
Follows rules	26	24	26	76	.44
Comprehension of oral language	24	29	33	86	.81
Expression of oral language	29	24	36	88	.77
Pursuit of understanding	38	19	26	83	.77
Task persistence	31	14	17	62	.57
Memory/knowledge	40	19	29	88	.69
Cause and effect relationships	36	19	29	83	.64
Problem-solving	33	21	21	76	.58
Demonstrates inventiveness/inventive play	29	19	33	81	.69
Number sense: Basic math skills (operations)	48	14	14	76	.70
Measurement	48	7	24	79	.67
Shapes	35	26	22	83	.45
Time	48	22	13	83	.80
Interest in literacy	40	24	26	90	.73
Writing	38	19	19	76	.65
Decoding (word recognition and use)	65	17	17	100	.84
Comprehension of written materials	9	30	13	52	.38
Movement and coordination (gross motor skills)	43	14	29	86	.54
Dexterity (fine motor skills)	40	19	17	76	.62
Personal Care Routines	52	10	21	83	.45
Safety	40	12	29	81	.50
Understanding healthy lifestyle	26	12	33	71	.37
Exercise and fitness	36	26	21	83	.36

School-age agreements (% R+E+A) at the indicator level ranged from 73% (Self Regulation and Learning) to 88% (Literacy). Agreements at the measure level ranged from

52% (*Comprehension of written materials*) to 100 percent (*Decoding – word recognition and use*).

7.4 Conclusion

As anticipated, inter-rater reliability coefficients at the indicator levels and measure levels were low to medium using raw matching. By combining raw matching with emerging and adjacent results, the inter-rater reliability coefficients are higher. These results are in agreement with the expected results given the stage of development of the instrument. Overall for the Infant/Toddler, Preschool, and School-Age instruments, the inter-rater reliability results are mixed. Among all the age groups, the strongest agreement was found in IT, where at the indicator level mean raw agreement was 44%, mean agreement with emerging categories was 69%, and 87% if adjacent categories were also included. At the measure level, mean agreement was higher for raw agreement (58%) and for raw + emergent + adjacent (93%). The mean ICCs were 0.69 and 0.68 at the indicator and measure levels, respectively. The PS agreement levels were slightly below that of IT. For example, mean raw and emergent agreement was 64%, and mean ICC was 0.63, for both the indicator and measure levels. The SA instrument had somewhat lower agreement levels (mean raw and emergent agreement of about 57% and mean ICCs of 0.63 and 0.59 for the indicator and measure levels, respectively). Additionally, SA had the only negative ICC value for the LIT measure *Comprehension of written materials*. Finally, SA had lower mean agreement at the measure level than the indicator level, which is the opposite pattern from the IT and PS instruments. This suggests that the SA instrument may not be as cohesive as PS and IT. Indeed, it covers the largest age span (seven years).

Overall, the results confirmed our expectation for low to intermediate agreement levels. The data used for this analysis was collected from teachers who were not used to rating with the DRDP-R. It is anticipated that raters will become more consistent over time as they become more familiar with the instrument and have opportunities to both observe children and assess them via the DRDP-R.

8 General Discussion

8.1 Overview of Findings

The DRDP-R is an embedded observational assessment system designed to be used to identify and follow children's progress from birth through age 12 in programs funded by the state of California. The DRDP-R assessment system provides a framework for educational practitioners (e.g., infant/toddler caregivers, pre-school teachers, and before/after school teachers) to observe, document, and rate children's development in socio-emotional, cognitive, physical and behavioral domains. The DRDP-R instruments are designed to collect evidence of children's development using multiple measures of progress in three age groups: Infant/Toddlers, Preschoolers and School-Age (i.e., kindergarten through 12 years). The DRDP-R provides practitioners, policy makers and researchers with information about children's socio-emotional, cognitive, physical and behavioral development.

An extensive literature review, as discussed in Chapter 2, was conducted to develop the measures so that they would be representative of the domains, and be based on leading child development research for the different age groups. Researchers in early-childhood education and developmental psychology and seasoned practitioners in Infant/Toddler, Preschool, and School-Age care programs were an integral part of the development of the DRDP-R. The materials were reviewed by measurement researchers to ensure alignment of the measure and levels within the measures to the intended constructs. The literature review provided a research basis for the development and later refinement of the measures represented on the DRDP-R. Measures were placed in six domains according to the child development literature. The six domains are (1) Self and Social Development, (2) Self Regulation and Self Care, (3) Language and Literacy, (4) Cognitive Development, (5) Mathematical Development, and (6) Motor

Development. The literature review is grouped in these domains as a natural extension of the current state of the child development field of study.

After the literature review was complete, research began in the form of the Calibration Study to collect empirical evidence regarding the validity and reliability of the DRDP-R. The findings from the study are summarized below. For more information see Chapter 3.

In the spring of 2005, the DRDP-R assessment system was tested across the state of California in the Calibration Study. The purpose of the Calibration Study was to calibrate the three DRDP-Rs together and to provide data to enable the investigation of the reliability, validity, and fairness of the instruments. In order to calibrate the instruments, the study was designed to collect rating data on a large sample of children. The data provided valuable evidence for scaling the developmental levels of each indicator within and across age-groups. A representative sample of children across age groups, gender, and ethnicity was gathered so that the calibrated instruments could be used to make valid inferences about the population. Moreover, the selection of a representative sample was important to the evaluation of the sensitivity of the instrument to ethnicity and gender. A secondary goal of the study was to examine the differences and similarities between ratings made by raters who have primary contact with the child (e.g., teachers, caregivers), versus raters who have only secondary contact with the child (e.g., teacher aid, administrator). Because of the developmental perspective of the DRDP-R, there is a need to address the articulation between age-contiguous instruments (e.g., IT and PS, PS and SA). Many CDD programs serve children whose ages collectively span two or more instruments; the continuity of our developmental measurement indicators from one instrument to the next has to be clear and obvious to teachers. A confirmation that this requirement is satisfied would be to find empirically that teachers mark the developmental level a child has reached on a measure at consistent points on contiguous instruments. Establishing this continuity is also essential in constructing scales for our indicators that span the entire DRDP-R age range. Establishing this linkage structure across instrument is important evidence to support the calibration of the DRDP-R assessment system.

As discussed in the previous chapters, the Calibration Study had 4 goals: to calibrate the three DRDP-Rs (Chapter 4), to examine the sensitivity of the DRDP-Rs to ethnicity and gender (Chapter 5), to study the developmental linkage structure across the three DRDP-Rs (Chapter 6), and to evaluate the inter-rater reliability of the DRDP-R (Chapter 7). To accommodate these goals, the Calibration Study included three conditions under which teachers rated children on the DRDP-R. The chapters identified above provide the results and more detailed information about the different parts of the study although brief summaries of findings are included below.

In Chapter 4, the results indicate that the DRDP-R assessment system was reasonably well calibrated using the Calibration Study sample. The instruments were shown to have desirable technical properties that support the objective of validity evidence. The calibration analysis produced a usefully reliable set of estimates for developmental levels' difficulties which appropriately span the underlying developmental dimension across all three age groups. Moreover, the instruments may be used to obtain reliable estimates of children's abilities on six indicators of Desired Results (DR) so long as the standard procedures are followed.

In Chapter 5, an investigation to identify the validity evidence based on internal structures was discussed. This investigation showed that, as anticipated, the ability distributions generally moved forward for each age range. Correlations among developmental domains are not atypical. For example, the domains on the PISA (a test of 15 year olds' academic achievement) are also correlated.

As a part of the internal structures investigation, a DIF analysis was completed to look for items that might show signs of interaction with demographic characteristics of children in the sample. The main conclusion from the DIF analysis is that, for the most part, DRDP-R measures do not show bias to a large degree towards certain demographic groups. Even when statistically significant DIF was detected at the indicator level, actual DIF was found only in a handful of measures out of the whole instrument. Overall, there are gender DIF for the COG indicator in the infant toddler instrument, gender DIF for the LRN/COG and MOT indicators, African-American/European-American DIF for the REG/SH and LANG/LIT indicators, Hispanic-American/European-American DIF for the SELF/SOC, LANG/LIT, LRN/COG and MATH indicators in the preschool instrument, and gender DIF for the SELF/SOC and REG/SH indicators in the school age instrument. After item-wise inspection, only *Socio-Dramatic Play* of LRN/COG indicator for gender DIF and *Comprehends Meaning*, *Uses Language in Conversation*, and *Emerging Writing* of LANG/LIT indicator, *Memory and Knowledge* and *Cause and Effect* of LRN/COG indicator, and *Patterning* and *Time* of MATH indicator for Hispanic-American/European-American DIF in the preschool instrument, and *Exercise and fitness* of REG/SH indicator for gender DIF in the school age instrument were found to exhibit at least intermediate DIF.

In Chapter 6, information regarding the articulation of the DRDP-R across age groups was presented. Psychometric analysis sheds light on the underlying unidimensional latent variable that aligns measures of the same indicator across successive instruments. The analysis shows that for the majority of the constructs, the DRDP-R linkage structure is either supported or partially supported by the data. In other words, levels that we identified to mean the same thing were estimated to be similarly difficult to attain. In the rest of the cases, the observed link was either lower or higher than the expected link, usually by no more than one level. Overall, measures of the same indicator were generally linked between the last two and first two levels in every pair of adjacent instruments. The analysis shows that the linkage structure between measures is generally an accurate expectation for teachers' rating tendencies. When expectations were not met by the data, the link usually shifted, so the first level in the older-age instrument is linked to one level lower or higher than expected in the younger-age instrument (although the DRDP-Rs did not show any specific bias to either direction of misalignment).

There could be many reasons why this shift occurs. First, we may have identified the wrong linkage structure. While this might be true in some cases, it is probably not true in others. Specifically, some linked levels are worded in exactly the same way, but still were estimated to have somewhat different difficulties. In this case, it may be the presence of the other categories that has made the difference. Second, the presence of relatively low- or high-achieving children in the sample may have caused many to be rated at the lower two levels of the older-age instrument. This would have a downward or upward "pulling" effect on the thresholds, causing the shifted links. As we have occasionally seen this pattern in other contingency tables of teachers' ratings, we suspect that this could be one of the main reasons for level misalignment. Finally, teachers may have misunderstood the behaviors described in the levels or misunderstood the purpose of the double-instrument study. Such misunderstanding may have caused some teachers to rate children in the same manner (i.e., at the exact same numerical level) on both instruments. These issues will be addressed in future studies of the DRDP-R linkage structure.

The main conclusion from the study examining the articulation of the DRDP-R across age groups is that the three parts of the DRDP-R are linked so that within each age group, children at various levels of development can be readily assessed. Future versions of the DRDP-R will include alignment to state standards, newer research findings and feedback from practitioners. It is crucial that any revisions to one instrument be applied to the linked levels of the other instruments, to the extent possible. Establishing the theoretical and empirical linkage structure of the three instruments is a central validity check of the DRDP-R assessment system as a whole.

Chapter 7 presented information regarding the consistency of ratings of children across primary and secondary raters. As anticipated, inter-rater reliability coefficients at the indicator levels and measure levels were low to medium using raw matching. By combining raw matching with emerging and adjacent results, the inter-rater reliability coefficients are higher. These results are in agreement with the expected results given the stage of development of the instrument. Overall for the Infant/Toddler, Preschool, and School-Age instruments, the inter-rater reliability results are mixed. Among all age groups, the strongest agreement was found in IT, where at the indicator level mean raw agreement was 44%, mean agreement with emerging categories was 69%, and 87% if adjacent categories were also included. At the measure level, mean agreement was higher for raw agreement (58%), and for raw + emergent+ adjacent (93%). The mean ICCs were 0.69 and 0.68 at the indicator and measure levels, respectively. The PS agreement levels were slightly below that of IT. For example, mean raw and emergent agreement was 64%, and mean ICC was 0.63, for both the indicator and measure levels. The SA instrument had somewhat lower agreement levels (mean raw and emergent agreement of about 57% and mean ICCs of 0.63 and 0.59 for the indicator and measure levels, respectively). Additionally, SA had the only negative ICC value for the LIT measure *Comprehension of* *written materials*. Finally, SA had lower mean agreement at the measure level than the indicator level, which is the opposite pattern than the IT and PS instrument. This suggests that the SA instrument may not be as cohesive as PS and IT. Indeed, it covers the largest age span (seven years). Overall, the results confirmed our expectation for low to intermediate agreement levels. The data used for this analysis was collected from teachers who were not used to rating with the DRDP-R. It is anticipated that raters will become more consistent over time as they become more familiar with the instrument and have opportunities to both observe children and assess via the DRDP-R.

8.2 Limitations and Extensions

There are limitations and extensions of the DRDP-R. First, because very young children cannot comment on or complete assessments themselves, observations of child behavior need to be conducted to collect information about their development. Direct assessment is not possible; hence it was necessary to develop an indirect assessment.

In addition, data were collected by caregivers and teachers and not by trained researchers. However, caregivers and teachers participating in the study received training by DRDP experts.

8.3 Future Research

In addition to the research presented here, evidence based on response processes, relation to other variables, and consequence are important aspects of validity that should also be examined. Evidence that is based on response processes is required to support and bridge evidence based on test content. According to the *Standards for Educational and Psychological*

Testing (APA, AERA, NCME, 1999, p.12), "Theoretical and empirical analyses of the response processes of test takers can provide evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinee." Evidence based on response processes generally comes from a close examination of the individual respondent and his/her individual responses (Wilson, 2005). For example, if an assessment is designed to assess scientific reasoning it becomes important to verify that scientific reasoning is taking place as the student is taking the exam and he/she is not using a test-writing strategy or other prop to answer questions. Messick (1987) writes that it has long been presumed that the score a person attains on a test is determined by relevant responses to the specific content and in addition, it is usually taken for granted that this score reflected the respondent's knowledge on achievement tests (for example). However, he suggests that "evidence in support of these presumptions is critical in establishing the meaning or construct validity of the scores" (Messick, 1988, p. 1). Snow and Lohman (1989) suggest that cognitive science can assist researchers by helping to improve understanding of constructs of interest, by illuminating how both content and format of tests can be made to conform better to the underlying cognitive processes, and to serve as inspiration for new theories of aptitude, learning, and achievement. In their article, Snow and Lohman (1989) discuss current cognitive theories and educational measurement's relation to them and ultimately decide that "it is theory-based assessments that demonstrably do good, for students and educational institutions, individually and collectively" (p. 321). When a student completes an assessment, the student engages in a variety of psychological processes that produce a set of responses to items (Snow and Lohman, 1989). By investigating evidence based on response processes we observe the student in the moment as s/he goes through the psychological processes necessary to respond to items. Evidence based on response processes should be included in further studies of the DRDP-R.

Evidence based on the relation of a measure to other variables addresses questions about the degree to which relationships between the assessment outcomes and external variables (such as other tests hypothesized to measure the same constructs) are consistent with hypothesized relationships based on theory or previous results.

Evidence based on consequences is an important aspect of validity that might not be readily available during a validity investigation because this aspect of validity relates to actual use of an instrument. Wilson (2005, p. 170) writes that "if a general category of use of a particular instrument is found to have negative consequences that is an overriding consideration that should be taken into account when deciding whether to use an instrument." Evidences based on consequences that can be traced to either construct irrelevant variance or construct underrepresentation are valid issues to consider within the validation process. However, if evidence cannot be traced to construct irrelevant variance or construct underrepresentation it might be an issue of policy rather than validity (*Standards*, APA, AERA, and NCME, 1999). Evidence based on consequences of the DRDP-R should be examined after the instrument is put into use in the field.

Appendixes

A. Example of the Formatting of the DRDP Instrument Before Revision

B. Example of the Formatting of the DRDP-R Instrument

C. DRDP-R Measures and Indicators for Three Age Groups

D. Graphs of Cross-Age Threshold Alignment of DRDP-R Measures

E. Psychometric Information of DRDP-R Measures

F. Test Information Plots

G. Disattenuated Correlations Among the Six DRDP-R Domains

H. Analyses Codes (*ConQuest* Codes)

Appendix A – Example of the Formatting of the DRDP Instrument Before Revision

Theme	Child Desired Result 2: Children are effective learners.	Observation	Not Yet	Emerging	Almost Mastered	Fully Mastered	Comments/Observations
	Indicator 1: Children are interested in learning new things.						
Interest in learning	16. Observes and examines natural phenomena through senses (e.g., notices different types of bugs, asks why it rains)	A	❍	❍	❍	❍	
		B	❍	❍	❍	❍	
	17. Combines activities, materials, and equipment in new ways (e.g., builds tent by using sheet or blanket around table, uses Play-Doh to make pretend food)	A	❍	❍	❍	❍	
		B	❍	❍	❍	❍	
	Indicator 2: Children show cognitive competence and problem-solving skills through play and daily activities.						
Cognitive competence	18. Acts out plays, stories, or songs (e.g., uses body and sounds to express rhythm; makes up plays or songs about common fables, stories, or familiar characters)	A	❍	❍	❍	❍	
		B	❍	❍	❍	❍	
	19. Completes increasingly complex puzzles (e.g., single, cut-out figures to four-piece interlocking to eight- or ten-piece puzzles)	A	❍	❍	❍	❍	
		B	❍	❍	❍	❍	
	20. Stays with or repeats a task (e.g., finishes a puzzle, asks that block structure be left to work on after snack, makes a really long Play-Doh snake out of many pieces)	A	❍	❍	❍	❍	
		B	❍	❍	❍	❍	

Appendix B – Example of the Formatting of the DRDP-R Instrument

▽ Desired Result 2: Children are effective learners

▽ Indicator: COG – Preschoolers show cognitive competence and problem-solving skills through play and daily activities

Preschool

▸ Measure 21: Socio-dramatic play

Definition: Child learns to play with others using organized role-playing and symbolic play

1. Mark the highest developmental level the child has mastered.

Exploring	Developing	Building	Integrating
◯	◯	◯	◯
Engages in brief pretend play on own ◯ Not yet at first level	Engages in brief pretend play with a peer, sharing materials or ideas	Takes a role in a play situation with other children, but without planning the role or the pretend play	Takes a role in a play situation with other children where they have agreed on roles and how they will pretend play
Examples			
▸ Sits in a box or on a chair and pretends to drive. ▸ At the sandbox, pretends to make a cake by mixing sand and water in a pail. ▸ Uses plastic banana as telephone and pretends to call Grandma.	▸ Sits in box and says to another child, *"I'm driving the bus to take kids to school."* ▸ Pretends to pour milk into cups and gives a cup to a peer. ▸ Stirs with a spoon in a bowl, pretends to taste, and says to a child who is also cooking, *"It's not ready yet."* ▸ Pretends to be a gas station attendant and pumps gas for trikes.	▸ Joins in when he sees two children pretending to drive a bus, but does not talk to them about what role he will play. ▸ In a dinnertime dramatic play sequence with peers, plays the parent or child having dinner at the small table. ▸ In a 'visit to the doctor' dramatic play sequence, plays the doctor using the stethoscope and placing bandages on another child. ▸ Plays superhero game, rescuing another child.	▸ Plays school bus, with one child playing the driver, another playing the child, and another the mommy helping her child. ▸ In block area, children create a zoo and assign roles such as zookeeper, cage cleaners, tour guide/bus driver, and bird keeper. ▸ Plays school with other children and assigns roles—*"I'll be the teacher, you be the calendar helper, and you be the snack helper."*

2. Record evidence for this rating here. (Use back for more space.)

3. Mark here if child is emerging to the next level. ◯

4. If you are unable to rate this measure, explain why.

Measure 21 **Socio-dramatic play** **COG 4 (of 4)**

© 2005 by California Department of Education – All rights reserved

Appendix C – DRDP-R Measures and Indicators for Three Age Groups

Measures of each indicator set are clustered by their age group. For brevity, this list does not include the developmental levels' descriptors and examples for each measure (these can be found at http://www.cde.ca.gov/sp/cd/ci/drdpforms.asp.). The last two columns show the numbering and labeling used in the 2005 Calibration Study (the letters i, p and s in the CS label column refer to IT, PS and SA, respectively).

