A parentally-administered cognitive development assessment for children from 10 to 24 months

Maryhan Baker (1, 2) *

Graham Schafer (1)

Katie J Alcock (3)

Shelley Bartlett (1, 4)

(1) University of Reading, Reading, UK

(2) Bournemouth University, Bournemouth, UK

(3) Lancaster University, Lancaster, UK

(4) Avon and Wiltshire Mental Health Partnership NHS Trust, UK

* Corresponding author: School of Design, Engineering, & Computing, Bournemouth University, Bournemouth, Poole House, Talbot Campus, Poole, Dorset BH12 5BB, UK.

A parentally-administered cognitive development assessment for children from 10 to 24 months

Abstract

The Cognitive Development Questionnaire (CDQ) allows accurate assessment of cognitive development of children from 10 to 24 months by parents and caregivers in the home. It takes between one and two hours to complete over about a week. Three phases of work are described, in which the instrument is progressively refined to improve its validity and reliability. This resulting version of the CDQ shows excellent correlation with age, and with the Mental Scale of the Bayley Scales of Infant Development (Bayley, 1993). The CDQ thus offers researchers and clinicians a useful alternative to professionally-administered cognitive assessment in infancy.

A parentally-administered cognitive development assessment for children from 10 to 24 months

Introduction

Cognitive assessment of children in the first year or two of life is undertaken for health surveillance, diagnosis, and/or research purposes. It typically involves administration of standardized tests by trained assessors. Table 1 lists some of the most commonly used professionally-administered instruments. Some measure cognitive ability only; others encompass domains such as motor, and personal-social skills, and general behavior. Yet, these assessor-administered tests do not utilize one of the richest sources of information on that infant's cognitive ability: the infant's parent or caregiver (Bricker & Squires, 1989; Bricker, Squires, Kaminski, & Mounts, 1988; Fenson, Bates, Reznick, Bates, Thal, & Pethick 1994; Saudino et al., 1998). Parental reports have been extensively used in other domains to assess e.g. attachment, social skills and behavioral problems (Achenbach & Edelbrock, 1983; Goldsmith, 1996; Hogan, Scott, & Bauer, 1992; Waters & Deane, 1985), and communicative development (Dale, Bates, Reznick, & Morisset, 1989; Fenson et al., 1994; Nelson, 1973).

INSERT TABLE 1 ABOUT HERE

Saudino et al. (1998) give a number of reasons to prefer parental[1] reports to professionally-administered tests. First, in a test in the home, a broader sample of infant behavior may contribute to the final assessment of cognitive ability than would be the case with a professionally administered test. Second, parent report measures are more economical than professionally-administered tests, which can involve a trained assessor in several hours of testing. Third, parental report measures can provide relatively economical data for research studies using large samples (see also Bricker et al., 1988; Oliver, et al., 2002). Fourth, parents draw on a broader sample of infant behavior in their assessment, and are thus more likely to observe and to report more accurately on emerging skills, than are professional researchers (Saudino et al., 1998). Fifth, the unnaturalness inherent in testing by a stranger may result in underestimation of ability (Gradel, Thompson, & Sheehan, 1981; Sheehan, 1988).

Strong positive correlations are commonly reported between parental and professional assessment (Dinnebeil & Rule, 1994), including when parents and professionals assess the same child with the same assessment instrument (Beckman, 1984; Blacher-Dixon & Simeonsson, 1981; Bricker & Squires, 1989; Bricker et al., 1988). Parental and professional agreement has also been found to be high when used for the purposes of developmental screening (Glascoe, Altemeier & MacLean, 1989). The accuracy of parental report has also been shown to be unaffected by socio-demographic factors such as level of education and experience in child rearing (Johnson, Marlow, Wolke, Davidson, Marston, O'Hare, Peacock & Schulte, 2007; Glascoe et al., 1989). Yet despite its apparent utility, little use has been made of parent report in evaluating cognitive ability below 2 years. To our knowledge, no parental report instrument focuses solely on *cognitive* development across a *range* of

ages in infancy. Nonetheless, five parentally-administered instruments include measures of cognitive ability and are now briefly reviewed.

