Article

Journal of Psychoeducational Assessment
1–14

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0734282918781431
journals.sagepub.com/home/jpa
SAGE

Agreement Between Bayley-III Measurements and WISC-IV Measurements in Typically Developing Children

Johanna Månsson[1,2], Karin Stjernqvist[1], Fredrik Serenius[3,4], Ulrika Ådén[5], and Karin Källén[1]

Abstract

The study aim was to explore the relationship between a developmental assessment at preschool age and an intelligence quotient (IQ) assessment at school age. One hundred sixty-two children were assessed at 2.5 years with the Bayley Scales of Infant and Toddler Development—Third Edition (Bayley-III) and then at 6.5 years with the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV). The Bayley-III Cognitive Index score was the Bayley entity that showed the highest correlation with WISC-IV Full-Scale IQ (FSIQ; r = .41). There was a significant difference between the individual WISC-IV FSIQ and the Bayley-III Cognitive Index scores. Analyses showed an average difference of −4 units and 95% limits of agreement of −18.5 to 26.4 units. A multivariate model identified the Bayley-III Cognitive Index score as the most important predictor for FSIQ and General Ability Index (GAI), respectively, in comparison with demographic factors. The model explained 24% of the total FSIQ variation and 26% of the GAI variation. It was concluded that the Bayley-III measurement was an insufficient predictor of later IQ.

Keywords

developmental assessment, Wechsler tests, intelligence/cognition, assessment, preschool, participants, scale development/testing, measurement

Introduction

Assessments of cognitive ability during infancy and early preschool years are generally regarded an insufficient predictor of later intelligence quotient (IQ) in normal populations (McCall & Carriger, 1993). Possible reasons have been discussed in research and clinical settings. Tests during infancy typically investigate the acquirement of specific developmental milestones. Constructions may build upon extensive and systematic observations of typical development in

[1]Lund University, Sweden
[2]Skåne University Hospital, Lund, Sweden
[3]Uppsala University, Sweden
[4]Umeå University, Sweden
[5]Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden

Corresponding Author:
Johanna Månsson, Department of Psychology, Lund University, Box 213, 221 00 Lund, Sweden.
Email: Johanna.Mansson@psy.lu.se

infants and toddlers. Hence, they rely on the assumption that a particular ability may or may not be present at a certain stage. This distinguishes developmental tests from intelligence tests. The intelligence construct is conceptualized in terms of mental factors, reflecting different facets of intelligence. For example, the Cattell–Horn–Carroll theory (Horn & Cattell, 1966) proposes that a general intelligence factor comprises both abilities shaped by experience (crystallized intelligence), with facets such as verbal comprehension and long-term memory, and abilities less sensitive to prior experience (fluid intelligence), with facets such as logic reasoning and abstract problem solving. Some of the intelligence facets are not yet fully developed in infancy, and therefore, not assessable. Nevertheless, infant paradigms such as habituation, novelty preference, and reaction time have been identified as useful predictors of later intelligence (Domsch, Lohaus, & Thomas, 2009; Ellingsen, 2016). These tasks reflect underlying cognitive functions, such as information processing and memory, which appear to be associated with a person's intelligence.

Normal development comprises discontinuities, that is, changes in development not only as an increment of abilities but as qualitative progress. The developing child continuously gains knowledge and integrates more complex skills. However, the interplay between internal factors (e.g., genetic and hereditary) and external factors (e.g., environmental and socioeconomic) leads to great variability with regard to developmental progress. Tucker-Drob, Briley, and Harden (2013) proposed that genetic influences on cognitive ability are maximized by environmental advantage. Children in families with high socioeconomic status (e.g., high parental educational attainment) have more access to stimulating activities and positive learning experiences. In low status environments, children are less likely to receive such opportunities and experiences. Gene × Socioeconomic status interaction effects on cognitive ability have been identified in infants and young children (Tucker-Drob et al., 2013). Children in social risk homes tend to score lower on IQ tests than peers with high socioeconomic status (von Stumm & Plomin, 2015).

Although cognitive ability is more variable in young childhood, IQ is considered a rather stable feature, given that circumstances are not changed dramatically. A developing person continuously improves his or her vocabulary, and understanding of complex concepts; however, his or her IQ does not usually change in comparison with that of other individuals of the same age (Neisser et al., 1996). This enhances predictability from one time point to another. In a longitudinal study by Deary and Brett (2015), IQ measures were collected at 11 years and then again at 77 years. Although many years apart, robust correlations (r > .70) between individual IQ scores were identified.