Age	Label	Measure name	Measure Definition	CS #	CS label
IT	SELF1	Identity of Self and Connection to Others	Child shows awareness that self is distinct from and also connected to others	1	i_SELF1
IT	SELF2	Recognition of Ability	Child evaluates own ability to do things and shows interest in others' evaluation of self	2	i_SELF2
IT	SELF3	Self Expression	Child explores own action, makes presence known in situations, and outwardly expresses feelings to others	3	i_SELF3
IT	SELF4	Awareness of Diversity	Child shows awareness of similarities and differences between self and others, as well as awareness of similarities and differences between people	4	i_SELF4
IT	SOC1	Empathy	Child shows awareness of other's feelings and responds to expressions of feelings by others	5	i_SOC1
IT	SOC2	Interactions with Adults	Child interacts effectively with both familiar and somewhat familiar adults	6	i_SOC2A
IT	SOC3	Relationships with Familiar Adults	Child forms close relationships or attachments with familiar adults	7	i_SOC2B
IT	SOC4	Relationships with Familiar Peers	Child forms relationships with specific peers	9	i_SOC3B
IT	SOC5	Interaction with peers	Child interacts effectively with a peer or small groups of peers	8	i_SOC3A
PS	SELF1	Identity of self	Child shows increasing awareness of own preferences and experiences as separate from those of others	10	p_SELF1
PS	SELF2	Recognition of own skills and accomplishments	Child evaluates and takes pleasure in own ability to perform skillfully	11	p_SELF2
PS	SOC1	Expressions of	Child shows awareness of other's	12	p_SOC1

Age	Label	Measure name	Measure Definition	CS #	CS label
		empathy	feelings and responds to expressions of feelings by others in ways that are appropriate to the other person's needs		
PS	SOC2	Building cooperative relationships with adults	Child works cooperatively with adults, through sharing and joint planning and problem solving	13	p_SOC2
PS	SOC3	Developing friendships	Child forms close relationships with specific peers, sharing experiences and activities	15	p_SOC3B
PS	SOC4	Building cooperative play with other children	Children interacts with children through play that becomes increasingly cooperative and towards a shared purpose	14	p_SOC3A
PS	SOC5	Conflict negotiation	Child learns how to understand the needs of other children and to negotiate constructively within the constraints of rules and values	16	p_SOC4
PS	SOC6	Awareness of diversity in self and others	Child acknowledges and responds to similarities and differences between self and others, and appreciates the value of each person in diverse communities	17	p_SOC5
SA	SELF1	Identity of self and connection to others	Child shows increasing awareness or understanding of self and his or her connection to others	18	s_SELF1
SA	SELF2	Self-esteem	Child makes positive judgments about self and his/her own abilities in increasingly broad contexts	19	s_SELF2
SA	SOC1	Empathy	Child shows increasing awareness of others' feelings and experiences and responds appropriately through words or actions	20	s_SOC1
SA	SOC2	Interactions with adults	Child develops positive relationships with increasingly larger groups of adults and acknowledges adult's perspective, while expressing clear sense of own self	21	s_SOC2
SA	SOC3	Friendship	Child develops one or more close relationships with peers and extends concept of friendship beyond his/her community	22	s_SOC3
SA	SOC4	Conflict negotiation	Child resolves conflicts by proposing solutions that consider the needs of others and extends concept of negotiation beyond his/her community	23	s_SOC4
SA	SOC5	Awareness of diversity: Appreciation of differences and similarities	Children show awareness, acceptance, understanding, and appreciation of others' special needs, genders, family structures, ethnicities, cultures, and languages	24	s_SOC5
IT	REG1	Impulse Control	Child responses to internal and external stimuli	1	i_REG1

Age	Label	Measure name	Measure Definition	CS #	CS label
IT	REG2	Seeking Other's Help to Regulate Self	Child manages needs through seeking or relying on assistance from other people	2	i_REG2A
IT	REG3	Responsiveness to Other's Support	Child is responsive to other's assistance with self regulation	3	i_REG2B
IT	REG4	Self Comforting	Child comforts self in response to distress from either internal or external stimulation	4	i_REG3
IT	REG5	Attention Maintenance	Child attends to things or the environment when interacting with others or exploring play materials	5	i_REG4
IT	SH1	Personal Care Routines	Child responds to and initiates personal care routines	6	i_SH1
IT	SH2	Safety	Child awareness of safety	7	i_SH2
PS	REG1	Impulse Control	Child develops strategies for responses in an increasingly socially appropriate way	8	p_REG1
PS	REG2	Taking turns	Child develops increased understanding of taking turns and begins to propose strategies for taking turns	9	p_REG2A
PS	REG3	Shared use of space and materials	Child develops the ability to with others, and begins to initiate of space and objects	10	p_REG2B
PS	SH1	Personal Care Routines	Child responds to and initiates personal care routines that support healthy growth and help prevent the spread of infection	11	p_SH1
PS	SH2	Personal safety	Child awareness of safety practices that minimize risk and support healthy growth	12	p_SH2
PS	SH3	Understanding healthy lifestyle	Child awareness of safety practices that minimize risk and support healthy growth	13	p_SH3
SA	REG1	Impulse Control	Child shows ability to regulate responses to internal and external stimuli in increasingly broad settings	14	s_REG1
SA	REG2	Follows rules	Child shows ability to follow rules in increasingly broad settings and understands the purpose of having rules	15	s_REG2
SA	SH1	Personal Care Routines	Child shows increasing independence in following personal care routines	16	s_SH1
SA	SH2	Safety	Child shows increasing independence in following rules for personal safety	17	s_SH2
SA	SH3	Understanding healthy lifestyle	Child shows increasing independence in making healthy lifestyle choices	18	s_SH3
SA	SH4	Exercise and fitness	Child shows increasing independence in participating in exercise and fitness activities	19	s_SH4
IT	LANG1	Language	Child shows understanding of that	1	i_LANG1A

Age	Label	Measure name	Measure Definition	CS #	CS label
		Comprehension	represents ideas		
IT	LANG2	Responsiveness to Language	Child acts or communicates in response to language	2	i_LANG1B
IT	LANG3	Communication of Needs, Feelings, and Interests	Child uses and nonverbal communication to convey needs, feelings, and interests	3	i_LANG2A
IT	LANG4	Reciprocal Communication	Child engages in back-and-forth communication or conversation	4	i_LANG2B
IT	LIT1	Interest in Literacy	Child shows interest in books, songs, rhymes, finger plays, and stories	5	i_LIT1
IT	LIT2	Recognition of Symbols	Child shows awareness that symbols and pictures represent people, objects, and actions	6	i_LIT2
PS	LANG1	Comprehends meaning	Child receives, decodes, and responds to oral and understands increasingly complex words, phrases, and ideas	7	p_LANG1A
PS	LANG2	Follows increasingly complex instructions	Child receives, decodes, and responds to directions to complete one or more steps	8	p_LANG1B
PS	LANG3	Expresses self through language	Child uses to communicate with increasingly complex words and grammar	9	p_LANG2A
PS	LANG4	Uses language in conversation	Child engages in back-and-forth communication or conversations following the appropriate social use of language	10	p_LANG2B
PS	LIT1	Interest in literacy	Child shows interest in books, songs, rhymes, stories, writing, and other activities, and seeks information in written text	11	p_LIT1A
PS	LIT2	Letter and word knowledge	Child shows awareness of symbols, letters, and words in the environment, and their relationship to sounds	13	p_LIT2A
PS	LIT3	Emerging writing	Child begins to use scribbles, symbols, letters, and words to represent meaning	15	p_LIT3
PS	LIT4	Concepts of print	Child shows awareness of the conventions of printed material and routines, and the physical organization of text and meaning	12	p_LIT1B
PS	LIT5	Phonological awareness	Child shows awareness of the sounds that make up including the segmentation of sounds in words, and recognition of word rhyming and alliteration	14	p_LIT2B
SA	LANG1	Comprehension of oral language	Child shows understanding of increasingly varied and complex oral language by responding appropriately (acting or communicating)	16	s_LANG1
SA	LANG2	Expression of oral language	Child uses oral language to communicate clearly and	17	s_LANG2

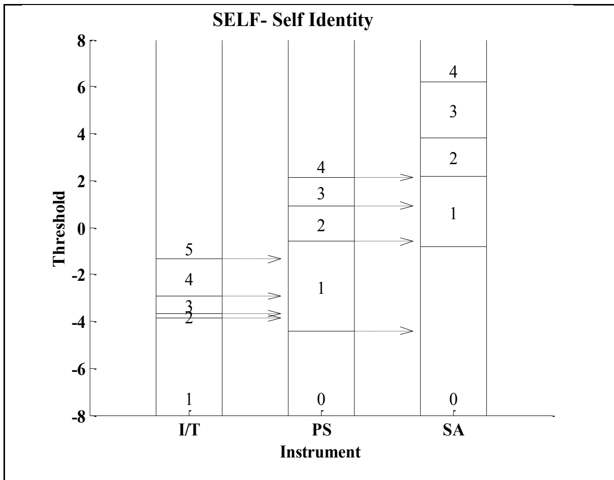
Age	Label	Measure name	Measure Definition	CS #	CS label
			effectively		
SA	LIT1	Interest in literacy	Child shows increasing interest in literacy activities (stories, books, writing, reading, maps)	18	s_LIT1
SA	LIT2	Decoding (word recognition and use)	Child shows increasing recognition and understanding of letters and words	19	s_S1LIT2
SA	LIT3	Writing	Child demonstrates increasing understanding and achievement of written communication skills	20	s_LIT3
SA	LIT4	Comprehension of written materials	Child shows increasing understanding of written materials and applies this knowledge in increasingly broad settings	21	s_S2LIT2
IT	COG1	Memory	Child shows awareness of past experiences and remembers information about people or things	1	i_COG1
IT	COG2	Cause and Effect	Child shows understanding of the connection between cause and effect	2	i_COG2
IT	COG3	Problem Solving	Child uses strategies to solve problems or make discoveries	3	i_COG3
IT	COG4	Symbolic Play	Child uses objects to represent other objects or ideas	4	i_COG4
IT	COG5	Curiosity	Child actively explores people and things, especially new ones	5	i_COG5
PS	LRN1	Curiosity and initiative	Child pursues knowledge or understanding of new materials or activities	6	p_LRN1
PS	LRN2	Engagement and persistence	Child persists in mastering and understanding an activity of his/her choice even in the face of difficulty or challenge	7	p_LRN2
PS	COG1	Memory and knowledge	Child stores, retrieves, and uses information about both familiar and unfamiliar events, past experiences, people, and things	8	p_COG1
PS	COG2	Cause and effect	Child logically relates events to what made them happen	9	p_COG2
PS	COG3	Engages in problem solving	Child reasons logically or uses strategies in order to reach a solution when confronted by a challenge	10	p_COG3
PS	COG4	Socio-dramatic play	dramatic play -Child learns to play with others using organized role-playing and symbolic play	11	p_COG4
SA	LRN1	Pursuit of understanding	Child uses strategies and resources to pursue knowledge about new materials, topics, or ideas	12	s_LRN1
SA	LRN2	Task persistence	Child persists in an activity of his/her choice even in the face of difficulty or challenge	13	s_LRN2
SA	COG1	Memory/knowledge	Child shows awareness of past experiences and remembers	14	s_COG1

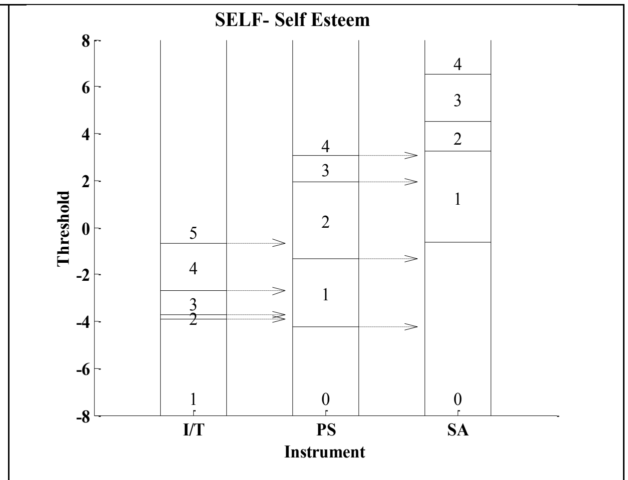
Age	Label	Measure name	Measure Definition	CS #	CS label
			information about people or things that can be used as a basis for making logical predictions about new situations or experiences		
SA	COG2	Cause and effect relationships	Child shows awareness and understanding of the connection between causes and effects in increasingly complex settings	15	s_COG2
SA	COG3	Problem-solving	Child uses logical and effective strategies to solve problems in increasingly broad settings	16	s_COG3
SA	COG4	Demonstrates inventiveness/inventive play	Child shows creativity and inventiveness in play and problem-solving in increasingly broad settings	17	s_COG4
IT	MATH1	Number	Child shows understanding of the concept of number or quantity	1	i_MATH1
IT	MATH2	Space and Size	Child shows understanding of how things move in space or fit in different spaces	2	i_MATH2
IT	MATH3	Time	Child shows understanding of the sequence of routine actions or events	4	i_MATH4
IT	MATH4	Classification and Matching	Child compares, matches, and categorizes different people or different things	3	i_MATH3
PS	MATH1	Number sense: Understands quantity and counting	Child recognizes, represents, names, and counts quantities	5	p_MATH1A
PS	MATH2	Number sense: Math operations	Child compares, combines, and separates simple quantities	6	p_MATH1B
PS	MATH3	Shapes	Child shows understanding of the characteristics of shapes and the placement of objects in space	7	p_MATH2
PS	MATH4	Time	Child understands and uses time-related vocabulary for routine actions, sequences and durations of events	11	p_MATH4
PS	MATH5	Classification	Child compares, matches, and groups objects according to some common characteristic	8	p_MATH3A
PS	MATH6	Measurement	Child shows understanding of measurable properties such as length, weight, and volume and how to quantify those properties	9	p_MATH3B
PS	MATH7	Patterning	Child recognizes and produces patterns of varying complexity	10	p_MATH3C
SA	MATH1	Number sense: Basic math skills (operations)	Child shows understanding of, and correctly performs, math operations (addition, subtraction, multiplication, and division)	12	s_MATH1
SA	MATH2	shapes	Child shows understanding of 2 and 3 dimensional shapes and manipulates them	13	s_S1MATH2

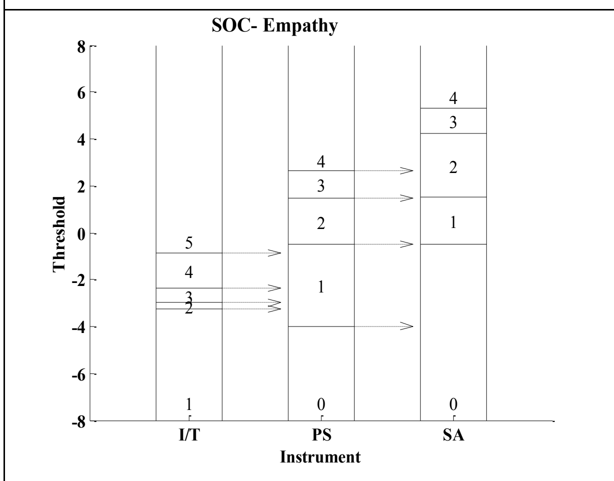
Age	Label	Measure name	Measure Definition	CS #	CS label
SA	MATH3	Time	Child shows understanding of concept of time and increasing ability to measure and tell time	15	s_S1MATH4
SA	MATH4	Measurement	Child shows understanding of measurement units, tools and techniques, and uses measurement to solve problems involving length, weight or volume	14	s_MATH3
IT	MOT1	Gross motor	Child moves different parts of body or whole body	1	i_MOT1A
IT	MOT2	Fine Motor	Child uses hands to reach or manipulate objects	3	i_MOT2A
IT	MOT3	Balance	Child maintains stability of body in various positions	2	i_MOT1B
IT	MOT4	Eye-Hand Coordination	Hand Coordination -Child uses eyes and hands together to perform an action or accomplish a task	4	i_MOT2B
PS	MOT1	Gross motor movement	Child refines the ability to move in a coordinated way using large muscles (e.g., arms and legs)	5	p_MOT1A
PS	MOT2	Fine motor skills	Child refines the ability to plan and coordinate use of grasp, release, strength, control of fingers and hands for functional and play activities	7	p_MOT2A
PS	MOT3	Balance	Child refines the ability to balance self in space	6	p_MOT1B
SA	MOT1	Movement and coordination (gross motor skills)	Child moves different parts of body or whole body with increasing coordination and integration	8	s_MOT1
SA	MOT2	Dexterity (fine motor skills)	Child demonstrates ability to manipulate small objects with his or her hands with increasing coordination and integration of movements	9	s_MOT2

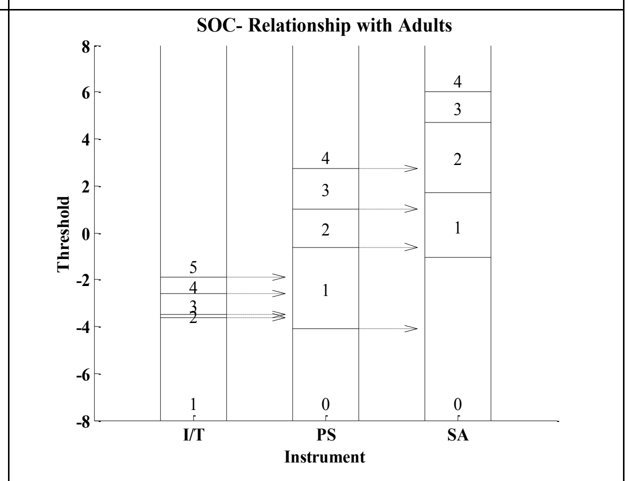
Appendix D – Graphs of Cross-Age Threshold Alignment of DRDP-R Measures

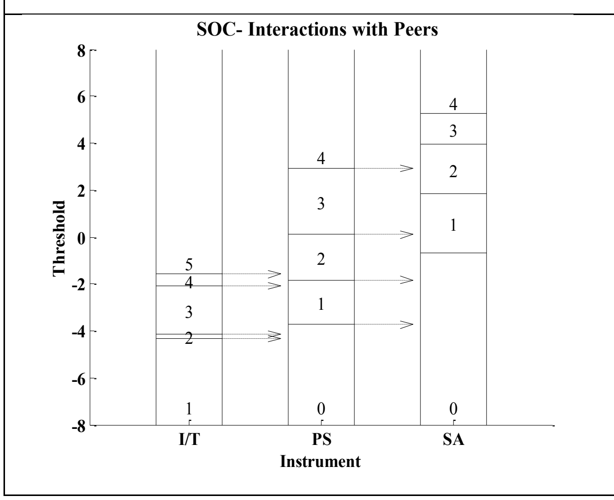
See the *Alignment of Developmental Level Difficulties* section in Chapter 6 for details on how to read and interpret the threshold alignment graphs.

