Kamppi, Dorian & Gilmore, Linda (2010) ***Assessing cognitive development in early childhood : a comparison of the Bayley-III and the Stanford-Binet fifth edition.*** Australian Educational and Developmental Psychologist, 27(2), pp. 70-75.

Author copy of the paper published in:
The Australian Educational and Developmental Psychologist, 27(2), 2010.

Assessing Cognitive Development in Early Childhood: A Comparison of the Bayley-III and the Stanford-Binet Fifth Edition

Dorian Kamppi and Linda Gilmore

Queensland University of Technology

Address correspondence to the second author at:

School of Learning and Professional Studies
Queensland University of Technology
Victoria Park Road
Kelvin Grove QLD 4059 Australia
E-mail l.gilmore@qut.edu.au

Keywords: Bayley Scales of Infant and Toddler Development; Bayley-III; Stanford-Binet; SB5; cognitive assessment

Abstract

The Bayley Scales of Infant Development, Third Edition (Bayley-III) and Stanford-Binet Intelligence Scale, Fifth Edition (SB5) were administered in a sample of 26 typically developing children (12 males and 14 females) aged 24 – 42 months. Children completed the assessments in two separate sessions, counterbalanced for order of administration. Scores on the two instruments were not significantly related, with the exception of the SB5 Knowledge score, which was moderately correlated with the Language score on the Bayley-III (r = .41, p = 0.04). Despite no other significant correlations, for 22 of the 26 children, scores were very consistent across the two instruments. Implications for test selection are discussed.

One of the decisions facing psychologists who assess cognitive functioning is which test to choose when two or more similar instruments are available and appropriate for the age group being assessed. This decision usually depends on factors such as the technical qualities reported in test manuals and the appropriateness of available norms. Published reports of relevant research are also useful, with comparative studies that report the results of concurrent administrations of similar instruments being of particular value.

In the early childhood years, a major task for educational and developmental psychologists is the assessment of children with developmental delay, and cognitive functioning is an important part of comprehensive assessment. There is often a peak in referrals around 2 years of age, when language and other milestones are not achieved age-appropriately. Two instruments used for assessing cognitive functioning in early childhood are the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III; Bayley, 2006) and the Stanford-Binet Scales of Intelligence, Fifth Edition (SB5; Roid, 2003). These two instruments overlap in the age range 2 years to 3 years 6 months.

A major difference between the Bayley-III and the SB5 is the fact that the Bayley is a developmental assessment, rather than an intelligence test. While the SB5 yields Full Scale, Verbal and Nonverbal IQ scores, the Bayley-III provides information about a broader range of cognitive, language, motor, social-emotional and adaptive skills. Historically, however, it has not been uncommon for the Bayley's cognitive index to be interpreted as a measure of intelligence (see, for example, Blaga et al., 2008; Molfese & Acheson, 1997; Ramey & Haskins, 1981).

To date, no comparative studies of the current editions of the Bayley and Stanford-Binet have been published. With the exception of one study that found a

significant concurrent relationship between earlier editions of the Bayley and Stanford-Binet (Fagen, Ohr, Fleckenstein, & Singer, 1987, April), research using these two instruments has focused on correlations between scores on the Bayley and performance one to two years later on the Stanford-Binet. Significant relationships have been found between Bayley scores at 2 years of age and the Stanford-Binet at age 3 for both typically developing children (Lewis, Jaskir, & Enright, 1986; Molfese & Acheson, 1997) and atypical samples, where greater stability of scores has been reported (e.g., Dezoete, MacArthur, & Tuck, 2003; Maisto & German, 1986).

As there are virtually no published comparisons of concurrent administrations of the Bayley Scales and Stanford-Binet, practitioners have no evidence base for decisions about instrument choice. The present study aimed to provide preliminary evidence about the relationships between Bayley-III and SB5 scores by administering the two instruments to a sample of typically developing children in the age range where the two instruments overlap – that is, 2 to 3 ½ years of age.