Historically, the Bayley Scales of Infant Development have been regarded the gold standard tool for developmental assessment of infants. The first and second editions (Bayley, 1969, 1993) used combined measures of cognitive and language development, expressed as the Mental Developmental Index. They also provided the Psychomotor Developmental Index, measuring early motor functioning. The latest revision, the Bayley Scales of Infant and Toddler Development—Third Edition (Bayley-III), led to substantial changes in the instrument structure (Bayley, 2006). It separates the former Mental Developmental Index into the Cognitive Index and the Language Index. The Cognitive Index is explicitly influenced by information processing theories and research, reflected in items measuring attention and habituation to visual and auditory stimuli, novelty preference, and processing speed. The Language Index contains items documenting social aspects of communication (e.g., the ability to attend to and initiate social routines), word comprehension, and verbal skills. The Motor Index of the Bayley-III is similar to the former Psychomotor Developmental Index. Early items measure the development of motor and sensory systems, reflected in, for example, the infant's manipulation of objects. Later items increasingly demand motor–cognitive integration (e.g., imitating body postures and completing puzzles).

The standardization procedure of the Bayley-III included a heterogeneous sample containing 10% having, or being at risk of, developmental difficulties (e.g., prematurity, cerebral palsy, and Down syndrome), consistent with the general U.S. population. Clinicians and researchers have questioned the inclusion of children with deficiencies in a normative sample, raising the concern that it may have had the unintentional effect of making the norms too forgiving (Johnson, Moore, & Marlow, 2014). This may result in overestimations of abilities, especially when examining children with low functioning levels (Anderson, De Luca, Hutchinson, Roberts, & Doyle, 2010; Vohr et al., 2012).

Like its predecessors, Bayley-III is widely used both in clinical and research settings. Only a few studies so far have investigated the instrument's predictive validity for later cognitive abilities, and the results are conflicting. Spencer-Smith, Spittle, Lee, Doyle, and Anderson (2015) found Bayley-III Cognitive and Language scores at 24 months to be poor predictors of cognitive impairments evaluated by the Differential Ability Scale—Second Edition at the age of 4 years. Results by Bode, D'Eugenio, Mettelman, and Gross (2014) indicated that Bayley-III Cognitive and Language scores strongly predicted IQ at 4 years, as measured by the Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III). Both studies were performed with children born < 30 weeks' gestation. The relationship between developmental test scores and later IQ is usually higher in clinical groups (e.g., in children born preterm), although low predictability has been demonstrated as well (Doyle & Casalaz, 2001; Luttikhuizen dos Santos, de Kieviet, Königs, van Elburg, & Oosterlaan, 2013; McGrath, Wypij, Rappaport, Newburger, & Bellinger, 2004; Tideman, 2007; Voss, Neubauer, Wachtendorf, Verhey, & Kattner, 2007).

Bode et al. (2014) also investigated correlations between Bayley-III cognitive score ($r = .63$), language score ($r = .67$), and IQ scores in their full-term control group. The Bayley-III has been examined in relation to the WPPSI-III, as part of its validation process as well ($n = 57$; 28-42 months of age; Bayley, 2006). The Bayley-III Cognitive score correlated highly with the Full-Scale IQ (FSIQ; $r = .79$). The Bayley-III Language score correlated highest with WPPSI-III Verbal IQ ($r = .83$) and the FSIQ ($r = .82$). The Motor score correlated moderately with the WPPSI-III FSIQ ($r = .55$). However, some argue that it is insufficient only taking into account information regarding the strength of the relationship when assessing comparability between two measures (Giavarina, 2015). The predictive validity of Bayley-III to subsequent cognitive abilities or IQ, in a normal population, is yet to be explored in more detail. Also, more sophisticated analyses of method agreement are needed.

The aim of the present study was to explore the agreement of the Bayley-III assessment at 2.5 years and later the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003) assessment at 6.5 years, with typically developing children born full term in Sweden between 2004 and 2007.

Method

Sample and Participation Selection

Study participants are children included as controls in the national, multidisciplinary project Extremely Preterm Infants in Sweden Study (EXPRESS). The EXPRESS examines survival and developmental outcomes in children born extremely preterm. It includes all children born < 27 gestational weeks between April 1, 2004, and March 31, 2007, in Sweden. For each extremely preterm child included, a control child was selected from the Swedish Medical Birth Registry. Selection criteria for controls were singleton birth at term with 5-min Apgar score > 3, matched by place of domicile, sex, day of birth, and ethnicity. So far, the EXPRESS has conducted follow-up studies at 1 and 2.5 years corrected age and 6.5 years (Fellman et al., 2009; Serenius et al., 2016; Serenius et al., 2013). The present study includes EXPRESS controls only, undergoing both assessments with the Bayley-III at corrected age 2.5 years and the WISC-IV, at age 6.5 years, a total of 162 children. Table 1 illustrates parental and infant characteristics at birth of the

Table 1. Parental and Infant Characteristics at Birth Among 162 Control Children With Available Bayley-III Results at 2.5 Years and Available WISC-IV Results at 6.5 Years.