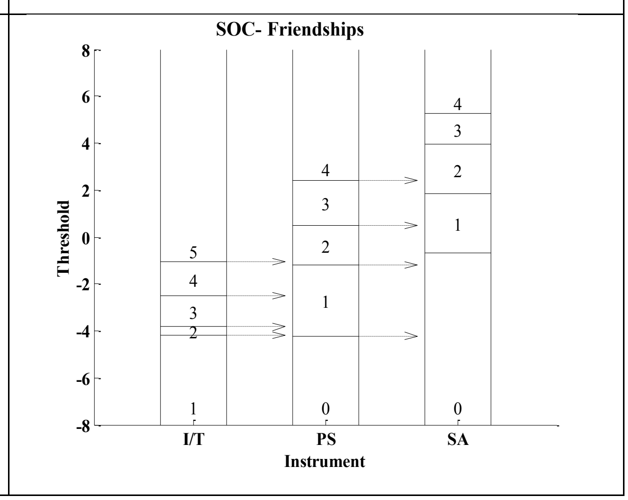


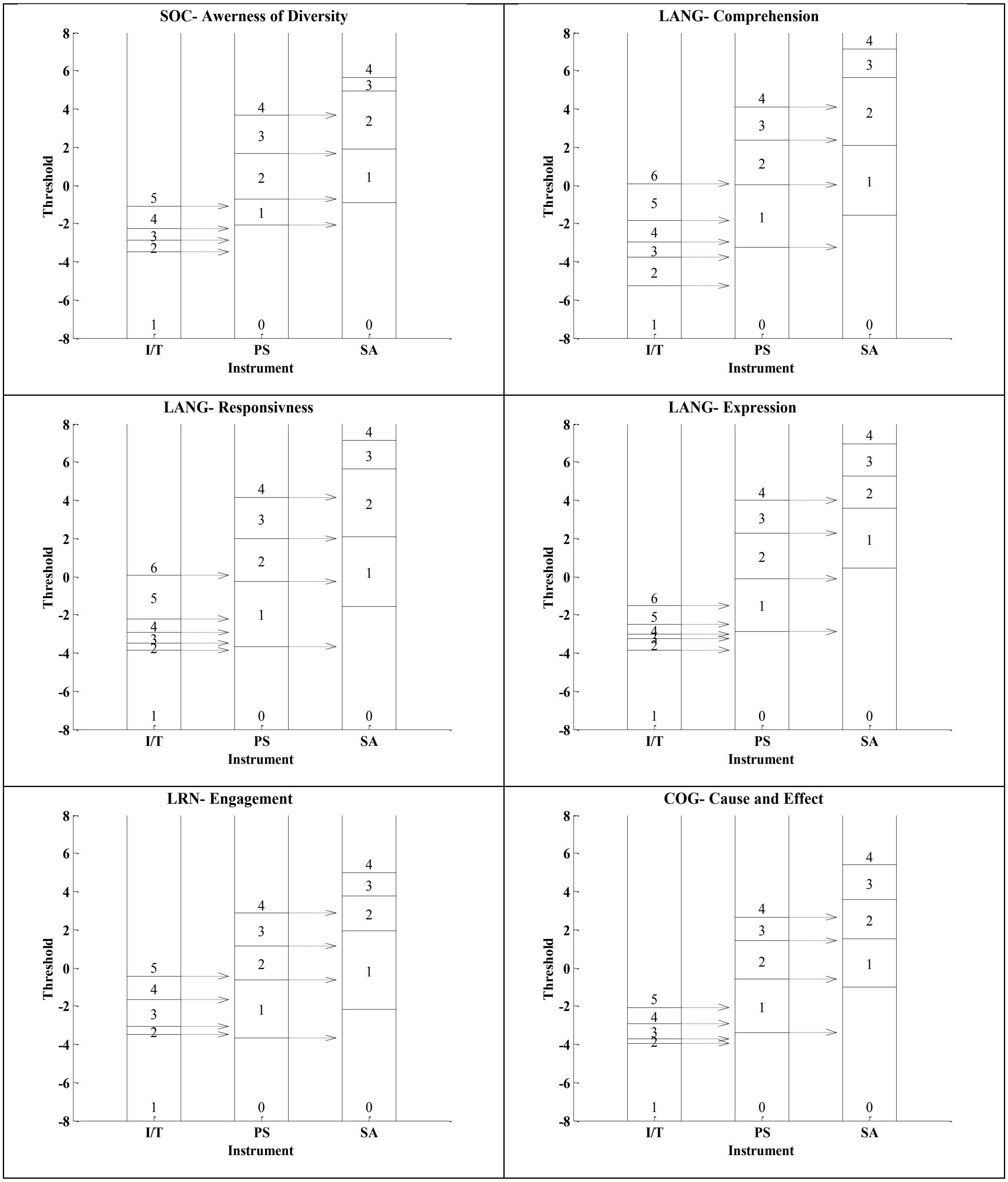


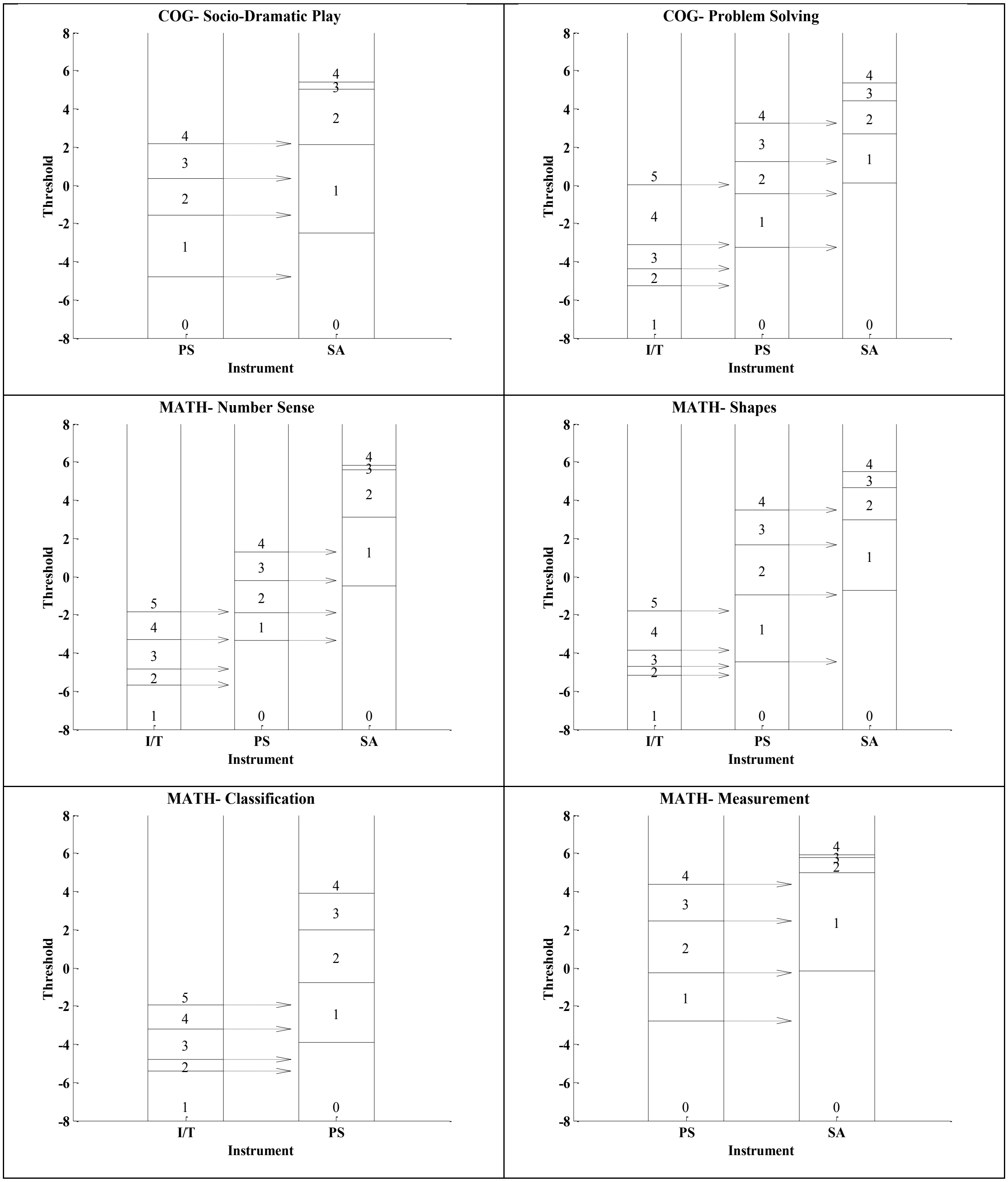


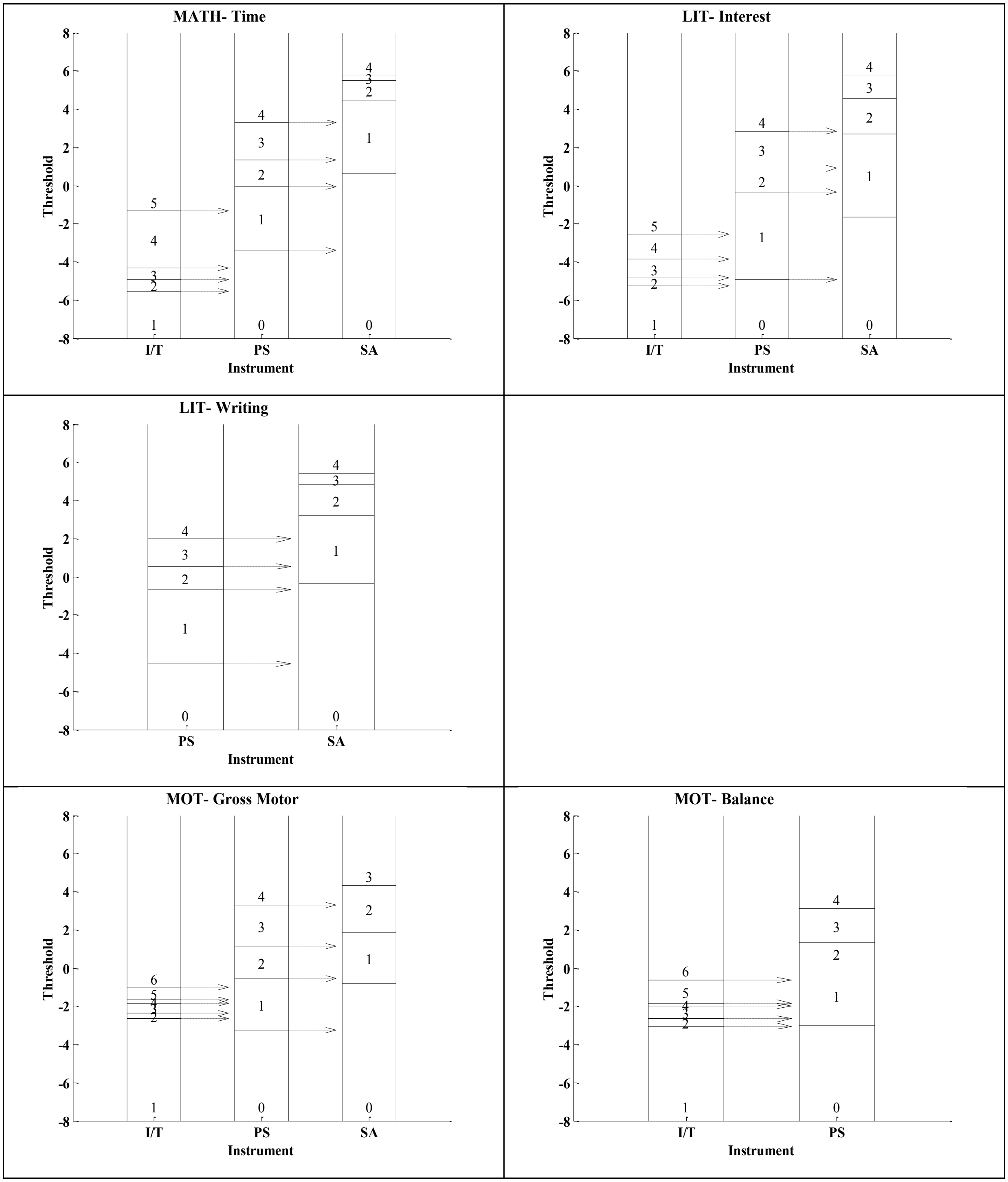


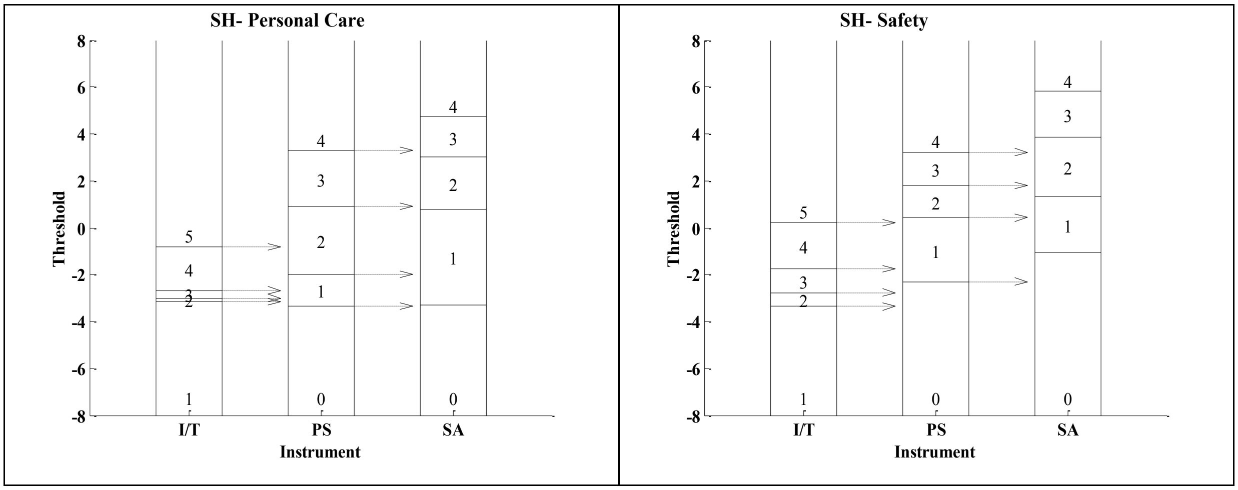





SOC- Awerness of Diversity
LANG- Comprehension
LANG- Responsivness
LANG- Expression
LRN- Engagement
COG- Cause and Effect
Threshold
Instrument
I/T
PS
SA


COG- Socio-Dramatic Play
Threshold
8
6
4
2
0
-2
-4
-6
-8
PS
SA
Instrument
COG- Problem Solving
Threshold
I/T
PS
SA
Instrument
MATH- Number Sense
Threshold
I/T
PS
SA
Instrument
MATH- Shapes
Threshold
I/T
PS
SA
Instrument
MATH- Classification
Threshold
I/T
PS
Instrument
MATH- Measurement
Threshold
PS
SA
Instrument


MATH- Time
Threshold
Instrument
I/T
PS
SA
LIT- Interest
Threshold
Instrument
I/T
PS
SA
LIT- Writing
Threshold
Instrument
PS
SA
MOT- Gross Motor
Threshold
Instrument
I/T
PS
SA
MOT- Balance
Threshold
Instrument
I/T
PS


SH- Personal Care
SH- Safety
Threshold
Instrument
I/T
PS
SA
8
6
4
2
0
-2
-4
-6
-8
1
0
0
2
3
4
5
1
2
3
4
1
2
3
4

Appendix E – Psychometric Information About DRDP-R Measures

The following tables show psychometric information about the DRDP-R measures, by indicator set. The tables contain the numbers and labels of measures from the three DRDP-Rs (labels for IT measures begin with "i", PS begins in "p" and SA begins with "s"). For each measure, the tables show the average measure difficulty (Delta), its corresponding standard error of measurement, the WMS fit statistics (here labeled as MNSQ), the corresponding t statistics, and the same information for each step difficulty (a parameter used to estimate difficulty of attaining each developmental level). The measure numbers and labels (the first two columns) refer to the 2005 Calibration Study measures. The Table in Appendix C provides the measure names and definitions that correspond to these measure labels.