The Child Development Inventory (CDI) (Ireton, 1998) is a parent report questionnaire for the screening and assessment of children at risk of development delay. It consists of 270 statements describing skills across eight developmental areas that parents are likely to observe in daily interactions: social, self-help, gross motor, fine motor, expressive language, language comprehension, letters, and numbers. The CDI provides information about the child's current development, their weaknesses and also their strengths. Parents' scores correlate well with standardized tests such as Clinical Adaptive Test/Clinical Linguistic Ability Milestone Scale (Accardo & Capute, 1996) and the Bayley Scales of Infant Development 2nd Edition (Bayley, 1993; Doig, Macias, Saylor, Craver, & Ingram, 1999). However, the number of items in the CDI makes its completion a lengthy process, restricting its utility.

The Infant Development Inventory (IDI; Ireton, 1994) was developed by Ireton to cover the period from birth to 21 months of age. Parents describe their infant in terms of current behaviors using an Infant Development Chart, describing patterns of behavior expected across the five domains of fine motor, gross motor, social, self-help and language. The IDI is a descriptive tool which health professionals ask concerned caregivers to complete—at-risk infants are referred for subsequent assessment. The IDI has yet to be validated against the BSID or other standardized test.

The Parent Report of Children's Abilities (PARCA; Saudino et al., 1998) assesses non-verbal cognitive ability in 2-year-old children and takes approximately

one hour to complete. The PARCA is divided into two parts: a parent-report section and a parent-administered section. The parent-report section comprises 26 yes-or-no items assessing quantitative skills, spatial abilities, symbolic play, planning & organizing, adaptive behaviors, and memory. The parent-administered section comprises 24 items assessing design drawing, match to sample, block building and imitation. These 24 test items were drawn from existing measures of cognitive ability e.g. the Bayley Scales of Infant Development (Bayley, 1969) and the McCarthy Scales of Children's Abilities (McCarthy, 1972). The instrument is not intended for use across a range of ages.

The Ages and Stages Questionnaires (ASQ; Bricker et al., 1995) is a screening tool which identifies infants and young children at risk of developmental delay between 4 months and 5 years. It is quick to complete (10-15 minutes), and covers a much wider range of domains than merely cognition (communication, gross motor, fine motor, problem solving, and person-social). The ASQ comprises 30-item questionnaires for use at designated ages e.g., 4, 6, 8, 10 and 12 months. Responses are scored to determine whether the child is at high or low risk of developmental delay in each domain. This tool has not been standardized against any existing test although its screening properties have been assessed. Additionally, Johnson et al. (2004) validated a parent report measure of cognitive development for use with very preterm infants, but only at 24 months.

Thus, the only parent report instruments to assess cognitive ability in a range of ages under 2 years are the ASQ (Bricker, Squires, & Mounts, 1995), the CDI (Ireton, 1998), and the IDI (Ireton, 1994). The Child and Infant Developmental Inventories measure cognition only within the language domain, whilst the ASQ measures the cognitive domains of communication and problem solving.

A parentally-reported instrument based on a wide range of cognitive constructs, which was easy to administer and score, and was reliable, would clearly benefit both research and clinical communities. If the instrument afforded parents the opportunity to observe patterns of behavior first hand, then the instrument would likely be more accurate than relying on memory of previous behaviors alone. It would provide researchers with a cost-effective means of examining relationships between infants' cognitive ability and their performance in other domains. In the longer term, it might also be useful as a screen for cognitive delay, or as a means of assessment and follow up. We now describe the development of such an instrument, termed the Cognitive Development Questionnaire ('CDQ').

We focused on the ages 8 to 24 months—a period of enormous change in cognitive performance, and arguably when delay might first be observed by parents. Although our aim was not to develop a test with subscales, adopting the guidelines of Dale et al. (1989), we aimed to develop an instrument which would measure non-verbal cognition under the broad constructs of: Objects, People, Communication, Imitation, Problem Solving, and Categorization.