Method

Participants

Thirty-three children were recruited from childcare centres and playgroups. One participant was excluded because of significant speech delays, and six withdrew from the study for reasons that included illness of the child or parent, or the family's inability to attend the second testing session. The final sample comprised 26 children (14 girls and 12 boys) ranging in age from 24 to 42 months (M = 34.54; SD = 4.50).

Instruments

The **Bayley Scales of Infant and Toddler Development, Third Edition** (Bayley-III) (Bayley, 2006) is the most recent version of a well-established developmental test. The instrument was designed for the age range 16 days to 42

months 15 days, and was normed on 1,700 children in the USA. The subtests combine to form several scales – Cognitive, Language (Receptive and Expressive), Motor (Fine and Gross), Social-Emotional, and Adaptive Behaviour – thereby providing a measure of general development. The Bayley-III's technical manual provides evidence of good reliability and validity for the preschool age range. For 24 to 42 month old children, internal consistency using the split-half method ranged from .92 to .97 for the Cognitive Scale, from .93 to .97 for the Language Scales, and from .87 to .95 for the Motor Scales. Test-retest reliability showed a high degree of stability over time with a slight increase in stability across age groups.

The **Stanford-Binet Intelligence Scales, Fifth Edition** (SB5) (Roid, 2003) is an individually administered test that is used to assess the general thinking and reasoning abilities of people aged 2 to 85+ years. Normed on 4,800 individuals in the USA, the instrument comprises five cognitive factors - Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing and Working Memory, in both verbal and nonverbal domains. Verbal IQ (VIQ) and Nonverbal IQ (NVIQ) scores are combined to produce a Full Scale IQ (FSIQ). For preschoolers, reliability coefficients ranged from .79 to .98. Test-retest reliability analyses for ages 2 to 5 showed a high degree of stability over time.

Post-assessment ratings were completed by the examiner after each test administration. Items were set on a 5-point scale, and were derived from the Behaviour and Attitude Checklist (Sattler, 2001) and the Behaviour Rating Scale from the Bayley Scales of Infant Development, 2nd Edition. The items covered ease of administration, testing environment, rapport, child responsiveness, cooperativeness, interest in test materials, enthusiasm, attention, persistence and frustration.

Procedure

An intern psychologist (the first author) administered the two tests in a counterbalanced order. For each child, both tests were conducted in the same environment, either the university clinic or the family home, and at a similar time of day. The interval between the two sessions ranged from 1 week to 13 weeks (M = 3.08; SD = 2.99). Standardised testing procedures outlined in each test manual were followed and a parent was present throughout both administrations.

Children completed the Cognitive, Receptive Language, Expressive Language, Fine Motor and Gross Motor subtests of the Bayley-III and all age-appropriate subtests on the SB5. The total administration time, including breaks, ranged from 60 to 110 minutes (M = 86.15; SD = 12.67) for the Bayley-III, and from 40 to 75 minutes (M = 50.19; SD = 10.05) for the SB5.

Results

Correlations Between Bayley-III and SB5 Scores

Table 1 presents correlations of Bayley-III and SB5 composite scores. Analyses using Pearson correlation coefficients indicated that the correlation between the SB5 Full-Scale IQ (M = 103.65, SD = 7.04) and the Bayley-III Cognitive score (M = 100.19, SD = 5.19) was not significant. None of the composite scores on the Bayley-III was significantly correlated with composite scores on the SB5, with the exception of SB5 Knowledge, which was significantly related to Bayley-III Language, r = .41, p = .04.

[Insert Table 1 about here]

Order of Administration

Independent samples t tests were conducted to examine whether test scores were related to order of test administration. Children scored significantly higher on the

SB5 subtest Fluid Reasoning when the SB5 was the first test to be administered, $t(24) = 2.46$, $p = .02$. No other test scores were related to order of presentation.

Length of Time Between Assessments

The interval between the two testing sessions ranged from 1 to 13 weeks ($M = 3.08$; $SD = 2.99$). Due to the extended period between the test administrations for two participants (9 and 13 weeks), partial correlations were conducted to explore time effects. There was little or no change in correlations after controlling for the length of time between assessments.