SELF/SOC

#	measure	Delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*
1	iself1	-4.69	0.07	0.72	-5.00	-2.84	0.12	0.82	-1.20	-1.75	0.11	0.79	-4.50	1.33	0.10	0.91	-3.00				
2	iself2	-4.36	0.07	0.81	-3.20	-1.97	0.12	0.90	-0.50	-2.23	0.11	0.90	-2.10	0.95	0.10	0.93	-2.50				
3	iself3	-4.86	0.08	0.73	-4.20	-3.67	0.12	0.75	-2.30	-0.12	0.11	0.94	-0.60	-0.32	0.10	0.88	-3.00				
4	iself4	-3.38	0.06	0.87	-2.20	-2.40	0.11	0.92	-1.40	-0.83	0.10	0.94	-1.60	1.21	0.11	1.00	-0.10				
5	isoc1	-3.68	0.06	0.80	-3.50	-1.81	0.11	0.94	-0.50	-1.45	0.10	0.85	-3.80	1.10	0.10	0.95	-1.10				
6	isoc2a	-4.03	0.07	0.70	-5.60	-2.95	0.12	0.80	-2.40	-1.20	0.10	0.83	-4.10	1.06	0.10	0.92	-2.50				
7	isoc2b	-4.11	0.06	0.82	-2.90	-2.12	0.12	0.89	-1.10	-1.36	0.10	0.97	-0.70	1.66	0.11	0.96	-0.60				
8	isoc3a	-4.23	0.07	0.84	-2.60	-1.78	0.11	0.90	-0.80	-1.53	0.10	1.00	0.00	1.20	0.11	0.97	-0.60				
9	isoc3b	-4.19	0.07	1.05	0.70	-1.74	0.12	0.93	-0.40	-2.14	0.11	0.99	-0.20	1.34	0.10	0.98	-0.40				
10	pself1	-1.84	0.05	0.96	-0.80	-3.62	0.10	1.00	0.00	0.20	0.08	0.99	-0.30	1.42	0.10	0.99	-0.30				
11	pself2	-1.52	0.06	0.98	-0.40	-3.70	0.12	0.99	-0.10	-1.03	0.09	0.96	-1.30	2.27	0.10	0.97	-0.50				
12	psoc1	-1.18	0.05	0.89	-2.10	-3.77	0.11	0.99	-0.20	-0.21	0.08	0.92	-3.00	1.76	0.10	0.97	-0.60				
13	psoc2	-1.32	0.05	0.89	-2.20	-3.76	0.11	0.96	-0.80	-0.11	0.08	0.93	-2.00	1.22	0.09	0.94	-1.70				
14	psoc3a	-1.70	0.06	0.99	-0.10	-2.93	0.11	0.99	-0.10	-1.08	0.09	1.00	0.10	0.68	0.08	1.01	0.40				
15	psoc3b	-1.70	0.05	1.01	0.20	-3.53	0.11	1.00	0.00	-0.33	0.08	1.00	0.00	1.09	0.09	1.00	0.10				
16	psoc4	-0.65	0.06	1.02	0.50	-3.58	0.13	1.05	1.10	-0.72	0.09	0.97	-1.10	1.08	0.09	0.92	-2.10				
17	psoc5	-0.46	0.05	0.99	-0.30	-2.31	0.11	0.94	-1.20	-1.44	0.09	0.95	-1.80	1.02	0.09	0.96	-1.10				
18	sself1	1.14	0.05	0.99	-0.20	-4.88	0.15	1.09	1.40	-1.63	0.10	0.96	-1.00	-0.09	0.09	0.95	-1.50	2.56	0.11	0.95	-1.10
19	sself2	1.64	0.05	0.88	-2.50	-4.95	0.15	1.03	0.50	-1.19	0.10	0.95	-1.30	-0.13	0.09	0.90	-3.00	2.47	0.12	0.91	-1.60
20	ssoc1	1.63	0.05	0.92	-1.60	-4.65	0.15	1.01	0.20	-1.89	0.10	0.92	-2.50	-0.02	0.09	0.91	-2.60	2.73	0.13	0.94	-0.90
21	ssoc2	1.75	0.06	0.86	-3.00	-4.82	0.16	0.97	-0.50	-2.00	0.10	0.93	-2.30	0.18	0.09	0.88	-3.60	2.54	0.13	0.93	-1.20
22	ssoc3	1.68	0.05	0.93	-1.50	-4.36	0.16	1.02	0.40	-1.79	0.11	0.98	-0.50	0.03	0.10	0.93	-1.60	1.16	0.11	0.89	-2.30
23	ssoc4	2.19	0.06	0.83	-3.50	-5.68	0.19	1.05	1.00	-1.44	0.10	0.95	-1.70	0.38	0.10	0.88	-2.90	2.19	0.13	0.85	-2.60
24	ssoc5	2.20	0.06	0.94	-1.20	-4.93	0.16	1.02	0.40	-1.87	0.10	0.94	-2.00	0.32	0.10	0.92	-2.10	2.51	0.14	1.04	0.60

REG/SH

#	measure	delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*
1	ireg1	-3.63	0.07	0.88	-2.10	-3.29	0.12	0.81	-1.80	-1.46	0.11	0.92	-1.90	1.82	0.10	0.96	-1.10				
2	ireg2a	-4.20	0.07	0.77	-4.00	-3.53	0.11	0.81	-1.80	-0.72	0.10	0.88	-2.10	1.02	0.10	0.90	-3.30				
3	ireg2b	-3.12	0.07	0.86	-2.50	-3.62	0.13	0.90	-1.20	-1.13	0.11	0.90	-2.60	1.21	0.10	0.96	-1.20				
4	ireg3	-3.06	0.07	1.04	0.70	-2.87	0.12	0.87	-1.50	-1.57	0.11	0.94	-1.60	2.12	0.11	1.03	0.40				
5	ireg4	-3.65	0.07	0.90	-1.70	-3.49	0.12	0.82	-1.70	-1.27	0.11	0.99	-0.30	1.48	0.10	0.97	-0.90				
6	ish1	-4.15	0.07	0.83	-2.80	-3.44	0.11	0.92	-0.80	-0.02	0.10	0.90	-1.30	0.54	0.10	0.96	-1.20				
7	ish2	-3.45	0.07	0.83	-3.00	-2.28	0.11	0.85	-1.40	-1.39	0.10	0.90	-2.00	0.82	0.10	0.94	-1.80				
8	preg1	-0.46	0.05	1.02	0.40	-2.55	0.11	0.93	-1.40	-0.77	0.09	0.97	-1.10	0.95	0.09	0.97	-0.70				
9	preg2a	-0.65	0.05	0.87	-2.60	-3.16	0.11	0.92	-1.90	-0.40	0.08	0.95	-1.90	1.38	0.10	0.95	-1.20				
10	preg2b	-1.31	0.05	0.95	-0.90	-3.11	0.10	0.94	-0.90	-0.28	0.08	0.98	-0.50	0.73	0.08	0.97	-1.10				
11	psh1	-1.64	0.06	1.03	0.50	-2.54	0.11	0.98	-0.10	-1.78	0.10	1.01	0.30	1.16	0.08	1.00	0.00				
12	psh2	-0.66	0.05	0.98	-0.40	-2.75	0.10	0.99	-0.20	-0.09	0.08	1.01	0.20	0.89	0.09	0.98	-0.50				
13	psh3	-0.79	0.05	1.02	0.30	-3.08	0.11	1.03	0.60	-0.39	0.09	0.97	-0.70	0.56	0.08	0.97	-1.00				
14	sreg1	1.16	0.05	1.09	1.70	-3.05	0.12	1.05	1.10	-1.44	0.09	1.00	0.10	0.50	0.09	1.01	0.30	1.28	0.12	0.96	-0.80
15	sreg2	0.83	0.05	0.94	-1.10	-4.17	0.19	0.95	-0.70	-2.04	0.11	0.97	-1.10	0.07	0.09	0.97	-0.90	1.83	0.11	0.96	-0.70
16	ssh1	-0.47	0.06	1.00	0.00	-3.97	0.11	1.00	0.00	-0.43	0.09	0.98	-0.70	1.09	0.08	0.99	-0.40				
17	ssh2	0.21	0.06	0.92	-1.70	-3.35	0.13	0.97	-0.40	-1.18	0.09	0.94	-2.20	1.04	0.09	0.91	-3.00				
18	ssh3	0.94	0.06	0.99	-0.20	-3.83	0.14	0.96	-1.10	-0.89	0.09	0.97	-1.00	1.11	0.10	0.96	-1.00				
19	ssh4	-0.02	0.06	1.52	8.20	-3.28	0.14	1.19	2.50	-1.06	0.10	1.12	2.70	-0.04	0.08	1.27	6.90				

LANG/LIT

#	measure	delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*	Step5	*SE*	MNSQ	*t*
1	ilang1a	-4.33	0.06	0.77	-4.00	-3.16	0.13	0.76	-2.40	-1.48	0.11	0.86	-2.10	-0.04	0.10	1.00	-0.10	1.38	0.10	0.98	-0.60				
2	ilang1b	-4.55	0.06	0.70	-5.10	-2.96	0.12	0.74	-2.60	-1.03	0.11	0.91	-1.00	-0.43	0.10	0.95	-1.00	1.16	0.10	0.94	-1.80				
3	ilang2a	-4.64	0.06	0.74	-4.20	-3.62	0.12	0.71	-3.40	0.19	0.10	1.01	0.10	-0.34	0.10	0.91	-1.60	1.54	0.11	0.94	-1.20				
4	ilang2b	-4.36	0.06	0.74	-4.60	-3.79	0.13	0.80	-2.00	-1.17	0.11	0.86	-2.10	-0.06	0.10	0.89	-3.00	2.03	0.11	0.97	-0.70				
5	ilit1	-4.54	0.07	0.85	-2.30	-1.65	0.11	0.88	-1.00	-1.26	0.10	1.00	0.10	0.66	0.10	0.94	-1.50								
6	ilit2	-4.28	0.07	0.78	-3.90	-2.45	0.12	0.82	-1.60	-1.62	0.11	0.88	-2.50	0.88	0.10	0.96	-1.20								
7	plang1a	-1.23	0.05	0.82	-3.80	-3.45	0.11	0.93	-1.40	-0.54	0.08	0.93	-2.40	1.42	0.09	0.94	-1.60								
8	plang1b	-1.46	0.06	0.92	-1.60	-3.65	0.11	0.98	-0.40	-0.54	0.09	0.96	-1.20	1.24	0.09	0.96	-1.40								
9	plang2a	-1.22	0.05	0.86	-2.80	-3.12	0.11	0.93	-1.40	-0.70	0.09	0.92	-2.90	1.34	0.09	0.93	-1.90								
10	plang2b	-1.24	0.05	0.93	-1.30	-2.94	0.10	0.99	-0.20	-0.38	0.08	0.99	-0.40	0.89	0.09	0.95	-1.30								
11	plit1a	-1.03	0.05	0.92	-1.50	-4.17	0.12	0.90	-2.20	0.39	0.09	0.96	-0.80	0.93	0.10	0.98	-0.60								
12	plit1b	-0.82	0.06	0.78	-4.50	-4.18	0.12	0.83	-4.00	-0.19	0.09	0.90	-3.60	1.68	0.10	0.94	-1.10								
13	plit2a	-0.70	0.05	0.93	-1.40	-3.28	0.11	0.96	-1.00	-0.49	0.09	0.97	-0.90	1.51	0.11	0.97	-0.50								
14	plit2b	-0.16	0.06	0.89	-2.20	-4.25	0.13	0.92	-2.00	0.18	0.09	0.96	-0.80	0.72	0.10	0.88	-2.70								
15	plit3	-1.29	0.05	1.14	2.50	-3.50	0.10	1.08	1.70	0.33	0.08	1.01	0.20	1.02	0.10	1.08	1.90								
16	slang1	1.88	0.05	0.76	-4.90	-6.24	0.20	0.95	-0.70	-2.37	0.11	0.95	-1.80	0.50	0.10	0.93	-1.20	0.79	0.11	0.89	-2.50	2.89	0.14	0.90	-1.70
17	slang2	2.09	0.05	0.76	-4.80	-4.79	0.14	0.98	-0.40	-1.72	0.10	0.94	-2.00	0.11	0.10	0.93	-1.40	1.06	0.11	0.91	-1.40	1.72	0.13	0.89	-2.30
18	slit1	2.31	0.05	0.79	-4.20	-5.29	0.16	0.96	-0.80	-1.31	0.10	0.96	-1.10	-0.15	0.10	0.93	-1.30	1.15	0.12	0.90	-1.40	1.70	0.13	0.86	-2.80
19	s1lit2	0.82	0.06	1.04	0.80	-3.22	0.15	1.08	1.50	-0.61	0.10	1.03	1.00	0.87	0.11	0.95	-1.20								
20	slit3	2.27	0.05	0.87	-2.60	-4.91	0.15	1.08	1.60	-1.69	0.10	0.98	-0.50	-0.05	0.10	0.94	-1.20	0.62	0.11	0.85	-3.10	2.59	0.14	0.90	-1.50
21	s2lit2	2.97	0.09	0.81	-2.00	-1.51	0.19	0.98	0.00	-0.69	0.18	0.99	0.10	-2.25	0.17	0.89	-1.90	1.61	0.15	0.92	-1.90				

LRN/COG

#	measure	delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*
1	icog1	-3.71	0.07	0.82	-3.10	-2.82	0.12	0.80	-2.30	-1.41	0.11	0.89	-2.60	0.97	0.10	0.99	-0.20				
2	icog2	-4.11	0.07	0.80	-3.50	-2.30	0.11	0.78	-2.40	-0.99	0.10	0.94	-1.10	0.80	0.10	0.96	-1.10				
3	icog3	-4.05	0.08	0.84	-2.80	-3.29	0.14	0.71	-2.50	-1.95	0.12	0.88	-2.80	0.98	0.10	0.94	-1.70				
4	icog4	-3.67	0.07	0.88	-2.20	-3.16	0.12	0.80	-2.70	-1.15	0.10	0.92	-2.10	1.51	0.10	1.02	0.60				
5	icog5	-4.20	0.07	0.82	-3.10	-2.56	0.11	0.89	-1.10	-1.20	0.10	0.92	-1.60	0.76	0.10	0.95	-1.80				
6	plrn1	-1.64	0.05	0.94	-1.10	-4.60	0.12	0.94	-1.10	0.08	0.08	0.97	-0.90	1.55	0.09	0.97	-0.90				
7	plrn2	-1.41	0.05	1.08	1.60	-3.72	0.12	1.07	1.40	-0.43	0.09	0.99	-0.40	1.29	0.09	0.99	-0.40				
8	pcog1	-1.22	0.05	0.93	-1.30	-3.10	0.10	0.96	-0.80	-0.02	0.08	0.97	-0.80	0.83	0.09	0.99	-0.20				
9	pcog2	-0.98	0.05	0.91	-1.70	-3.11	0.11	0.93	-1.60	-0.47	0.08	0.95	-1.60	1.51	0.10	0.96	-0.80				
10	pcog3	-0.81	0.05	0.92	-1.60	-3.19	0.11	0.91	-2.20	-0.37	0.09	0.94	-1.70	0.88	0.09	0.97	-1.00				
11	pcog4	-1.92	0.05	1.06	1.20	-3.67	0.11	0.99	-0.20	-0.35	0.09	1.00	0.10	1.30	0.09	0.99	-0.20				
12	slrn1	1.67	0.05	0.85	-3.10	-5.44	0.17	1.04	0.90	-1.35	0.10	0.94	-1.80	0.11	0.10	0.86	-3.80	2.44	0.13	0.90	-1.80
13	slrn2	1.39	0.06	0.91	-1.90	-5.78	0.18	1.04	0.70	-1.40	0.10	0.95	-1.40	0.22	0.10	0.92	-2.40	2.44	0.12	0.91	-1.60
14	scog1	1.28	0.05	0.92	-1.70	-5.48	0.18	0.92	-1.50	-1.38	0.10	0.93	-2.00	0.17	0.10	0.96	-1.10	1.84	0.11	0.94	-1.20
15	scog2	1.20	0.06	0.91	-1.80	-5.17	0.18	1.00	-0.10	-1.87	0.11	0.93	-2.10	-0.17	0.10	0.94	-2.00	2.23	0.11	0.91	-1.80
16	scog3	2.19	0.05	0.84	-3.20	-4.46	0.15	1.01	0.10	-1.69	0.10	0.96	-1.30	-0.11	0.11	0.88	-3.10	1.56	0.13	0.87	-2.20
17	scog4	1.68	0.06	0.83	-3.60	-5.62	0.19	0.98	-0.40	-1.84	0.10	0.89	-3.80	0.24	0.10	0.85	-4.30	2.37	0.13	0.94	-1.00

MATH

#	measure	delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*	Step5	*SE*	MNSQ	*t*
1	imath1	-3.87	0.06	0.89	-1.90	-3.14	0.13	0.85	-2.20	-1.05	0.10	0.94	-1.60	1.68	0.11	1.00	0.00								
2	imath2	-4.57	0.07	0.90	-1.60	-3.02	0.12	0.73	-3.00	-1.11	0.10	0.97	-0.60	0.78	0.10	0.99	-0.30								
3	imath3	-3.91	0.06	0.84	-2.90	-3.34	0.12	0.80	-3.20	-0.68	0.10	0.94	-1.60	1.50	0.11	1.01	0.20								
4	imath4	-4.13	0.07	0.83	-2.90	-3.08	0.13	0.83	-2.30	-0.71	0.11	0.97	-0.50	0.23	0.10	1.00	-0.10								
5	pmath1a	-1.80	0.05	0.97	-0.50	-1.67	0.09	1.02	0.40	-0.77	0.09	1.01	0.10	0.80	0.09	0.96	-0.80								
6	pmath1b	-0.55	0.05	0.79	-4.40	-2.78	0.10	0.89	-3.10	-0.49	0.09	0.89	-3.60	1.54	0.11	0.94	-0.90								
7	pmath2	-1.02	0.06	0.82	-3.60	-3.76	0.12	0.91	-2.10	-0.69	0.09	0.93	-2.60	1.59	0.10	0.94	-1.30								
8	pmath3a	-0.68	0.06	0.82	-3.80	-3.58	0.12	0.87	-3.00	-0.89	0.09	0.92	-3.10	1.53	0.10	0.93	-1.60								
9	pmath3b	-0.10	0.06	0.80	-4.30	-3.06	0.11	0.89	-3.10	-1.10	0.09	0.91	-3.40	1.33	0.11	0.96	-0.90								
10	pmath3c	-0.28	0.05	1.00	0.10	-2.20	0.10	0.96	-0.90	-0.77	0.09	0.98	-0.50	0.94	0.10	0.96	-0.90								
11	pmath4	-0.70	0.05	0.91	-1.70	-3.14	0.10	0.92	-2.00	0.01	0.09	0.99	-0.20	0.70	0.10	0.95	-1.30								
12	smath1	1.71	0.05	0.85	-2.90	-4.75	0.13	1.11	2.00	-1.92	0.10	0.96	-1.20	0.77	0.09	0.95	-0.80	1.26	0.11	0.93	-0.90	2.00	0.13	0.98	-0.30
13	s1math2	0.52	0.06	0.98	-0.30	-3.55	0.14	1.06	1.10	-0.42	0.10	0.97	-0.90	1.14	0.11	0.95	-1.30								
14	smath3	2.62	0.05	0.82	-3.40	-4.89	0.14	0.95	-1.10	-0.73	0.10	0.96	-0.70	0.07	0.11	0.97	-0.40	0.25	0.11	0.89	-2.50	2.87	0.18	0.96	-0.30
15	s1math4	1.77	0.06	0.91	-1.40	-3.48	0.14	0.96	-1.20	-0.09	0.11	0.98	-0.50	1.28	0.15	0.98	-0.30								