Development of the CDQ took place in 3 stages: (1) We piloted our draft CDQ1 with a volunteer sample, to establish item correlations with age, and allow problematic items to be removed; (2) a revised CDQ2 was tested on a large postal sample, modified to give CDQ3, and performance of this revised instrument was compared with the Mental Scale of the Bayley Scales of Infant Development (BSID-II[3]) (Bayley, 1969, 1993); (3) inter-rater effects were assessed using CDQ3.

Phase 1: Pilot postal phase – CDQ1[2]

Method

Participants

A total of 35 families (16 with girls) participated. Most participants came from white, middle class households. Children ranged in age between 8 and 24 months (mean 16.6 months, SD 4.8 months).

CDQ1 Content

We drafted CDQ1 with 2 sections: *Games* (somewhat analogous to activities in the Mental Scale of BSID-II), and *Questionnaire* (analogous to the approach in Fenson et al., 1994, and in the PARCA: Saudino et al., 1998). The *Games* section consisted of 35 scripted games, for parents to play with their infant, and then to respond to a statement with a simple yes or no, or a simple piece of information (e.g., "number of blocks used"). The *Questionnaire* section consisted of 28 yes-or-no statements about the infant's everyday behavior. Test items were designed to increase progressively in difficulty, cognitive competence and age. In the manner of development of the PARCA scale (Saudino et al., 1998), test items were either drawn from existing instruments and modified for use in the home, or new items written specifically for the CDQ. Test items were modified from: the Ages and Stages Questionnaire (Bricker et al., 1995), the Bayley Scales of Infant Development II (Bayley, 1993), and the Cognitive Ability Questionnaire (Alcock, 2002). The CDQ was designed to take no more than two hours in total. Parents were told that it was not necessary to complete the CDQ in a single session, but were asked to take no longer than one week over it.

Vocabulary measures

Parents completed a modified version of the MacArthur Communicative Inventory (Fenson et al., 1994) adapted for British English, containing 662 words, but no gesture or grammar sections. Unlike the original, our instrument asked parents to report on comprehension as well as production to 24 months.

Results

CDQ1 data were collected for all 35 infants, along with measures of vocabulary for 32 infants. Because some items in the CDQ1 contained more than one question, the maximum total score was 72, being 44 for the *Games* section and 28 for the *Questionnaire* section. Total CDQ1 scores in the sample ranged from 15 to 69 with a mean of 43.8 (SD=15.4).

We used the Kuder-Richardson Reliability test coefficient (KR20) to assess the extent to which our dichotomous data were associated (Anastassi & Urbina, 1997; Loewenthal, 2001). Across the 63 dichotomous items, the KR20 statistic was 0.95, indicating high internal consistency. The KR20 statistic for the *Games* section alone was 0.91 and for the *Questionnaire* section was 0.89. Recall that we had constructed CDQ1 with six constructs in mind, as listed in the Introduction. Accordingly, because each of these was tested by a specific subset of the items, KR20 coefficients were calculated for each. A reliability coefficient between 0.70 and 0.80 is an appropriate level of reliability in standardized test construction (Anastassi & Urbina, 1997; Kline, 1993; Loewenthal, 2001). All constructs except that for Objects showed acceptable reliability. Reliability for this construct was increased to .75 by deletion of four items from the *Games* section and two from the *Questionnaire* section. Further analysis in

this section refers to this reduced item set. Following these deletions sub-total scores for the *Games* and *Questionnaire* sections were positively correlated r (35) = .84, p < 0.001.

Across the range of infants piloted, CDQ1 total scores showed a good spread of scores, with a positive roughly linear relation between score and age, r (35) = 0.85, p < .001 (two-tailed), see Figure 1. Mean receptive vocabularies were larger than productive vocabularies (175 and 69 respectively). Spearman correlations between comprehension and production vocabularies and total CDQ1 scores were .86 and .81 respectively (both ps < .01).

Parents were asked for comments on the ease of administration, and infants' levels of enjoyment, of CDQ1. Twenty-two had additional comments. Overall parents reported that CDQ1 was easy to administer and children cooperated well: if their infant was uncooperative with a particular item, they moved on and returned to it another time. Some parents commented that CDQ1 provided them with ideas of games to play with their child. In line with suggestions from parents about instructions, we made some minor changes to the wording of CDQ1.