Gender Differences

Independent samples *t* tests showed that girls scored significantly higher than boys on the Bayley-III Cognitive Standard Score, $t(24) = 2.23$, $p = .04$ ($M = 102.14$, $SD = 5.45$ for girls; $M = 97.92$, $SD = 3.97$ for boys). There were no other significant gender differences.

Post-Assessment Ratings

Paired-sample t-tests were conducted to explore differences and similarities between the post-assessment ratings for the two tests. Results indicated that children displayed significantly more interest in the Bayley-III test materials than in the SB5 test materials, $t(25) = -2.94$, $p = .007$. There were no significant differences between other ratings for the two tests.

Within-Group Differences

Comparisons of individual differences between Bayley-III Cognitive scores and SB5 FSIQ scores revealed that for 22 children (85% of the sample), the scores on the two tests were sufficiently similar to have overlapping confidence intervals at the 95% level (see Table 2). Only four children had differences between the two test scores

which were sufficiently large that the confidence intervals did not overlap. In three of the four cases, the SB5 FSIQ yielded the higher score.

[Insert Table 2 about here]

Discussion

The current study was designed to address an important issue in early childhood assessment, namely which test to choose when assessing the cognitive functioning of children in the age range 2 to 3½ years. On the one hand, correlations of Bayley and Stanford-Binet scores in the current sample are lower than might be expected on the basis of the limited previous research that has been undertaken with earlier editions of the two instruments. Yet on the other hand, the scores of most children are sufficiently similar that their confidence intervals overlap.

Given the different structure and content of the Bayley-III and the SB5, there are few direct comparisons that can be made between subtests or clusters of items. The significant correlation between SB5 Knowledge and Bayley-III Language scores is not surprising since the SB5 verbal knowledge subtest consists of vocabulary items that require young children to name body parts, objects and pictures. Although the nonverbal knowledge items do not require expressive language, an understanding of verbal instructions is needed.

There are some important differences between the Bayley-III and the SB5 that have implications for practitioners' decisions about which instrument to use. First, the Bayley-III provides broader information about a child's developmental skills and abilities which may be of considerable value when making recommendations for early intervention. Second, the greater interest children show in the Bayley-III's toys and brightly coloured materials may contribute to making this the instrument of choice for young children whose participation is more difficult to engage or sustain. On the other

hand, the considerably shorter administration time for the Stanford-Binet (50 minutes on average, compared with 86 minutes for the Bayley) would be an advantage for restless children or when time for assessment was restricted.

This paper has presented what appears to be the first comparative data from the most recent revisions of the Bayley and Stanford-Binet scales. While the findings are likely to have value for practitioners who undertake cognitive assessments in the early childhood years, several limitations must be acknowledged. Because Bayley-III scores are obtained in 5 point intervals, rather than along a continuous scale as used for the SB5, analyses must be viewed with some caution. A notable limitation is the small sample size. Studies examining predictive relationships between the Bayley and Stanford-Binet for typically developing children, have used samples of between 89 and 164 children (Lewis et al., 1986; Molfese & Acheson, 1997). More important than the small sample size, however, is the fact that the children who participated in the study were very average in ability. There was limited variance, with all scores on both measures being within one standard deviation of the mean, and half of the Bayley-III scores being 100. Given that the Bayley-III and SB5 are generally used to assess children whose development is delayed or atypical, future comparative studies with atypical samples are needed to provide the necessary evidence base for professionals who are making decisions about which instruments to use.

References

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

Blaga, O. M., Shaddy, D. J., Anderson, C. J., Kannass, K. N., Little, T. D., & Colombo, J. (2008). Structure and continuity of intellectual development in early childhood. *Intelligence, 37,* 106-113.

Dezoete, J. A., MacArthur, B. A., & Tuck, B. (2003). Prediction of Bayley and Stanford-Binet scores with a group of very low birthweight children. *Child: Care, Health & Development, 29,* 367-372.