MOT

#	measure	delta	*SE*	MNSQ	*t*	Step1	*SE*	MNSQ	*t*	Step2	*SE*	MNSQ	*t*	Step3	*SE*	MNSQ	*t*	Step4	*SE*	MNSQ	*t*
1	imot1a	-4.34	0.07	0.75	-3.40	-1.98	0.11	0.81	-1.40	-0.28	0.11	0.88	-0.60	-0.61	0.11	0.88	-1.30	0.43	0.10	0.98	-0.60
2	imot1b	-4.49	0.07	0.77	-3.00	-2.08	0.11	0.79	-1.30	-0.49	0.11	0.88	-0.70	-0.27	0.10	0.89	-0.80	-0.12	0.10	0.93	-1.80
3	imot2a	-3.42	0.07	0.95	-0.70	-2.65	0.13	0.91	-0.70	-1.26	0.12	0.97	-0.20	-0.98	0.11	0.90	-2.00	1.07	0.10	1.03	0.90
4	imot2b	-3.71	0.06	0.89	-1.80	-2.30	0.12	0.92	-0.40	-1.83	0.11	0.88	-1.40	-0.33	0.10	0.96	-0.80	1.35	0.10	1.01	0.20
5	pmot1a	-2.20	0.06	0.97	-0.60	-3.58	0.10	0.93	-0.80	-0.45	0.09	0.98	-0.40	1.01	0.08	1.00	-0.20				
6	pmot1b	-1.92	0.05	0.95	-0.90	-3.54	0.10	0.98	-0.40	0.24	0.08	1.00	-0.10	0.77	0.09	1.00	0.10				
7	pmot2a	-1.86	0.05	1.12	2.10	-2.71	0.10	1.03	0.50	-0.10	0.08	1.03	0.60	0.67	0.09	1.02	0.70				
8	smot1	-0.55	0.05	0.90	-1.90	-5.25	0.14	1.16	1.60	-1.92	0.11	0.99	-0.40	0.36	0.09	0.92	-2.50	3.85	0.14	0.90	-1.10
9	smot2	-0.42	0.05	0.93	-1.20	-6.09	0.15	1.26	3.10	-1.26	0.10	1.00	0.10	0.68	0.09	0.94	-2.10	3.56	0.13	0.92	-1.00

Appendix F – Test Information Plots

The reliability of an assessment can also be evaluated through individual items' information, and overall using test information functions. Test information provides an index of the precision of measurement (i.e., the extent to which measurement is without error), at any given point on the ability scale. For example, higher information values between -2 to 0 mean that at this range of the scale, ability estimates derived based on this assessment are the most accurate. Technically, information is defined as the reciprocal of the squared measurement error associated with each ability estimate.

For the purpose of the DRDP-R we wish to see that each age-specific instrument is most sensitive at an ability range appropriate for that age group, with some overlap between instruments. This is because we expect only one age-specific instrument to be administered to a child according to his or her age. Figures F.1-F.6 show test information plots. Each figure contains three plots per indicator, one for each instrument. Higher information values mean that ability estimates at a given logit level are more accurate. Note that the scale of these distributions may differ for each age group.

SELF/SOC – Figure F.1 shows that for IT, the ability estimates between -7 and -1 appears the most accurate, and that the amount of information ranges approximately from 4 to 5.5. For PS, there seems a more distinct peak around -0.5. The ability estimates between -2 and +2 appear the most accurate, and the amount of information ranges between 4 and 5. For SA, the ability estimates between 0 and 5 look the most accurate, and the amount of information is between 3 and 3.5.

REG/SH – Figure F.2 shows that for IT, the ability estimates between -6 and -1 appear the most accurate, and that the amount of information ranges approximately from 3 to 4. For PS, there seems a more distinct peak around 0. The ability estimates between -2 and +2 appear the most accurate, and the amount of information ranges between 3 and 4. For SA, the ability estimates between -1 and 3 look the most accurate, and the amount of information is on average 3.

LANG/LIT – Figure F.3 shows that for IT, the ability estimates between -7 and -2 appear the most accurate, and that the amount of information ranges approximately from 3 to 5. For PS, there seems a more distinct peak around 0. The ability estimates between -2 and +2 appear the most accurate, and the amount of information ranges between 4 and 6. For SA, there seems a more distinct peak around 3. The ability estimates between 1 and 5 look the most accurate, and the amount of information is between 3.5 and 5.

LRN/COG – Figure F.4 shows that for IT, the ability estimates between -6 and -2 appear the most accurate, and that the amount of information ranges approximately from 2.5 to 3. For PS, there seems a more distinct peak around 0. The ability estimates between -2 and +2 appear the most accurate, and the amount of information ranges between 3 and 4. For SA, the ability estimates between 0 and 5 look the most accurate, and the amount of information is between 2.5 and 3.5.

MATH – Figure F.5 shows that for IT, the information plot looks a little flatter than the other plots. The ability estimates between -5 and -2 appear the most accurate, and the amount of information is on average 2. For PS, there seems a more distinct peak around 0. The ability estimates between -2 and +2 appear the most accurate, and the amount of information ranges between 3.5 and 4.5. For SA, there seems a more distinct peak around 3. The ability estimates between 2 and 4 look the most accurate, and the amount of information is between 2.5 and 3.5.

MOT – Figure F.6 shows that unlike the other indicators, there are clear peaks for each age-group plot. For IT, the peak appears around -5, and the amount of information is about 4.5. For PS, there seems a peak around -1; the amount of information is about 2. For SA, the peak seems around 3, and the amount of information is about 1.4. As they are close to both ends, the estimates are less accurate.

Overall, the information functions suggest that, as desired, each age-specific DRDP-R is especially sensitive for the age range it was designed for.

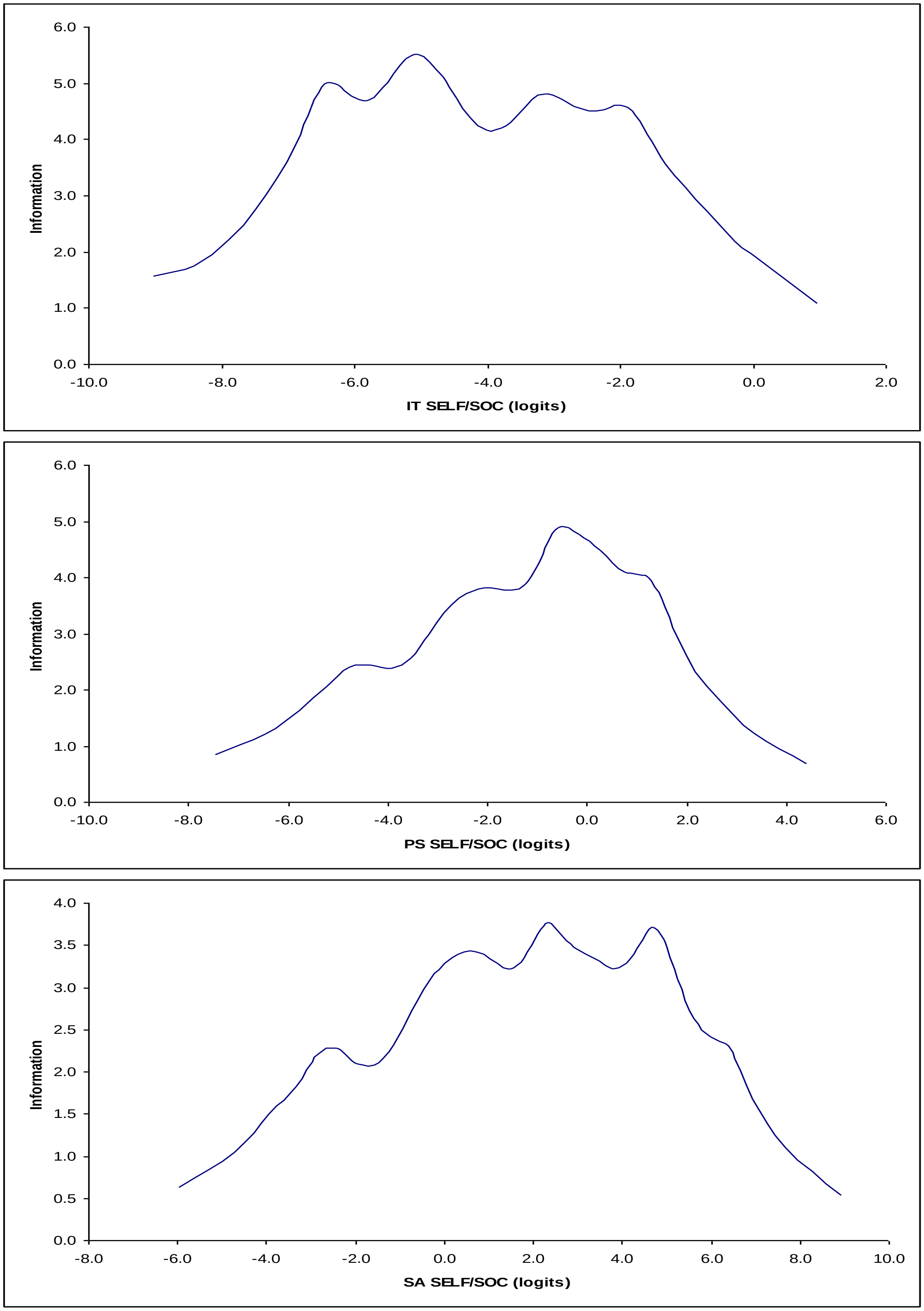


Figure F.1—Information for Self Concept and Social Interpersonal Skills.

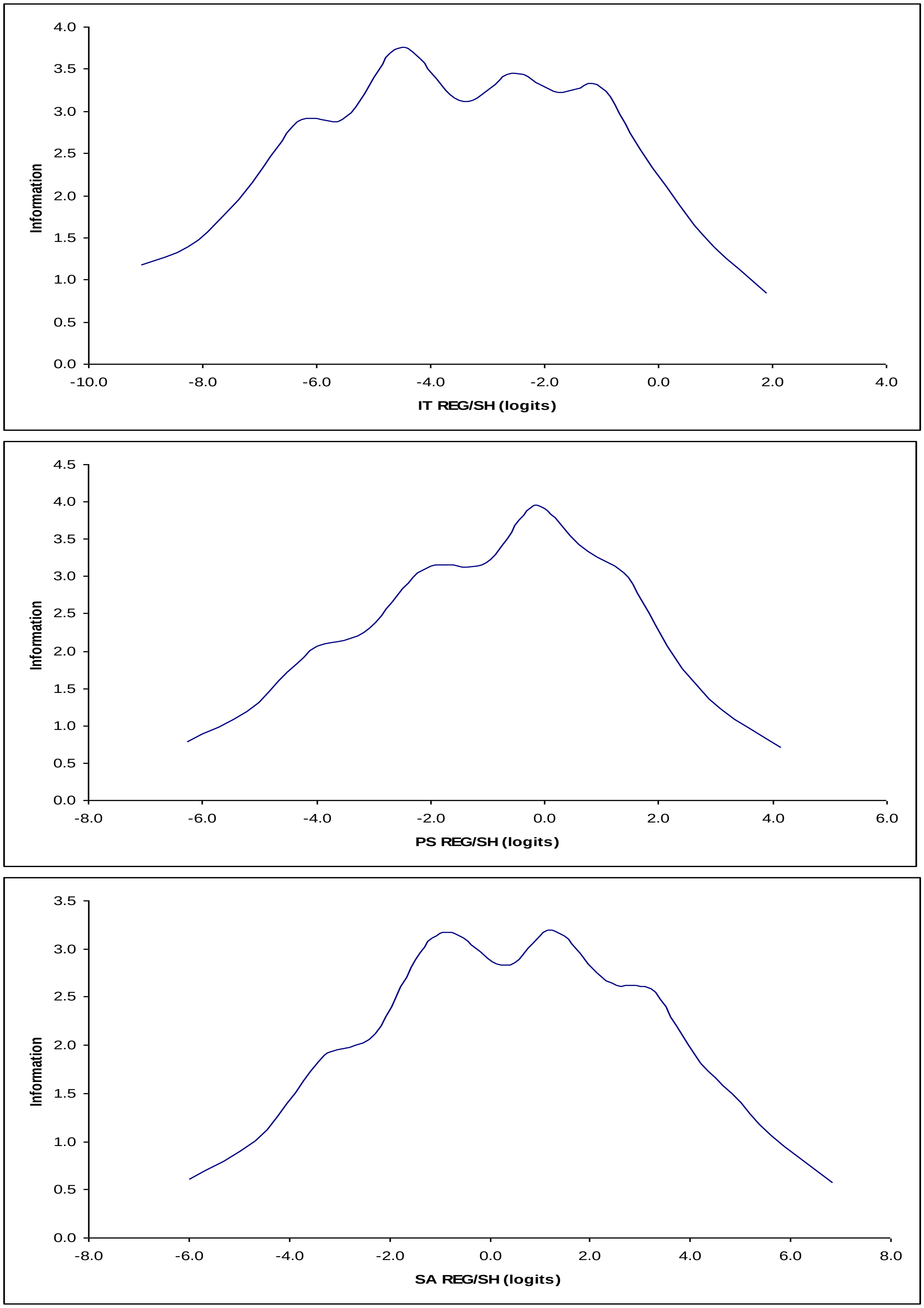


Figure F.2—Information for Self Regulation and Safety and Health.

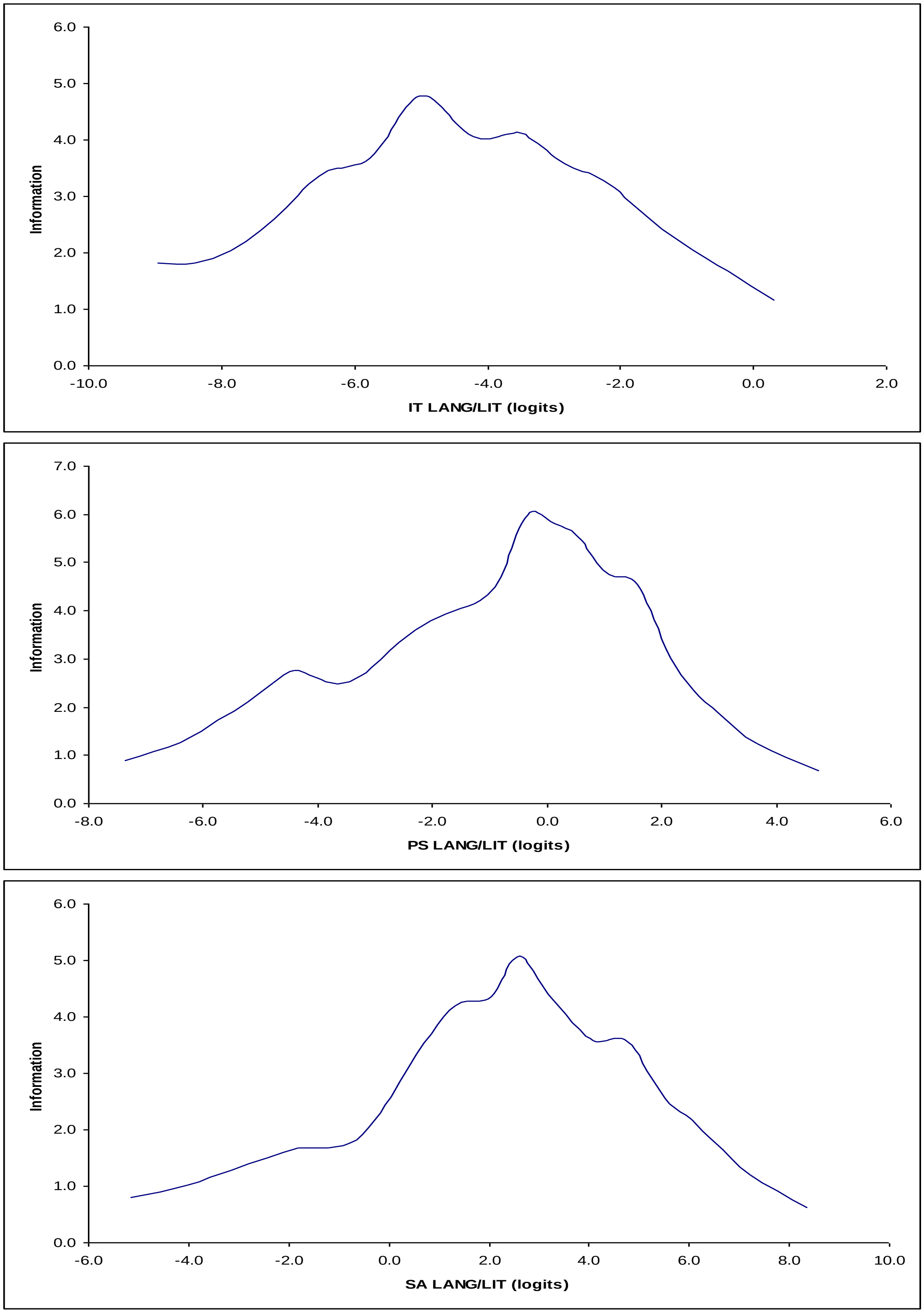


Figure F.3—Information for Language and Literacy.