Phase 2: Concurrent Validity of CDQ2 and generation of CDQ3

In this phase, we sought to refine the content of the CDQ by postal administration of CDQ2 to a larger sample than that of Phase 1. Statistical analysis of items then gave rise to a revision, CDQ3, whose performance was validated by comparison with a standard instrument for assessment of cognitive development (the Mental Scale of the Bayley Scales of Infants Development (Bayley, 1993).

Method

Following modification from Phase 1, CDQ2 consisted of 50 test items: 28 in *Games* and 22 in the *Questionnaire* section, each covering our six theoretical constructs. We sent CDQ2 to parents of 8- to 24-month-olds (greater than 37 weeks' gestation and no reported hearing or speech problems) to complete, using a laboratory database. Of 225 CDQ2s sent out, 162 (81 girls) were returned (140 complete), and 64 infants (33 girls) were recruited for administration of the Mental Scale of the Bayley Scales of Infants Development (BSID-II) (see below, and Table 2). Most participants came from white, middle class households. Administration of the standardized test instrument was video recorded.

INSERT TABLE 2 ABOUT HERE

Results

Maximum total score for CDQ2 was 50 (28 for *Games* and 22 for *Questionnaire*). The mean score in our sample was 30.9 (SD 9.4), range 9 – 47 (2.5 – 25.5 for *Games* and 5 – 22 for *Questionnaire*). Scores for the *Games* section and the *Questionnaire* were well correlated, r (162) = 0.78, p < .001. Reliability analyses were conducted as before. KR20 for all test items in CDQ2 was .93 (*Games* .91; *Questionnaire* .84).

Again, we examined subsets of items, which had been designed to tap the cognitive constructs listed in the Introduction. KR20 scores for these fell within the range .74 - 0.83, except those for the Person concept construct; the two items which comprised this construct were deleted and are omitted from subsequent analysis. We then examined internal reliability of individual items for each of the five remaining

theoretical constructs using Cronbach's alpha on item total. Ten further items were removed as a result (*Games* 4, *Questionnaire* 6). The resulting instrument, consisting of 24 *Games* and 16 *Questionnaire* items, we term CDQ3: all further analysis refers to this instrument. The reliability coefficient for CDQ3 was .93 (.92 for the *Games* section, and .84 for the *Questionnaire*). There was a strong positive relationship between age at testing and total CDQ3 scores, r (162) = .82, p < .001; these data are shown in Figure 2.

INSERT FIGURE 2 ABOUT HERE

The CDQ is designed to assess cognitive ability across a broad range of ages, and should therefore demonstrate an age-related range of response rates around 50% (Anastassi & Urbina, 1997). We calculated the percentage of infants who successfully completed each test item in each of three broad ranges: 8 to 13 months, 14 to 17 months and 18 to 24 months: see Figure 3. Inspection of Figure 3 suggests that we have been successful in generating a series of test items suitable for a range of ages.

INSERT FIGURE 3 ABOUT HERE

We investigated concurrent validity of CDQ3 against the Mental Scale of the Bayley Scales of Infant Development ('BSID-II' - Bayley, 1993), a widely-accepted standardized, assessor-administered assessment. Sixty-four infants (33 girls) were administered relevant age appropriate test items from the Mental Scale of the BSID-II within one week of parent completion of the CDQs. Two researchers (the first author and a colleague) conducted the assessments: one researcher tested 15 infants, the other 49. To establish inter-scorer reliability assessments for eight infants were scored by both researchers (with the second researcher scoring a video recording of the initial test session). These raw scores were highly and significantly correlated, _ (8) = 0.96.

12

Total CDQ3 scores and raw Bayley Mental Scale scores were well correlated, _ (64) = 0.88, $p < .001$, see Figure 4.