Fagen, J. W., Ohr, P. S., Fleckenstein, L. K., & Singer, J. M. (1987, April). *Infant Long-Term Memory for a Conditioned Response and Intelligence Test Performance at 2 Years of Age*. Paper presented at the Biennial meeting of the Society for Research in Child Development, Baltimore. Retrieved from http://search.ebscohost.com./login.asp

Lewis, M., Jaskir, J., & Enright, M. K. (1986). The development of mental abilities in infancy. *Intelligence, 10,* 331-354.

Maisto, A. A., & German, M. L. (1986). Reliability, predictive validity, and interrelationships of early assessment indices used with developmentally delayed infants and children. *Journal of Clinical Child Psychology, 15,* 327.

Molfese, V. J., & Acheson, S. (1997). Infant and preschool mental and verbal abilities: How are infant scores related to preschool scores? *International Journal of Behavioral Development, 20,* 595-607.

Ramey, C. T., & Haskins, R. (1981). The modification of intelligence through early experience. *Intelligence, 5*(1), 5-19.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside Publishing.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego: Jerome M. Sattler.

Table 1

Correlations Between Bayley-III and SB5 Scores

	Bayley-III Composites			SB5 IQ Composites			SB5 Factor Composites				
	Cognitive	Language	Motor	FSIQ	NVIQ	VIQ	FR	KN	QR	VS	WM
SB5											
FSIQ	.314	.308	.183	1.000							
NVIQ	.327	.178	.157	.854**	1.000						
VIQ	.234	.354	.142	.858**	.469*	1.000					
FR	-.050	-.181	-.166	.350	.304	.311	1.000				
KN	.141	.406*	-.041	.209	.191	.174	-.168	1.000			
QR	.205	.158	.202	.741**	.621**	.645**	.276	-.275	1.000		
VS	.199	.117	.216	.523**	.507**	.375	-.247	.179	.182	1.000	
WM	.300	.313	-.060	.654**	.533**	.598**	.224	.092	.478*	.054	1.000
Bayley-III											
Cognitive	1.000										
Language	.539**	1.000									
Motor	.428*	.016	1.000								

Note. FSIQ = Full Scale IQ; NVIQ = Nonverbal IQ; VIQ = Verbal IQ; FR = Fluid Reasoning, KN = Knowledge; QR = Quantitative Reasoning; VS = Visual-spatial Processing; WM = Working Memory.
*$p < .05$. **$p < .01$.

Table 2

Composite Cognitive Scores and Confidence Intervals on the Two Instruments for each Child

Participant	SB5 FSIQ	95% CI	Bayley-III Cognitive	95% CI	CI Overlap
1	97	93 - 101	100	92 - 108	Yes
2	108	104 - 112	100	92 - 108	Yes
3	113	109 - 117	100	92 - 108	No
4	104	100 – 108	100	92 - 108	Yes
5	111	107 -115	100	92 - 108	Yes
6	98	94 -102	100	92 - 108	Yes
7	100	96 -104	100	92 - 108	Yes
8	104	100 – 108	90	83 - 99	No
9	112	108 -116	100	92 - 108	No
10	115	111 – 119	105	97 - 113	Yes
11	114	110 - 118	105	97 - 113	Yes
12	106	102 – 110	95	87 - 103	Yes
13	89	85 – 93	100	92 - 108	Yes
14	110	106 – 114	105	97 - 113	Yes
15	98	94 – 102	100	92 - 108	Yes
16	102	98 – 106	95	87 - 103	Yes
17	112	108 – 116	115	106 - 122	Yes
18	94	90 – 98	95	87 - 103	Yes
19	93	89 – 97	95	87 - 103	Yes
20	96	92 – 100	110	101 - 117	No
21	102	98 – 106	100	92 - 108	Yes
22	108	104 – 112	100	92 - 108	Yes
23	100	96 – 104	105	97 - 113	Yes
24	106	102 – 110	100	92 - 108	Yes
25	100	96 – 104	95	87 - 103	Yes
26	103	99 – 107	95	87 - 103	Yes

Note. FSIQ = Full Scale IQ; CI = Confidence Interval.