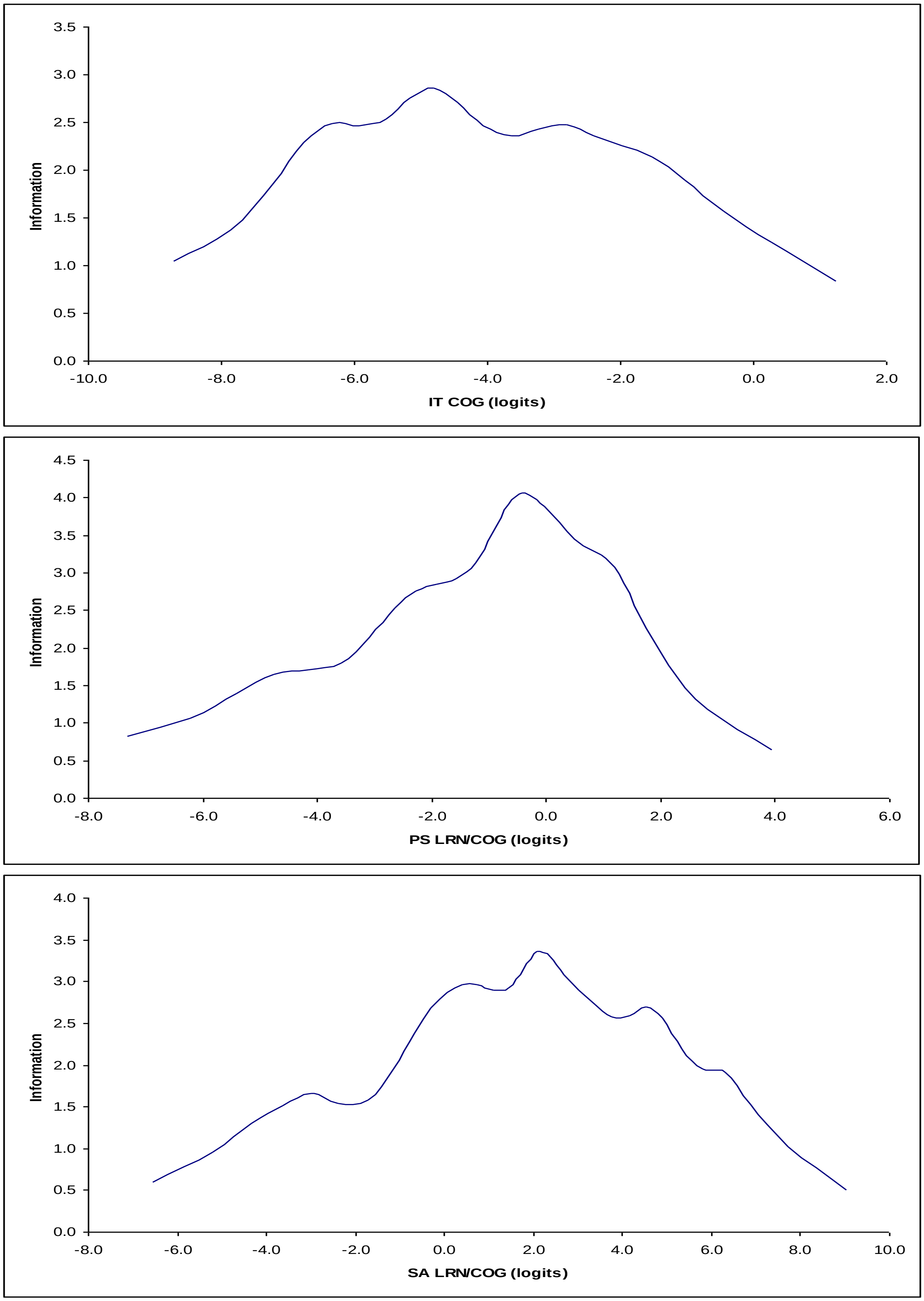


Figure F.4—Information for Learning and Cognitive Competence.

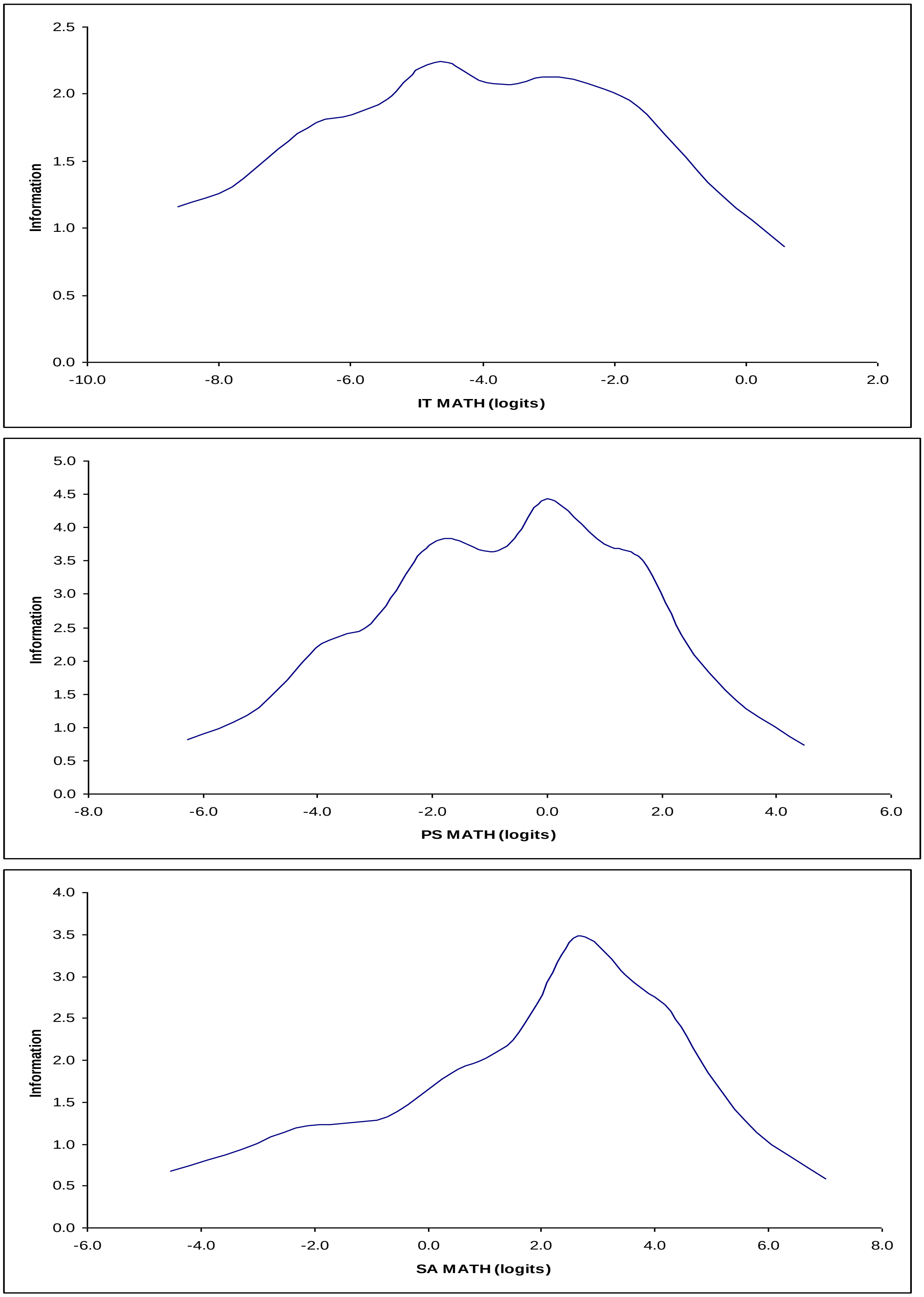


Figure F.5—Information for Math.

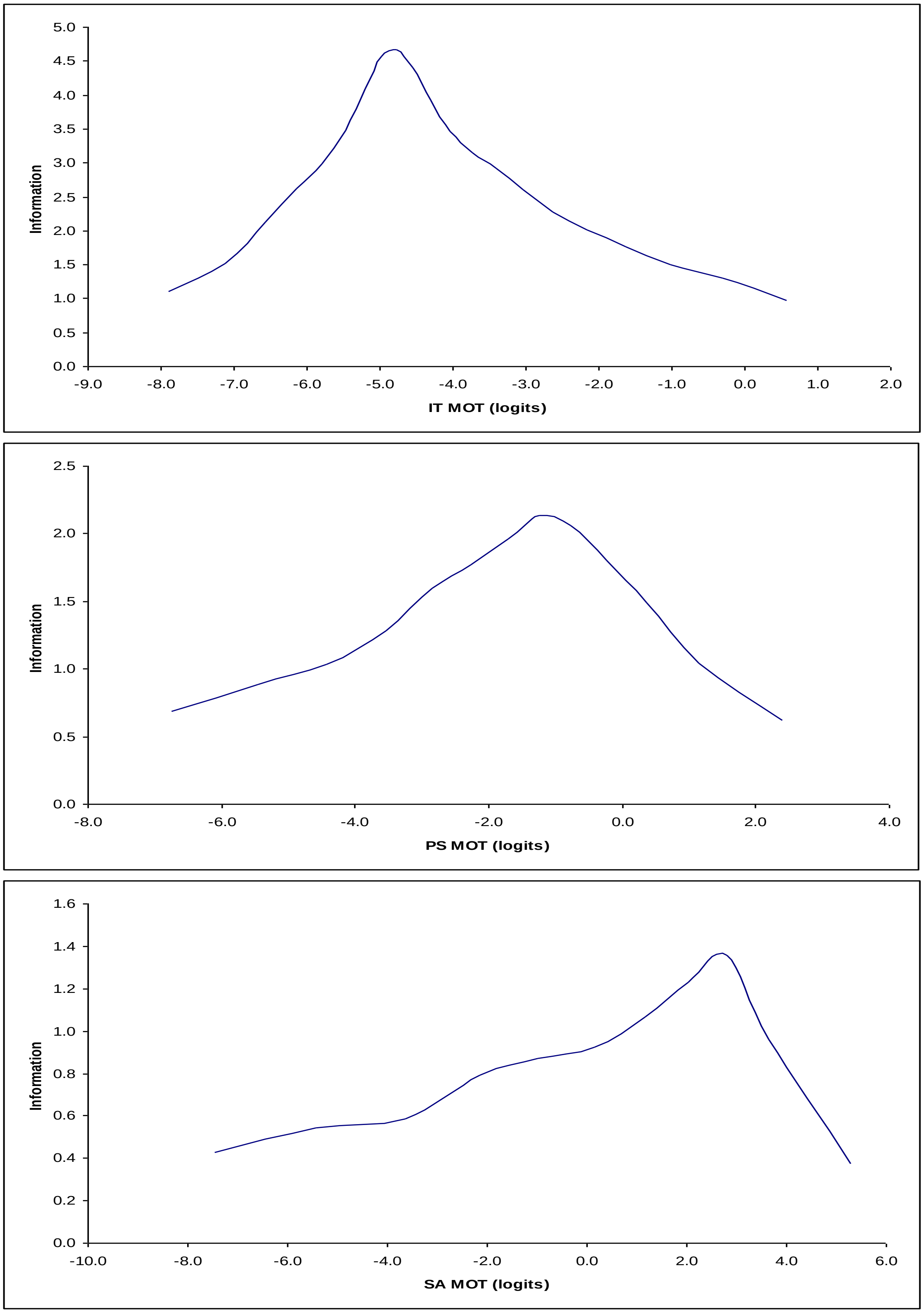


Figure F.6—Information for Motor Skills.

Appendix G – Disattenuated Correlations Among the Six DRDP-R Domains

Table G.1—Disattenuated Correlations Among the Six DRDP-R Domains

	LANG/LIT	MATH	LRN/COG	SELF/SOC	REG/SH	MOT
LANG/ LIT	1					
MATH	.97	1				
LRN/COG	.96	.96	1			
SELF/SOC	.92	.91	.95	1		
REG/SH	.86	.87	.89	.92	1	
MOT	.86	.85	.85	.81	.80	1

Appendix H – Analysis Codes (ConQuest Codes)

```
reset;
Title IT-PS-SA SELF/SOC PCM augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 21-29, 61-68, 101-107;
labels << augc3sa2ss.nam;
codes 0,1,2,3,4,5;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-9);
score (0,1,2,3,4,5)  (0,1,2,3,4,5)  ! items(1-24);
model item + item*step;
export log    >> augc3sa2ss.l,
       par       >>  augc3sa2ss.p,
       cov       >>  augc3sa2ss.c,
       reg_coeff >>  augc3sa2ss.r;
/*import init_par  <<  augc3sa2ss.p,
       init_cov  <<  augc3sa2ss.c,
       init_reg  <<  augc3sa2ss.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2ss.shw;
show cases ! estimates=wle >> augc3sa2ss.wle;
show cases ! estimates=eap >> augc3sa2ss.eap;
show parameters !tables=7 >> augc3sa2ss.thr;
equivalence wle >> augc3sa2ss.eqv;
itanal ! estimates=wle >> augc3sa2ss.itn;

reset;
Title  IT-PS-SA REG/SH PCM augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 30-34, 54-55, 69-71, 97-99, 108-109, 127-130;
labels << augc3sa2rs.nam;
codes 0,1,2,3,4,5;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-7);
score (0,1,2,3,4,5)    (0,1,2,3,4,5)    ! items(1-19);
model item + item*step;
export log       >>  augc3sa2rs.l,
       par       >>  augc3sa2rs.p,
       cov       >>  augc3sa2rs.c,
       reg_coeff >>  augc3sa2rs.r;
/*import init_par  <<  augc3sa2rs.p,
       init_cov  <<  augc3sa2rs.c,
       init_reg  <<  augc3sa2rs.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2rs.shw;
show cases ! estimates=wle >> augc3sa2rs.wle;
show cases ! estimates=eap >> augc3sa2rs.eap;
```

```
show parameters !tables=7 >> augc3sa2rs.thr;
equivalence wle >> augc3sa2rs.eqv;
itanal ! estimates=wle >> augc3sa2rs.itn;

reset;
Title  IT-PS-SA LANG/LIT PCM augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 35-38, 48-49, 72-75, 89-93, 110-111, 122-124,131;
labels << augc3sa2ll.nam;
codes 0,1,2,3,4,5,6;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-6,21);
score (0,1,2,3,4,5,6)  (0,1,2,3,4,5,6) ! items (1-21);
model item + item*step;
export log       >>  augc3sa2ll.l,
       par       >>  augc3sa2ll.p,
       cov       >>  augc3sa2ll.c,
       reg_coeff >>  augc3sa2ll.r;
/*import init_par  <<  augc3sa2ll.p,
       init_cov  <<  augc3sa2ll.c,
       init_reg  <<  augc3sa2ll.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2ll.shw;
show cases ! estimates=wle >> augc3sa2ll.wle;
show cases ! estimates=eap >> augc3sa2ll.eap;
show parameters !tables=7 >> augc3sa2ll.thr;
equivalence wle >> augc3sa2ll.eqv;
itanal ! estimates=wle >> augc3sa2ll.itn;

reset;
Title  IT-PS-SA LRN/COG augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 39-43, 76-81, 112-117;
labels << augc3sa2lc.nam;
codes 0,1,2,3,4,5;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-5);
score (0,1,2,3,4,5)    (0,1,2,3,4,5)   ! items (1-17);
model item + item*step;
export log       >>  augc3sa2lc.l,
       par       >>  augc3sa2lc.p,
       cov       >>  augc3sa2lc.c,
       reg_coeff >>  augc3sa2lc.r;
/*import init_par  <<  augc3sa2lc.p,
       init_cov  <<  augc3sa2lc.c,
       init_reg  <<  augc3sa2lc.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2lc.shw;
show cases ! estimates=wle >> augc3sa2lc.wle;
show cases ! estimates=eap >> augc3sa2lc.eap;
```

```
show parameters !tables=7 >> augc3sa2lc.thr;
equivalence wle >> augc3sa2lc.eqv;
itanal ! estimates=wle >> augc3sa2lc.itn;

reset;
Title  IT-PS-SA MATH PCM augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 44-47, 82-88, 118-121;
labels << augc3sa2ma.nam;
codes 0,1,2,3,4,5,6;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-4);
score (0,1,2,3,4,5,6)    (0,1,2,3,4,5,6)   ! items (1-15);
model item + item*step;
export log       >>  augc3sa2ma.l,
       par       >>  augc3sa2ma.p,
       cov       >>  augc3sa2ma.c,
       reg_coeff >>  augc3sa2ma.r;
/*import init_par  <<  augc3sa2ma.p,
       init_cov  <<  augc3sa2ma.c,
       init_reg  <<  augc3sa2ma.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2ma.shw;
show cases ! estimates=wle >> augc3sa2ma.wle;
show cases ! estimates=eap >> augc3sa2ma.eap;
show parameters !tables=7 >> augc3sa2ma.thr;
equivalence wle >> augc3sa2ma.eqv;
itanal ! estimates=wle >> augc3sa2ma.itn;

reset;
Title  IT-PS-SA MOT PCM augmented cond3 analysis;
set update = yes;
set warnings=no;
set constraints=cases;
set n_plausible = 1;
set mle_max=30;
set keeplast=yes;
datafile augcond3sa2.dat;
format responses 50-53, 94-96, 125-126;
labels << augc3sa2mo.nam;
codes 0,1,2,3,4,5;
recode (1,2,3,4,5,6) (0,1,2,3,4,5) ! items (1-4);
score (0,1,2,3,4,5)  (0,1,2,3,4,5) ! items (1-9);
model item + item*step;
export log       >>  augc3sa2mo.l,
       par       >>  augc3sa2mo.p,
       cov       >>  augc3sa2mo.c,
       reg_coeff >>  augc3sa2mo.r;
/*import init_par  <<  augc3sa2mo.p,
       init_cov  <<  augc3sa2mo.c,
       init_reg  <<  augc3sa2mo.r;*/
estimate ! method=quadrature;
show ! estimates=latent;
show ! estimates=mle;
show ! estimates=wle >> augc3sa2mo.shw;
show cases ! estimates=wle >> augc3sa2mo.wle;
show cases ! estimates=eap >> augc3sa2mo.eap;
```

```
show parameters !tables=7 >> augc3sa2mo.thr;
equivalence wle >> augc3sa2mo.eqv;
itanal ! estimates=wle >> augc3sa2mo.itn;
quit;
```

Bibliography

Adams, M. J. (1990). *Learning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adams, R. J., & Wilson, M. R. (1996). Formulating the Rasch model as a mixed coefficients multinomial logit. In G. Engelhard & M. Wilson (Eds.), *Objective measurement: Theory into practice (Vol. III*; pp. 143-166). Norwood, NJ: Ablex.

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

Adams, R.J., & Khoo, S.T. (1996). *Quest: The interactive test analysis system*. Melbourne: Australian Council for Educational Research.

Alexander, J.E., & Filler, R.C. (1976). *Attitudes and reading*. Newark, DE: International Reading Association.

American Academy of Pediatrics, American Public Health Association, & National Resource Center for Health and Safety in Child Care (2002). *Caring for our children: National health and safety performance standards: Guidelines for out-of-home child care program (2nd ed.)*. Elk Grove Village, IL: American Academy of Pediatrics.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington. DC: AERA.

Anderson, R. C. (1982). Allocation of attention during reading. In A. Falmmer & W. Kintsch (Eds.), *Discourse processing*. New York: North-Holland.