INSERT FIGURE 4 ABOUT HERE

To assess whether the CDQ is sensitive to relative level of development at a given age, the relationship between Bayley scores and total CDQ3 scores was examined with age partialled out. The significant partial correlation, r (61) = .40, $p <$ .001, suggests that the relationship between the two instruments is not simply based on age, but accurately reflects the cognitive abilities of infants of specific ages. Saudino et al. (1998) note that the different components (games, questionnaire items) of a test should each make a unique contribution to the measurement of cognitive ability. Controlling for *Games*, CDQ *Questionnaire* scores were well correlated with Bayley scores, r (61) = .36, $p = .004$; controlling for *Questionnaire*, *Games* scores were also well correlated with Bayley score, r (61) = .60, $p < .001$.

Vocabulary data were obtained for 115 infants, as described above. Mean receptive vocabulary was 197 words (range 0 – 662) SD = 159); mean productive vocabulary was 81 words (range 0 – 662) SD = 114). Both vocabularies correlated strongly with total CDQ3 scores, r (115) = .89, .87 respectively, ps < .001.

Phase 3: Inter-rater and test-retest reliability in CDQ3

In this final phase, our main purpose was to check the internal reliability of the instrument. We used a counterbalanced ABA design, in which CDQ3 was administered to two groups either (a) first by a parent, then by a researcher (the fourth author) and then by the parent again (the PXP group), or (b) first by the researcher,

then by the parent, and then the researcher (the XPX group). We also made some small textual changes to clarify some items, and replaced four images with copyright-free ones.

Method

Participants

Thirty parents with children aged 11 to 24 months participated, with roughly even age distribution, again from mostly white middle-class households. Of the 30 participating families, 25 completed two administrations (mean age = 16.9 months, SD= 4.1), and 18 completed all three administrations (mean age = 18.1 months, SD= 3.8).

Procedure

Parents completed the CDQ in the home within a week. The target interval between each administration was around a week (making two weeks target between T1 and T3; actual T1-T2: 0-17 days; T1-T3: 3-29 days). One item in the *Games* section required a delay of 24 hours for completion, and all but 5 items in the *Questionnaire* section asked about habitual behavior of the child. It was therefore not possible for the researcher to complete these items, and they are not included in the inter-rater statistics (but they contribute to test-retest calculations).

Results

To estimate inter-rater reliability, two-way mixed intraclass correlation coefficients (ICCs) were computed for the CDQ3 total scores of consecutive administrations of CDQ3, using the absolute agreement method (see Shrout & Fleiss, 1979; McGraw & Wong, 1996). ICCs between first and second administration, and between second and third, were .79 and .83, respectively ($ps < .001$) (i.e., in the 'excellent' range:

Cicchetti, 1994). The achieved power of both analyses was 0.99 (Faul, Erdfelder, Lang & Buchner, 2007).

Test-retest reliability was calculated using the first and third administrations of CDQ3, thus keeping administrator constant for each infant. Overall reliability, again by two-way mixed absolute agreement ICCs was excellent at .93, $p < .001$. Scores at third administration were slightly greater than at first administration, but not significantly so, $t(18) = 1.35$, $p = .194$.

We calculated inter-rater and test-retest reliabilities of the *Games* and *Questionnaire* sections separately. The *Games* section had good inter-rater reliability (ICC = .73, $p < .001$) and excellent test-retest reliability (ICC = .88, $p < .001$), with high levels of power (.99 for both analyses). The *Questionnaire* section also showed good inter-rater reliability (.70, $p < .001$) and excellent test-retest reliability (.99, $p < .001$), with achieved power of .98 and 1 respectively.

To examine how similarly parents and the researcher scored individual infants, we conducted a 2 (Group) X 3 (Administration) mixed model ANOVA, on CDQ3 scores with Greenhouse-Geisser correction. Neither main effect was significant, but there was a Group X Administration interaction, $F(1.9, 30.3) = 8.94$, $p = .001$. There were no effects of Group at any Administration, but parental scores were higher than researcher scores in both the XPX group and, less strongly, the PXP group ($p = .001$ and .093, respectively, by planned F tests).

Whilst the data collected represents a comparatively small sample, it was our intention to assess the internal reliability of the instrument rather than its validity. A smaller sample was therefore adequate to our purposes.