	Children with available Bayley-III and WISC-IV information *N* = 162
Maternal characteristics	*N* (%)
Maternal age, years, *M* (*SD*)	31.4 (4.6)
Age < 20 years	1 (0.6)
Age ⩾ 35 years	33 (20.4)
Primiparity	75 (46.3)
Smoking	4 (2.5)
Born outside Nordic countries	5 (3.1)
Education ⩽ 9 years	6 (3.7)
Education 10-11 years	13 (8.0)
Education 12-13 years	50 (30.9)
Education 14+ years	81 (50.0)
Paternal characteristics	
Education ⩽ 9 years	7 (4.3)
Education 10-11 years	22 (13.6)
Education 12-13 years	55 (34.0)
Education 14+ years	65 (40.1)
Infant characteristics	
Gestational age, weeks, mean (*SD*)	39.9 (1.2)
37-38 completed weeks	34 (21.0)
39-40 completed weeks	95 (58.6)
41-42 completed weeks	33 (20.4)
Birth weight, grams, *M* (*SD*)	3,613 (482)
< 3,000 g	14 (8.6)
> 4,000 g	39 (24.1)
Birth weight, z scores, mean (*SD*)	0.11 (0.95)
Small for gestational age	2 (1.2)
Large for gestational age	3 (1.9)

Note. WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; Bayley-III = Bayley Scales of Infant and Toddler Development—Third Edition.

participants included. The Regional Ethics Review Board at Lund University, Sweden, approved the present study. Parents provided written informed consent.

Measures

Bayley-III. The Bayley-III is an individually administered instrument assessing developmental functioning (Bayley, 2006). It provides the examiner with measures of cognitive, language, and motor abilities in children between 1 month and 42 months of age. The Language Index contains two subscales measuring receptive and expressive communication, respectively. The Motor Index provides separate assessments for fine motor and gross motor skills. The Bayley-III enables comparisons between a child's performances and same-age peers' through normed scores (index $M = 100$, $SD = 15$). Its main purposes are identifying children with developmental deficits at a very young age, investigating the effects of interventions, and documenting developmental progress over time. To date, the Bayley-III has been adapted and translated to Swedish; however, there are no Swedish norms (Bayley, 2009).

WISC-IV. The WISC-IV is one of the most widely used instruments for measuring intellectual ability in school-aged children (Wechsler, 2003). It rests on the Cattell–Horn–Carroll theory, and the conceptualization of intelligence as a multidimensional construct. The test age range is between 6 years and 16 years 11 months. Compared to its predecessors, the WISC-IV is regarded as being more allied with recent theories on intelligence, with an increased focus on, for example, working memory and processing speed (Tideman, 2007; Wechsler, 2003).

The WISC-IV includes 15 subtests, of which 10 are included in the ordinary test battery, and five are optional. This present study included the following 10 subtests: Similarities, Vocabulary, Comprehension, Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Arithmetic, Coding, and Symbol Search. The subtest Arithmetic is a supplemental Working Memory subtest that replaced the core subtest Letter–Number Sequencing, which requires knowledge of order of magnitude and alphabetic order. Our study cohort was tested at 6.5 years of age. In Sweden, children at this age have just started school. Therefore, knowledge of order of magnitude and/or alphabetic order may not be well-established. Raw scores are converted to scale scores ($M = 10$, $SD = 3$), which are summed up into four indices ($M = 100$, $SD = 15$), each representing distinct intellectual abilities: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), Processing Speed Index (PSI), and one FSIQ. FSIQ represents the sum of scale scores, that is, of all subtests included. Furthermore, the WISC-IV includes the optional General Ability Index (GAI), derived from the VCI and the PRI subtests. The GAI provides the examiner with an estimate of general intellectual capacity, with reduced emphasis on working memory skills and processing speed compared with FSIQ

Subtests included in the VCI (Similarities, Vocabulary, and Comprehension) assess verbal comprehension, verbal concept formation, and verbal reasoning. Subtests included in the PRI (Block Design, Picture Concepts, and Matrix Reasoning) assess nonverbal and logic reasoning, visual–spatial and visual–motor skills. The VCI is, compared with the PRI, more sensitive to the child's ability to draw upon earlier learning experiences (crystallized intelligence). Subtests included in the WMI (Digit Span and Arithmetic) measure working memory, attention, and executive functions. Subtests included in the PSI (Coding and Symbol Search) measure speed of information processing, attention, and differentiation of visual stimuli.