Anglin, J. (1995). Classifying the world through language: Functional relevance, cultural significance, and category name learning. *International Journal of Intercultural Relations*, *19*, 161-181.

Anthony, J. L., Lonigan, C. J., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*, 470-487.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Bloomfield, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82,* 65-92.

Arehart, D., & Haaf, R. A. (1989). *Detection of scene schema violations by 12-month-old infants.* Poster session presented at the meetings of the Society for Research in Child Development, Kansas City, MO.

Asher, S. R. (1979). Influence of topic interest on black children's and white children's reading compression. *Child Development, 50,* 686-690.

Ausubel, D. P., & Ausubel, P. (1966). Cognitive development in adolescence. *Review of Educational Research, 36*(4), 403-413.

Bamberg, M., & Moissinac, L. (2003). Discourse development. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 395-437). Mahwah, NJ:Erlbaum.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Barnes, K. E. (1971). Preschool play norms: A replication. *Developmental Psychology, 5*(1), 99-103.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bauer, P. J. (1997). Development of memory in early childhood. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 83-111). Hove, England: Psychology Press.

Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1- to 2-year-olds' recall of events. *Developmental Psychology, 25*(2), 197-206.

Bauer, P. J., & Shore, C. M. (1987). Making a memorable event: Effects of familiarity and organization on young children's recall of action sequences. *Cognitive Development, 2*(4), 327-338.

Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development, 20*, 247-258.

Becker, J. (1989). Preschoolers' use of number words to denote one-to-one correspondence. *Child Development, 60*(5), 1147-1157.

Beers, J. W., & Henderson, E. H. (1977). A study of developing orthographic concepts among first grade children. *Research in the Teaching of English, 11*, 133-148.

Bergen, D. (2002). The role of pretend play in children's cognitive development. *Early Childhood Research & Practice, 4*(1).

Bergman, A., & Wilson, A. (1984). Thoughts about stages on the way to empathy and the capacity for concern. In J. D. Lichtenberg, M. Bornstein & D. Silver (Eds.), *Empathy, Vol. 1* (pp. 59-80). Hillsdale, N.J.: Analytic Press.

Bernstein, M. R. (1955). Relationship between interest and reading comprehension. *Journal of Educational Research, 49*, 283-288.

Birch, L. L. (1990). Development of food acceptance patterns. *Developmental Psychology, 26*, 515-519.

Bissex, G. L. (1980). Patterns of development in writing: A case study. *Theory into Practice, 19*(3), 197-201.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York: Wiley.

Blue, N. (1995). *What causes causality? The development of causal reasoning in young children.* Poster session presented at the British Psychology Society Developmental Section Conference, Glasgow.

Braungart, J. M., & Stifter, C. A. (1991). Regulation of negative reactivity during the strange situation: Temperament and attachment in 12-month-old infants. *Infant Behavior and Development, 14*, 349-364.

Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functionalist perspective. *Child Development, 55*, 529-548.

Bullock, M., & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development, 50*(1), 89-96.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209-254). New York: Academic Press.

Burns, M. S., Griffin, P., & Snow, C. E. (1999). *Starting out right*. Washington, DC: National Academy Press.

Buss, K. A., & Goldsmith, H. H. (1998). Fear and anger regulation in infancy: Effects on the temporal dynamics of affective expression. *Child Development, 69*, 359-374.

Calkins, S., & Johnson, M. C. (1998). Toddler regulation of distress to frustrating events: temperamental and maternal correlates. *Infant Behavior and Development, 21*, 379-395.

Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser & T. P. Romberg (Eds.), *Addition and subtraction: A cognitive perspective*, (pp. 9-24). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school language in African American children and its relation to early reading achievement. *Child Development, 75(*5), 1340-1356.

Chen, Z., & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development, 65*(2), v-96.

Chomsky, C. (1972). Stages in language development and reading exposure. *Harvard Educational Review, 42*, 1–33.

Clarke, L. K. (1988). Invented versus traditional spelling in first graders' writings: Effects on learning to spell and read. *Research in the Teaching of English, 22*, 281-309.

Clay, M. M. (1977). Exploring with a pencil. *Theory into Practice, 16*(5), 334-341.

Clay, M. M. (1993). *Reading recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.

Clay, M. M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Portsmouth, NH: Heinemann

Clements, D. H. (2004a). Geometric and spatial thinking in early childhood. In D. H. Clements, A. DiBiase & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 267-298). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D. H. (2004b). Major themes and recommendations. In D. H. Clements, A. DiBiase & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 7-72). .Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D. H., & Sarama, J. (2000). Young children's ideas about geometric shapes. *Teaching Children Mathematics, 6*(8), 482-488.

Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements, A. DiBiase & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 299-317). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education, 30*(2), 192-212.

Cooper, R., & Waldau, R.C. (1999). *Development of Problem Solving Strategies: Strategy Shift in the Tower of London Task*. Paper presented at the 21st Annual Conference of the Cognitive Science Society. Retrieved July 25, 2009 from http://cogent.psyc.bbk.ac.uk/publications/cooperwaldau99.html

Corsaro, W. A. (2005). *The sociology of childhood* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-Processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74-101.

Cummings, E. M. (1987). Coping with background anger in early childhood. *Child Development, 58*, 976-984.

Curenton, S., & Justice, L. M. (2004). Low-income preschoolers' use of decontextualized discourse: Literate language features in spoken narratives. *Language, Speech, and Hearing Services in Schools, 35*, 240-253.

Curry, N. E., & Johnson, C. N. (1990). *Beyond self-esteem: Developing a genuine sense of human value.* Washington, DC, U.S.: National Association for the Education of Young Children.

Damon, W. (1988). The *moral child: Nurturing children's natural moral growth*. New York: Free Press.

Damon, W., & Hart, D. (1982). The development of self-understanding from infancy through adolescence. *Child Development, 53*(4), 841-864.

Das Gupta, P., & Bryant, P. E. (1989). Young children's causal inferences. *Child Development, 60*(5), 1138-1146.

DeLoache, J. (1996). Shrinking trolls and expanding minds: How very young children learn to understand and use symbols. *Zero to Three, 17*, 10-16.

DeLoache, J. S. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development, 4*(2), 121-139.

Devlin, K. J. (2000). *The math gene: How mathematical thinking evolved and why numbers are like gossip.* United States: Basic Books.

Diamond, K. E. (1993). Preschool children's concepts of disability in their peers. *Early Education and Development, 4*, 123-129.

Diamond, K. E. (2001). Relationships among young children's ideas, emotional understanding, and social contact with classmates with disabilities. *Topics in Early Childhood Special Education*, *21*, 104-113.

Dickinson, D. K., & Snow, C. E. (1987). Interrelationships among prereading and oral language skills in kindergartners from two social classes. *Early Childhood Research Quarterly, 2*, 1–25.

Dickinson, D. K., & Tabors, P. O. (1991). Early literacy: Linkages between home, school, and literacy achievement at age five. *Journal of Research in Childhood Education*, *6*, 30-46.

Donowitz, L.G. (Ed.) (1996). *Infection control in the child care center and preschool* (2nd ed.). Baltimore, ME: Williams & Wilkins.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., & Klebanov, P. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.

Dunn, J., & Herrera, C. (1997). Conflict resolution with friends, siblings, and mothers: A developmental perspective. *Aggressive Behavior, 23*, 343-357.

Durkin, K. (1995). *Developmental social psychology: From infancy to old age.* Malden, MA: Blackwell Publishing.

Durrell, D. D. (1980). Commentary: Letter name values in reading and spelling. *Reading Research Quarterly*, *16*, 159-163.

Dweck, C. S. (1989). Motivation. In A. Lesgold, & R. Glaser (Eds.), *Foundations for a Psychology of Education* (pp. 87-136). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development.* New York: Psychology Press.

Dweck, C. S., & Elliot, E. (1983). Achievement motivation. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 643-691). New York: Wiley.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256-273.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology, Vol. 3. Social, emotional and personality development* (5th ed., pp. 1017-1095). New York: Wiley.

Ehri, L. C. (1988). Movement in word reading and spelling: How spelling contributes to reading. In J. Mason (Ed.), *Reading and Writing Connections*. Newton, MA: Allyn & Bacon.

Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In T. Waller & Mackinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 1), 63-111. New York: Academic Press.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18(*2), 116-125.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading. Special Issue: The Contribution of Psychological Research, 18*(2), 116-125.

Ehri, L. C. (1998). Grapheme-morpheme knowledge is essential for learning to read words in english. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ: Erlbaum.

Eisenberg, A. R. (1985). Learning to describe past experiences in conversation. *Discourse Processes, 8*(2), 177-204.

Eisenberg, N., Spinrad, T. L., & Sadovsky, A. (2006). Empathy-related responding in children. In M. Killen, & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 517-549). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Elicker, J., Englund, M., & Sroufe, L. A. (1992). Predicting peer competence and peer relationships in childhood from early parent-child relationships. In R. D. Parke & G. W. Ladd (Eds.), *Family-peer relations: Modes of linkage* (pp. 77-106). Hillsdale, NJ: Erlbaum.

Embretson, S.E., & Reise S. P., (2000). *Item response theory for psychologists.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Endres, J. B., & Rockwell, R. E. (1994). *Food, nutrition, and the young child*. (4th ed.). New York, NY: Merrill Publishing Co.

Erikson, E. H. (1964). *Childhood and society* (2nd ed.). Oxford, England: W. W. Norton.

Espinosa, M. P., Sigman, M. D., Neumann, C. G., Bwibo, N. O., & McDonald, M. A. (1992). Playground behaviors of school-age children in relation to nutrition, schooling, and family characteristics. *Developmental Psychology*, *28*(6), 1188-1195

Estes , T. H., & Vaughan, J. L. (1973). Reading interest comprehension: Implications. *Reading Teacher, 27*, 149-153.

Farrar, M. J., & Goodman, G. S. (1992). Developmental changes in event memory. *Child Development, 63*(1), 173-187.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314.

Fein, G. G., & Stork, L. (1981). Sociodramatic play: Social class effects in integrated preschool classrooms. *Journal of Applied Developmental Psychology, 2*(3), 267-279.

Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Portsmouth, NH: Heinemann.

Fivush, R., & Haden, C. A. (1997). Narrating and representing experience: Preschoolers' developing autobiographical accounts. In P. W. van den Broek, P. J. Bauer & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events* (pp. 169-198). Hillsdale, NJ: Erlbaum.

Fivush, R., & Mandler, J. M. (1985). Developmental changes in the understanding of temporal sequence. *Child Development*, *56*(6), 1437-1446.

Fivush, R., & Slackman, E. (1986). The acquisition and development of scripts. In K. Nelson, & J. Gruendel (Eds.), *Event knowledge: Structure and function in development* (pp. 71-96). Hillsdale, N.J.: Erlbaum.

Fivush, R., Gray, J. T., & Fromhoff, F. A. (1987). Two-year-olds talk about the past. *Cognitive Development*, *2*(4), 393-409.

Fivush, R., Haden, C., & Adam, S. (1995). Structure and coherence of preschoolers' personal narratives over time: implications for childhood amnesia. *Journal of Experimental Child Psychology*. *Special Issue: Early Memory, 60*(1), 32-56.

Fivush, R., Kuebli, J., & Clubb, P. A. (1992). The structure of events and event representations: A developmental analysis. *Child Development, 63*(1), 188-201.

Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research, 4*, 331-342.

Fox, N.A. (1994). The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*, 240.

Friedman, W. J. (1990). Children's representations of the pattern of daily activities. *Child Development, 61*(5), 1399-1412.

Friedman, W. J. (1993). Memory for the time of past events. *Psychologica*l *Bulletin, 113*(1), 44-66.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag Publishing.

Fuson, K. C. (1992a). Relationships between counting and cardinality from age 2 to age 8. In J. Bideaud, C. Meljac & J. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 127-149). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fuson, K. C. (1992b). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 243-275). New York: Macmillan.

Garber, J., & Dodge, K. A., (Eds.). (1991). *The development of emotional regulation and dysregulation*. New York: Cambridge University Press.

Garvey, C. (1984). *Children's talk*. Cambridge, MA: Harvard University Press.

Garvey, C. (1990). *Play.* Cambridge, MA: Harvard University Press.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC, US: American Psychological Association.

Gelman, R., Bullock, M., & Meck, E. (1980). Preschoolers' understanding of simple object Transformations. *Child Development, 51*(3), 691-699.

Ginsburg, H. P., Inoue, N., & Seo, K. (1999). Young children doing mathematics: Observations of everyday activities. In J. V. Copley (Ed.), *Mathematics in the early years* (pp. 88-99). Washington, DC: National Association for the Education of Young Children.

Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools, 33*(4), 268-281.

Goswami, U. (1998). *Cognition in children*. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.

Goswami, U. C., & Bryant, P. E. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum.

Gowen, J. W. (1995). Research in review. The early development of symbolic play. *Young Children*, *50*(3), 75-84.

Graves, D.E., Suitor, C.W., Holt, K.A. (Eds.) (1997). *Making food healthy and safe for children: How to meet the National Health and Safety Performance Standards-Guidelines for out of home child care programs.* Vienna, VA: National Maternal and Child Health Clearinghouse.

Greenes, C. (1999). Ready to learn: Developing young children's mathematical powers. In J. V. Copley (Ed.), *Mathematics in the early years.* (pp. 39-47). Washington, DC, US: National Association for the Education of Young Children.

Grolnick, W. S., Bridges, L. J., & Connell, J. P. (1996). Emotion regulation in two-year-olds: strategies and emotional expression in four contexts. *Child Development*, *67*, 928-941.

Gunn, B. K., Simmons, D. C., & Kameenui, E. J. (1995). *Emergent literacy: Synthesis of the research* (Technical Report No. 19). Eugene: University of Oregon, National Center to Improve the Tools of Educators.

Hamond, N. R., & Fivush, R. (1991). Memories of Mickey Mouse: Young children recount their trip to Disneyworld. *Cognitive Development*, *6*(4), 433-448.

Harner, L. (1982). Talking about the past and future. In W. J. Friedman (Ed.), The *developmental psychology of time* (pp. 141-170). New York: Academic Press.

Harris, P. L. (1993). Understanding emotion. In M. Lewis and J. M. Haviland (Eds.), *Handbook of emotions* (pp. 237-246). New York: Guilford.

Harste, J. C., Woodward, V. A., & Burke, C. L. (1984). Examining our assumptions: A transactional view of literacy and learning. *Research in the Teaching of English, 18*(1), 84-108.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Brookes.

Harter, S. (1990). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In R. J. Sternberg, & J. Kolligian Jr. (Eds.), *Competence considered* (pp. 67-97). New Haven, CT: Yale University Press.

Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development* (5th ed., pp. 553-617). Hoboken: Wiley.

Harter, S. (1999). *The construction of the self*. New York: Guilford Press.

Hartup, W. W. (1983). The peer system. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 54-68). New York: Wiley.

Hartup, W. W. (1983a). Peer relations. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 120-163). New York: Wiley.

Hawks, D., Ascheim, J., Giebink, G.S., Graville, S., Solnit, A.J. (1994). Science, prevention, and practice VII: Improving child day care, a concurrent summary of the American Public Health Association/American Academy of Pediatrics National Health and Safety Guidelines for Child-Care Programs; *Featured Standards and Implementation. Pediatrics*, *95*(6): 1110-1112.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classroom.* New York: Cambridge University Press.

Henderson, E. H., & Beers, J. W. (Eds.) (1980). *Developmental and cognitive aspects of learning to spell: A reflection of word knowledge.* Newark, DE: International Reading Association.

Hestenes, L. L., & Carroll, D. E. (2000). The play interactions of young children with and without disabilities: Individual and environmental influences. *Early Childhood Research Quarterly, 15*(2), 229-246.

Heyman, G. D., & Dweck, C. S. (1998). Children’s thinking about traits: Implications for judgments of the self and others. *Child Development*, *69*(2), 391-403.

Hidi, S., & Baird, W. (1986). Interestingness—A neglected variable in discourse processing. *Cognitive Science*, *10*(2), 179-194.

Hidi, S., & Baird, W. (1988). Strategies for increasing text-based interest and students’ recall of expository text. *Reading Research Quarterly, 23*(4), 465-483.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*(4), 549-571.

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13*(3), 191-209.

Hiebert, J. (1997). *Making sense:Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hoff, E. (2005). *Language development* (3rd ed.). Belmont, CA: Wadsworth.

Hoffman, M. L. (1990). Empathy and justice motivation. Motivation and emotion. *Special Issue: Empathy, 14*(2), 151-172.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice.* New York: Cambridge University Press.

Howe, N., Rinaldi, C. M., Jennings, M., & Petrakos, H. (2002). ‘No! The lambs can stay out because they got cozies!’: Constructive and destructive sibling conflict, pretend play, and social understanding. *Child Development, 73*(5), 1460-1473.

Hudson, J. A., & Shapiro, L. (1991). Effects of task and topic on children's narratives. In A. McCabe & C. Peterson (Eds.), *New directions in developing narrative structure* (pp. 89-136). Hillsdale, NJ: Erlbaum.

Hudson, J. A. (1990). Constructive processing in children's event memory. *Developmental Psychology, 26(*2), 180-187.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82*, 2–28.

Hutt, S., Tyler, S., Hutt, C., & Christopherson, H. (1989). Play, exploration, and learning: A natural history of the pre-school. London: Routledge.

James, W. (1913). *The principles of psychology*. New York: Henry Holt

Johnson, J. E., & Ershler, J. (1981). Developmental trends in preschool play as a function of classroom program and child gender. *Child Development*, *52*(3), 995-1004.

Johnson, J. E., Christie, J. F., Yawkey, T. D., & Wardle, F. (1987). *Play and early childhood development.* Glenview, IL: Scott, Foresman.

Justice, L. M., & Ezell, H. K. (2004). Print referencing: An emergent literacy enhancement technique and its clinical applications. *Language, Speech, and Hearing Services in Schools, 35*, 185-193.

Justice, L. M., Pence, K., Bowles, R., & Wiggins, A. (2006). An investigation of four hypotheses concerning the order by which 4-year-old children learn the alphabet letters. *Early Childhood Research Quarterly, 21*(3), 374-389.

Katz, L. G. (1993). *Dispositions: Definitions and implications for early childhood practices. Perspectives from ERIC/EECE: A monograph series, No. 4.* Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED360104, p. A).

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* J.Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would Read a story anyway. *Poetics, 9*, 87–89.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13*(1), 113-148.