General Discussion

We set out to establish whether a questionnaire designed to report cognitive development in the period 10 to 24 months could be successfully administered by parents, and would be found to have good face, construct, and internal validity. We found that parents can indeed administer such a measure; furthermore, total scores on the instrument show good concurrent validity, by virtue of a strong positive relationships between infants' total CDQ scores and their raw score on the Mental Scale of BSID-II (Bayley, 1993), and a roughly linear relationship with age. The *Games* and the *Questionnaire* sections each made independent contributions to CDQ validity.

Our findings both substantiate the validity of the CDQ measure, and support the use of parental report for the assessment of infant ability in general. Parental report measures are often overlooked, perhaps because some studies have shown that parents can overestimate their infant's ability (e.g., Miller, 1986; Miller, Manhal, & Mee, 1991). However, such studies have typically asked parents to predict how their infants may perform on a particular task, rather than directly comparing parent and researcher reports of directly observed ability. Researchers who have made direct comparisons between parent- and researcher- administered test instruments have found high levels of agreement (e.g., Beckman, 1984; Blacher-Dixon & Simeonsson, 1981; Bricker & Squires, 1989; Bricker et al., 1988; Johnson et al., 2004).

One important difference between our CDQ measure of infant cognitive ability and other parental assessment tools such as Ages and Stages (Bricker et al., 1995) and the Infant Development Inventory (Ireton, 1994) is that the CDQ makes the parent the assessor by employing the *Games* section test items, allowing parents to

directly observe the behaviors in question rather than relying entirely on parents' memory of infant behavior which may not have been placed within a suitable context, or their predictions of infants' performance in a hypothetical situation. This feature presumably reduces susceptibility of the instrument to parental overestimation of the kind described by Miller and colleagues (Miller, 1986; Miller et al., 1991).

The five constructs (Objects, People, Communication, Imitation, Problem Solving, and Categorization) included in the final modified CDQ demonstrated high face and construct validity in their good individual correlations with age and Bayley scores. Taken as a whole, our data suggest that test items cumulate well to measure infants' overall cognitive ability. We are confident that our constructs did what we wanted them to, that is, aid us in designing a well-rounded cognitive assessment. We make no claims for subsets of CDQ item scores: We would not want to place authoritative-sounding labels on subsets of this instrument at this stage. However we note that it may be possible to use them in the future (see for instance Alcock & Krawczyk, 2010). Therefore, currently, the CDQ measure outcome is simply the total score[4].

Our aim to design a parental report cognitive development questionnaire, accessible to researchers and clinicians, has proved successful. The next task in the development of our Cognitive Development Questionnaire might be to standardize test scores at a range of specific ages by means of a large-scale validation study. The standardization process should draw on a more diverse infant sample; ensuring representations are made across a range of demographic variables. Nonetheless, CDQ3 already is a potential alternative instrument for researchers or clinicians

interested in gathering cognitive ability data from large samples of infants in a cost effective manner. As such it forms a viable alternative to such instruments as the Bayley Scales of Infant and Toddler Development (Bayley, 2005: but see Footnote 3).

A further development of the CDQ instrument would be to examine whether it can successfully detect cognitive delay. Its ease of administration clearly makes it a potentially powerful tool for screening or surveillance of at-risk infants. More research is required to establish the sensitivity and specificity of the CDQ in this context.

Overall, the CDQ represents not simply a cost-effective and time-efficient method of assessing non-verbal cognitive development in infancy. It has also been shown to have good measurement properties and, importantly, good validity when compared to existing measures. It is often implied that direct laboratory-administered testing is the "gold standard" and that parent-administered testing or parent report measures are adequate where laboratory testing cannot take place. However, in the case of infant cognition, parent questionnaires may be superior, for the same reasons as apply for instance in the case of infant language (Fenson et al., 1994). Parents can sample a much larger range of children's behavior than can assessors. Importantly in the case of the CDQ with its directly administered games, children may also cooperate in testing much more readily with a parent than with an unfamiliar tester. Thus the CDQ has great promise in experimental as well as clinical and screening settings.

Footnotes

1. The term 'parent' may at all times be read as 'parent or caregiver'.

2. Where the version of the CDQ is material, it will be given a numeric suffix.

3. The latest version of the Bayley Scales is the 3rd Edition ('BSID-III': Bayley, 2006). Part of the research described in the article was done before BSID-III became widely available. Throughout the research, we therefore used Bayley II. We have no reason to believe that the advent of a new version of Bayley changes the substantive conclusions of this report (but see Moore, Johnson, Haider, Hennessy, & Marlow, 2011).

4. The latest version of the CDQ, and details of how to score it, will be available copyright-free by contacting Dr XXX at xxx@xxx. Permission will be granted for free use so long as the authors can be reassured that such use will be appropriate and not for profit. [Note to Editor: Supplementary materials will be made available for publication on the journal website if requested, with the proviso that we would rather retain control of the CDQ to ourselves by making it available to those who ask for it. Otherwise, for example with publication on line—past experience suggests that that corrupted, non-standard versions will emerge].

5.

References

Accardo, P. J., & Capute, A. J. (1996). *The Capute Scales: Cognitive Adaptive Test / Clinical Linguistic and Auditory Milestone Scale*: Baltimore MD: Brookes Publishing Company.

Achenbach, T. M., & Edelbrock, C. (1983). *Manual for Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, VT: Achenbach.

Alcock, K. (2002). Cognitive Ability Questionnaire: Unpublished manuscript.

Alcock, K. J., & Krawczyk, K. (2010). Individual differences in language development: Relationship with motor skill at 21 months. *Developmental Science, 13*(5), 677-691.

Anastassi, A., & Urbina, S. (1997). *Psychological Testing. Seventh Edition*. Upper Saddle River, New Jersey: Prentice Hall.

Bayley, N. (1969). *Bayley Scales of Infant Development.* New York,NY: The Psychological Corporation.

Bayley, N. (1993). *Bayley Scales of Infant Development, 2nd ed.*. San Antonio, TX: Psychological Corporation.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development*. San Antonio, TX: Harcourt Assessment, Inc.

Beckman, P. J. (1984). Perceptions of children with handicaps: A comparison of mothers and program staff. *Mental Retardation, 22*, 176-181.

Blacher-Dixon, J., & Simeonsson, R. J. (1981). Consistency and correspondence of mothers' and teachers' assessments of young handicapped children. *Journal of the Division for Early Childhood, 3*, 64-71.

Bradley-Johnson, S., & Johnson, C.M. (2001). Cognitive Abilities Scale- Second Edition. Sutin, TX: Pro-Ed

Bricker, D., & Squires, J. (1989). The Effectiveness of parental screening of at-risk infants: The infant monitoring questionnaires. *Topics in Early Childhood Special Education, 9*(3), 67-85.

Bricker, D., Squires, J., Kaminski, R., & Mounts, L. (1988). The validity, reliability and cost of a parent-completed questionnaire system to evaluate at-risk infants. *Journal of Pediatric Psychology, 13*, 55-68.

Bricker, D., Squires, J., & Mounts, L. (1995). *Ages and Stages Questionnaires: A Parent-Completed, Child Monitoring System.* Baltimore, MS: Brookes.

Cattell, P. (1940). *Cattell Infants Intelligence Scale*: The Psychological Corporation.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Dale, P. S., Bates, E., Reznick, S. J., & Morisset, C. (1989). The validity of a parental report instrument of child language at twenty months. *Journal of Child Language, 16*, 239-249.

Dinnebeil, L. A., & Rule, S. (1994). Congruence between parents' and professionals' judgments about the development of young children with disabilities: a review of the literature. *Topics in Early Childhood Special Education, 14*(1), 1-25.

Doig, K. D., Macias, M. M., Saylor, C. F., Craver, J. R., & Ingram, P. E. (1999). The Child Development Inventory: A developmental outcome measure for follow-up of the high-risk infant. *Journal of Pediatrics 135*(3), 358-362.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). The MacArthur Communicative Development Inventories: user's guide and technical manual. *San Diego: Singular Publishing Group*.

Fenson, L., Dale, P. S., Reznick, S. J., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial No.242).