Klein, A., & Starkey, P. J. (2004). Fostering preschool children's mathematical knowledge: Findings from the Berkeley Math Readiness Project. In D. H. Clements, A. DiBiase & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp.343-360). Hillsdale, NJ: Erlbaum

Kochanska, G., Padavich, D. L., & Koenig, A. L. (1996). Children's narratives about hypothetical moral dilemmas and objective measures of their conscience: Mutual relations and socialization antecedents. *Child Development, 67*, 1420-1436.

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*, 263-277.

Kolen, M. J. (2004). Linking assessments: Concept and history. *Applied Psychological Measurement, 28*(4), 219-226.

Kopp, C. B., & Gralinski, J. H. (1993). Everyday rules for behavior: Mothers' requests to young children. *Developmental Psychology, 29*(3), 573-584.

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. *Developmental Psychology*, *25*, 343-354.

Kotch, J. B. (1990). *Reduction in transmission of disease in child care settings: Final report*. Chapel Hill, NC: University of North Carolina at Chapel Hill, Department of Maternal and Child Health.

Krapp, A., Hidi, S., & Renniger, A. (1992). Interest, learning and development. In R.A. Renniger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development*. Hillsdale, NY: Erlbaum.

Kress, G. N. (1994). *Learning to write*. New York: Routledge.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotive, W. (1988). *The development of scientific thinking skills.* San Diego, CA, US: Academic Press.

Ladd, G. W. (1992). Themes and theories: Perspectives on processes in family-peer relationships. In R. D. Parke & G. W. Ladd (Eds.), *Family-peer relations: Modes of linkage* (pp. 3-34). Hillsdale, NJ: Erlbaum.

Lally, J.R., Griffin, A., Fenichel, E., Segal, M., Szanton, E., Weissbourd, B. (1995). *Caring for infants and toddlers in groups: Developmentally appropriate practice*. Washington, D.C.: Zero to Three.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landry, S. H., Miller-Loncar, C. L., Smith, K. E., & Swank, P. R. (1997). Predicting cognitive-language and social growth curves from early maternal behaviors in children at varying degrees of biological risk. *Developmental Psychology, 33,* 1040-1053.

Lerner, R. M., & Hultsch, D. F. (1983). *Human development: A life-span perspective*. New York: McGraw-Hill.

Lesiak, J. L. (1991). Research based answers to questions about emergent literacy in kindergarten. *Psychology in the Schools, 34*, 143-160.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self.* New York: Plenum Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Lillard, A. S. (1993). Young children's conceptualization of pretense: Action or mental representational state? *Child Development, 64*(2), 372-386.

Lillard, A. S. (1998). Playing with a theory of mind. In O. N. Saracho, & B. Spodek (Eds.), *Multiple perspectives on play in early childhood education* (pp. 11-33). Albany, NY: State University of New York Press.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education, 6*(1), 83-102.

Longford, N. T., Holland, P. W., & Thayer, D. T. (1993). Stability of the MH D-DIF statistics across populations. In P. W. Holland, & H. Wainer (Eds.), *Differential item functioning* (pp. 171-196). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Lonigan, C. (2006). Conceptualizing phonological processing skills in pre-readers. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research: Vol. 2.* New York: Guilford.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a Latent Variable Longitudinal Study. *Developmental Psychology, 36*, 596-613.

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in two- to five-year-old children. *Journal of Educational Psychology, 90*, 294-311.

MacLean, M., Bryant, P., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly, 33*, 255-282.

Mandler, J. M. (1984). Stories, scripts, and scenes: Aspects of schema theory. Hillsdale, N.J.: L. Erlbaum Associates, Publishers.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Mathewson, G. C. (1994). Toward a comprehensive model of affect in the reading process. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd Ed., pp. 841–856). Newark, DE: International Reading Association.

McDevitt, T. M., & Ormrod, J. E. (2002). *Child development and education*. Upper Saddle River, NJ: Merrill Prentice Hall.

McKenna, M. C., Kear, D. J., & Ellsworth, R. A., (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30*(4), 934-955.

McMullen, M. B., & Darling, C. A. (2003). Learning from their mistakes: Glimpses of symbolic functioning in two-and-a-half to three-year-old children. *Early Years: An International Journal of Research and Development, 23* (1), 55-66.

Messick, S. (1994, September). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning* (Research Report RR-94-45). Princeton, NJ: Educational Testing Service.

Metzuk, B., Parrish, D., & Mangione, P. (2006). *Overview of the Desired Results for Children and Families System*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359-381.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects.* Princeton, NJ: Educational Testing Service.

Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement, 13*(1), 57-75.

Moore D.S., & McCabe (1998). *Introduction to the practice of statistics, 3rd Edition*. New York: W.H. Freeman and Company.

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65,* 370–398.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004) Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665-681.

Myers, N. A., & Bluhm, C. (1985). *Preschool children's memory of a classroom event.* Paper presented at the meeting of the Society for Research in Child Development, Toronto.

National Association for Sport and Physical Education (1995). *Moving into the future: National standards for physical education*. St. Louis, MO: Mosby.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* J.Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests.* M. J. Feuer, P. W. Holland, B. F. Green, M. W. Bertenthal, & F. Cadell Hemphill (Eds.). Committee on Equivalency and Linkage of Educational Tests, National Research Council Washington, DC: National Academy Press.

Nelson, K. (1978). How children represent knowledge of their world in and out of language: A preliminary report. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 255-273). Hillsdale, NJ: Erlbaum.

Nelson, K. (1989). *Narratives from the crib*. Cambridge, MA: Harvard University Press.

Nelson, K., & Gruendel, J. (1986). *Event knowledge: Structure and function in development.* Hillsdale, NJ: Erlbaum.

Nelson, K., & Gruendel, J. (1986). Generalized event representations: Basic building blocks of cognitive development. In M.E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. 1, pp. 131-158). Hillsdale, NJ: Erlbaum.

Nelson, K., & Ross, G. (1980). The generalities and specifics of long-term memory in infants and young children. In M. Perlmutter (Ed.), *Children's memory: New directions for child development* (Vol. 10, pp. 87-101). San Francisco: Jossey-Bass.

Newborg, J. (2005). *Battelle developmental inventory* (Second Ed.). Itasca, IL: Riverside.

Niffenegger, J.P. (1997). Proper handwashing promotes wellness in child care. *Journal of Pediatric Health Care, 11*, 26-31.

Owens, R.E. (1996). *Language development: An introduction* (4th Ed.). Boston: Allyn and Bacon.

Paek, I. (2002). *Investigation of differential item function: Comparisons among approaches, and extension to a multidimensional context.* Unpublished doctoral dissertation, University of California, Berkeley.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative competence in young children. *Reading Research Quarterly, 38*, 36-42.

Parke, R. D., & Ladd, G. W. (Eds.). (1992). *Family-peer relations: Modes of linkage*. Hillsdale, NJ: Erlbaum.

Parten, M. B. (1932). Social participation among pre-school children. *The Journal of Abnormal and Social Psychology, 27*(3), 243-269.

Pellegrini, A. D., Galda, L., Jones, I., & Perlmutter, J. (1995). Joint reading between mothers and their head start children: Vocabulary development in two text formats. *Discourse Processes, 19*, 441–463.

Perfetti, C. A., Van Dyke, J., & Hart, L. (2001). The psycholinguistics of basic literacy. Annual *Review of Applied Linguistics, 21*, 127-149.

Peterson, C., & Mccabe, A. (1992). Parental styles of narrative elicitation: Effect on children's narrative structure and content. *First Language, 12*(36), 299-321.

Phillips, S. U. (1983). The Invisible Culture. New York: Longman.

Piaget, J. (1928). Judgment and Reasoning in the Child. Oxford, England: Harcourt, Brace.

Piaget, J. (1962). *Play, dreams, and imitation*. New York: Norton.

Piaget, J. (1963). *The origins of intelligence in children*. New York: W W Norton & Co.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theories, research, and application*. Engelwood Cliffs, NJ: Merrill/Prentice Hall.

Pipes, P. L., & Trahms, C. M. (1993). *Nutrition in infancy and childhood* (5th ed.). St. Louis, MO: Mosby.

Price, D. W., & Goodman, G. S. (1990). Visiting the wizard: Children's memory for a recurring event. *Child Development, 61*(3), 664-680.

Puckett, M. B., & Black, J. (2005). *The young child: Development from prebirth to age eight* (4th Ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Purcell-Gates, V. (1996). Stories, coupons, and the TV Guide: Relationships between home literacy experience and emergent literacy knowledge. *Reading Research Quarterly, 31*, 406-428.

Purcell-Gates, V., & Dahl, K. L. (1991). Low-SES children's success and failure at early literacy learning in skills-based classrooms. *Journal of Reading Behavior, 23*(1), 1-34.

Ramsey, P. G. (1991). Young children's awareness and understanding of social class differences. *Journal of Genetic Psychology, 152*, 71-82.

Ramsey, P. G., & Myers, L. C. (1990). Salience of race in young children's cognitive, affective, and behavioral responses to social environments. *Journal of Applied Developmental Psychology, 11*, 49-67.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press.

Ratner, H. H., Smith, B. S., & Dion, S. A. (1986). Development of memory for events. *Journal of Experimental Child Psychology, 41*(3), 411-428.

Roberts, L., Mapp, E., Smith, W., Jorm, L., Patel, M., Douglas, R.M., & McGilchrist, C. (2000). Effect of infection control measures on the frequency of upper respiratory infection in child care: A randomized, controlled trial. *Pediatrics, 105*(Suppl. 4, Pt. 1), 738-742.

Ross, H. S., & Conant, C. L. (1992). The social structure of early conflict: Interaction, relationships, and alliances. In C. U. Shantz & W. W. Hartup (Eds.), Conflict in Child and Adolescent Development. *Cambridge studies in social and emotional development; Conflict in child and adolescent development* (pp. 153-185). New York: Cambridge University Press.

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon (Ed.), *Handbook of child psychology: Vol. 3: Social, emotional, and personality development* (5th ed., pp. 105-176). New York: Wiley.

Rubin, K. H. (1980). *Children's play*. San Francisco: Jossey-Bass.

Rubin, K. H., Bukowski, W., & Parker, J. G. (1998). Peer interactions, relationships, and groups. In W. Damon, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. social, emotional, and personality development.* (5th ed., pp. 619-700). Hoboken, NJ: Wiley.

Rubin, K. H., Fein, G. G., & Vanderberg, B. (1983). Play. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 693-774.). New York: Wiley.

Rubin, K. H., Watson, K. S., & Jambor, T. W. (1978). Free-play behaviors in preschool and kindergarten children. *Child Development*, *49*(2), 534-536.

Ruble, D. N., Grosovsky, E. H., Frey, K. S., & Cohen, R. (1992). Developmental changes in competence assessment. In A. K. Boggiano & T. S. Pittman (Eds.), *Achievement and motivation: A social-developmental perspective*. (pp. 138-164). New York: Cambridge University Press.

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific Thinking: children's understanding of the hypothesis-evidence relation. *Child Development, 64*(6), 1617-1636.

Saarni, C. (1990). Emotional competence: How emotions and relationships become integrated. In R.A. Thompson (Ed.), *Nebraska Symposium on Motivation: Vol. 36. Socioemotional development* (pp. 115-182). Lincoln: University of Nebraska Press.

Saarni, C. (1999). *Developing emotional competence*. New York: Guilford.

Salkind, N.J. (2002). *Macmillan encyclopedia of child development.* New York: Macmillan.

Sanders, S. W. (2002). *Active for life: Developmentally appropriate movement programs for young children*. National Association for the Education of Young Children.

Sandy, S. V., & Cochran, K. M. (2000). The development of conflict resolution skills in children: Preschool to adolescence. In M. Deutsch & P. T. Coleman (Eds.), *The handbook of conflict resolution: Theory and practice* (pp. 316-342). San Francisco: Jossey-Bass.

Schank, R. C. (1979). Interestingness: Controlling inferences. *Artificial Intelligence, 12*, 273-297.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Oxford, England: Lawrence Erlbaum.

Schiefflin, B. B., & Cochran-Smith, M. (1984). Learning to read culturally: Literacy before schooling. In H. Goelman, A. Oberg, & F. Smith (Eds.) *Awakening to Literacy*. Portsmouth, NH: Heinemann.

Schneider, W., & Bjorklund, D. F. (1992). Expertise, aptitude, and strategic remembering. *Child Development*, *63*(2), 461-473.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior, 27*, 1-17.

Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington D.C.: National Academy Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development, 46*(2), 394-399.

Shultz, T. R., Pardo, S., & Altman, E. (1982). Young children's use of transitive inference in causal chains. British *Journal of Psychology, 73*, 235-241.

Sigelman, C. K., & Shaffer, D. R. (1991). *Life-span human development.* Pacific Grove, CA: Thomson Brooks/Cole.

Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script in story form by young children. *Child Development, 55*(2), 329-340.

Slawinski, J., & Best, D. (1995). *Effects of meta-memory on children's predictions of recall.* Paper presented at the meeting of the Society for Research in Child Development: Indiana.

Smilansky, S. (1968). *The effects of sociodramatic play on disadvantaged preschool children.* Oxford, England: Wiley

Smith, P. H., deSaint-Victor, C. M., & Arehart, D. M. (1988). *Evidence for schematic representation in 5-month-old infants*. Paper presented at the International Conference on Infant Studies, Washington, DC.

Smith, P. K. (1978). A longitudinal study of social participation in preschool children: Solitary and parallel play reexamined. *Developmental Psychology, 14*(5), 517-523.

Snow, C.E. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review, 53*, 165-189.

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development, 3*(2), 183-222.

Soto, J.C., Guy, M., & Belanger, L. (1994). Science, prevention and practice II: Preventing infectious diseases, abstracts on handwashing and infection control in day-care centers. *Pediatrics, 94*(6), 1030.

Spencer, M. B., & Markstrom-Adams, C. (1990). Identity processes among racial and ethnic minority children in America. *Child Development, 61*, 290-310

Sperry Smith, S. (2001). *Early childhood mathematics* (2nd ed.). Boston: Allyn and Bacon.

Steen, L. A., & National Research Council. (1990). *On the shoulders of giants :New approaches to numeracy*. Washington, D.C.: National Academy Press.

Stein, N. L., & Levine, L. J. (1989). The causal organisation of emotional knowledge: A developmental study. *Cognition and Emotion, 3,* 343-378.

Steinberg, L., & Meyer, R. (1995). *Childhood*. New York: McGraw-Hill.

Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development, 55*(6), 1983-1989.

Stipek, D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development, 57*(1), 100.

Sullivan, S. A., & Birch, L. L. (1990). Pass the sugar, pass the salt: Experience dictates preference. *Developmental Psychology, 26*, 546-551.

Sulzby, E. (1986). Writing and reading: Signs of oral and written language organization in the young child. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Reading and writing* (pp. 50-87). Norwood, NJ: Ablex.

Sulzby, E. (1987). Children's development of prosodic distinctions in telling and dictation modes. In A. Matsuhashi (Ed.), *Writing in real time: Modeling production processes* (pp. 133-160). Westport, CT: Ablex.

Tamis-LeMonda, C. S., & Bornstein, M. H. (1991). Individual variation, correspondence, stability, and change in mother and toddler play. *Infant Behavior & Development, 14*(2), 143-162.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767.

Taylor, D., & Dorsey-Gaines, C. (1998). *Growing up literate: Learning from inner-city families.* Portsmouth, N. H.: Heinemann.

Teale, W. H. (1987). Emergent literacy: Reading and writing development in early childhood. In J. Readence & R. S. Baldwin, (Eds.), *Research in literacy: Merging perspectives--Thirty-sixth yearbook of the National Reading Conference* (pp. 45-74). Rochester, NY: National Reading Conference.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. In N. A. Fox (Ed.), *The development of emotion regulation and dysregulation: Biological and behavioral aspects. Monographs of the Society for Research in Child Development, 59* (2-3), 25-52 (Serial no. 240).

Thompson, R. A. (1990). Emotion and self-regulation. In R. A. Thompson (Ed.) *Socioemotional development: Nebraska Symposium on motivation, Vol. 36* (pp. 383-483). Lincoln, NE: University of Nebraska Press.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

Torgesen, J. K., & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness. *Journal of Experimental Child Psychology, 63,* 1-21.

Treiman, R. (1992). The Role of intrasyllabic units in learning to read and spell. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 65–106). Hillsdale, NJ: Erlbaum.

Treiman, R., & Broderick, V. (1998). What's in a name: Children's knowledge about the letters in their own name. *Journal of Experimental Child Psychology, 70*(2), 97-116

Troseth, G. L., Pickard, M. E. B., & DeLoache, J. S. (2007). Young children's use of scale models: Testing an alternative to representational insight. *Developmental Science, 10*(6), 763-769.

U.S. Department of Agriculture (1995). *Child and adult care food program: Nutrition guidance for child care homes*. Washington, DC: Author, Family Child Services.

U.S. Department of Agriculture (1994). *Understanding healthy lifestyle, and exercise and fitness*. Washington, DC: Author, Family Child Services.

Van de Walle, J. A. (2001). *Elementary and middle school mathematics: Teaching developmentally* (4th ed.). New York: Longman.

van Kleeck, A. (1990). Emergent literacy: Learning about print before learning to read. *Topics in Language Disorders, 10*, 25–45.

Vaughn, B. E., Kopp, C. B., & Krakow, J. B. (1984). The emergence and consolidation of self-control from eighteen to thirty months of age: Normative trends and individual difference. *Child Development, 55,* 990-1004.

Want, S. C., & Harris, P. L. (2001). Learning from other people's mistakes: Causal understanding in learning to use a tool. *Child Development*, *72*(2), 431-443.

Wehren, A., DeLisi, R., & Arnold, M. (1981). The development of noun definition. *Journal of Child Language, 8,* 165-175.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergartners. Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Fall 1998* (NCES Publication No. 2000-2070). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Whitehurst, G. J. & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872.

Whitehurst, G., & Lonigan, C. (2001). Emergent literacy: Development from prereaders to readers. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11-29). New York: Guilford Press.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology, 13*(2), 231-247.

Wilkening, F. (1982). Children's knowledge about time, distance and velocity interrelations. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 82-112). New York: Academic Press.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*, 181-208.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97-116.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis. Rasch measurement*. Chicago: MESA.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). ConQuest: Generalized Item Response Modeling Software [Computer program]. Camberwell, VIC, Australia: ACER.

Wu, M.L. (1997). *The development and application of a fit test for use with marginal maximum likelihood estimation and generalised item response models*. Unpublished master's thesis, University of Melbourne, Melbourne, Victoria, Australia.

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and Emotion, 14*(2), 107-130.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology, 28*, 1038-1047.

Zelazo, P. D., & Reznick, J. S. (1991). Age-related asynchrony of knowledge and action. *Child Development, 62*(4), 719-735.