Franenburg, W. K., Didds, J. B., Fandal, A. W., Kazuk, E., & Cohrs, M. (1975). *Denver Developmental Screening Test*. Denver: Denver Developmental Materials.

Glascoe, F. P., Altemeier, M. D., & MacLean, W. E. (1989). The importance of parents' concerns about their child's development. *Archives of Pediatrics and Adolescent Medicine, 143*(8), 955-958.

Glascoe, F. P., Byrne, K. E., Ashford, L. G., Johnson, K. L., Chang, B., & Strickland, B. (1992). Accuracy of the Denver-II in Developmental Screening. *Pediatrics, 89*(6), 1221-1225.

Goldsmith, H. H. (1996). Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Development, 67*, 218-235.

Gradel, K., Thompson, M., & Sheehan, R. (1981). Parental and professional agreement in early childhood assessment. *Topics in Early Childhood Special Education, 1*(2), 31-39.

Griffiths, R. (1996). *Griffiths Mental Development Scales*. Henley: The Test Agency.

Hogan, A. E., Scott, K. G., & Bauer, C. R. (1992). The Adaptive Social Behavior Inventory (ASBI): A new assessment of social competence in high-risk three-year-olds. *Journal of Psychoeducational Assessment, 10*, 230-239.

Ireton, H. (1994). *Infant Development Inventory*. Minneapolis: Behavior Science Systems Inc.

Ireton, H. (1998). *Child Development Inventory*. Minneapolis: Behavior Science Systems Inc.

Johnson, S., Marlow, N., Wolke, D., Davidson, L., Marston, L., O'Hare, Peacock, J., et al. (2004). Validation of a parent report measure of cognitive development in very preterm infants. *Developmental Medicine & Child Neurology, 46*(6), 389-397

Kline, P. (1993). *The handbook of psychological testing.* London: Routledge.

Knobloch, J., Stevens, F., & Malone, A. F. (1980). *Manual of Developmental Diagnosis: Revised Gesell Developmental Schedules*. Hagerstown, M.D: Harper and Row.

Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales. 2nd Edition*. Hove, East Sussex: Psychology Press Limited.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities.* New York: The Psychological Corporation.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Miller, S. C. (1986). Parents' beliefs about their children's cognitive abilities. *Developmental Psychology, 22*(2), 276-284.

Miller, S. C., Manhal, M., & Mee, L. L. (1991). Parental Beliefs, Parental Accuracy, and Children's Cognitive Performance: A search for Causal Relations. *Developmental Psychology, 27*(2), 267-276.

Moore, T., Johnson, S., Haider, S., Hennessy, E., & Marlow, N. (2011). The Bayley-III cognitive and language scales: How do scores relate to the Bayley II? *Archives of Disease in Childhood, 96*(Suppl 1), A39.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*, (1-2 Serial No. 149).

Newborg, J., Stock, J., & Wnek, L. (1984). *Battelle Developmental Inventory*: Rolling Meadows, IL: Riverside Publishing Company.

Oliver, B., Dale, P., Saudino, K., Petrill, S., Pike, A., & Plomin, R. (2002). The validity of a parent-based assessment of cognitive abilities in three-year-olds. *Early Child Development and Care, 172*(4), 337-348.

Provence, S., Erikson, J., Vater, S., & Palmeri, S. (1995). *Infant-Toddler Developmental Assessment: IDA*. Chicago, IL: Riverside Publishing.

Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., et al. (1998). The validity of parent-based assessment of the cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology, 16*, 349-363.

Sheehan, R. (1988). Involvement of parents in early childhood assessment. In R. Sheehan (Ed.), *Assessment of Young Developmentally Disabled Children*. New York: Plenum.

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

Uzgiris, I. C., & Hunt, J. M. (1975). *Assessment in Infancy: Ordinal Scales of Psychological Development*. Urbana: University of Illinois Press.

Waters, E., & Deane, K. E. (1985). Defining and assessing individual differences in infant attachment relationships: Q methodology and the organization of behavior. In E. Waters (Ed.), *Growing Points of Attachment Theory and Research* (Vol. 50, pp. 41-65): Monographs of the Society of Research in Child Development.