The present study used the Swedish standardized version of the WISC-IV to assess intelligence. This version was translated according to typical procedures and items were culturally adapted. Factor analyses and validation analyses (e.g., of its relation to other Wechsler scales) were thereby performed using the Swedish normal sample, which was based on 290 children between 6:0 and 16:11 years (51% boys), all of which were fluent in Swedish and with no severe sensory impairments, intellectual disabilities, or mental health disorders (Wechsler, 2007).

Procedure

Participating children were tested with Bayley-III and WISC-IV by licensed psychologists at the seven perinatal centers in Sweden (Stockholm, Gothenburg, Uppsala, Lund, Umeå, Linköping, and Örebro). Psychologists were trained in the Bayley-III instrument by one of the authors (K.S.), who also were available for online supervision during both data collection periods, to strengthen assessment reliability.

Statistical Methods

Descriptive background data were presented using numbers and percentage, or using mean and standard deviation (*SD*) when appropriate. The summary statistics for Bayley-III and WISC-IV test results were presented with mean (with 95% confidence interval, CI), median, and the 25th and 75th percentiles, respectively.

Pearson's rho was calculated to describe the pairwise correlations between all Bayley-III scales and all WISC-IV subtests.

Mixed effect analyses were used to study possible differences between boys and girls regarding their individual trajectories (slopes) measured by Bayley-III Cognitive Index scores at 2.5 years, and by WISC-IV FSIQ and GAI at 6.5 years.

The agreement between Bayley-III Cognitive Index scores at 2.5 years and WISC-IV FSIQ at 6.5 years was visualized using a Bland–Altman plot.

Models to predict WISC-IV results were developed using univariate and multivariable linear regression analyses. In the first step, univariate analyses were performed to investigate the impact of maternal age, parity, smoking, country of birth, maternal and paternal educational level, infant birth weight, gestational age, and gender on WISC-IV FSIQ and GAI, respectively. To investigate to which degree the information on Bayley-III scores obtained at 2.5 years contributes to the overall prediction of WISC-IV results 4 years later, two multivariable models were used. The first model included all factors with p values < .2 in the univariate analyses except for Bayley-III Cognitive Index scores. In the full model, information on Bayley-III Cognitive Index scores was also included.

Results

Table 1 shows the parental and infant characteristics of the study cohort. The majority of parents had a high educational level, a low smoking rate, and the number of women who were born outside the Nordic countries was low. The children were all born between 37 and 42 weeks. The rates of children born small or large for their gestational age were low.

Table 2 shows descriptive statistics for the test results for Bayley-III and WISC-IV, by FSIQ and indices, respectively. For Bayley-III, the means for all three indices, and combinations, ranged from 104 to 112, and the *SD* ranged from 10.6 to 14.0. The medians were somewhat lower than the respective means, suggesting a mild right skewness of the score distributions. The mean and the median of the WISC-IV FSIQ were both very close to the stipulated 100, but the *SD* (10.5) was lower than expected (15), and there was a considerable heterogeneity over the WISC-IV subtests. From Table 2, it is evident that the mean score for WMI was more than one *SD* below that for VCI and PRI.

The correlations between all Bayley-III and WISC-IV indices are displayed in Table 3. There were weak positive correlations between most of the WISC-IV and the Bayley-III indices and combination scores. The highest correlation was seen between Bayley-III Cognitive Index (r = .41) and GAI (r = .42), and the lowest between PSI and the combination of Bayley-III Cognitive and Language indices (r = –.02). The Bayley-III Cognitive Index score was the Bayley entity that showed the highest correlation with WISC-IV FSIQ (r = .41).

A mixed effect analysis revealed that there was a significant difference between the individual WISC-IV FSIQ and the Bayley-III Cognitive Index scores (difference = −4.22; 95% CI = [–6.04, –2.43]). There was no difference between the slopes among males and females (p = .77), respectively. When the individual changes between Bayley-III Cognitive Index scores and GAI scores were evaluated, no significant difference was identified (difference = 1.30; 95% CI = [–0.52, 3.12]). The individual changes between Bayley-III Cognitive Index and WISC-IV FSIQ scores are illustrated in a Bland–Altman plot (Figure 1), showing an average difference of −4 units and 95% limits of agreement of −18.5 to 26.4 units.

Tables 4 and 5 show univariate and multivariate models predicting WISC-IV FSIQ at 6.5 years of age. The univariate analyses revealed significant positive linear associations between WISC-IV FSIQ at 6.5 years of age and maternal and paternal educational levels, whereas for gestational age (among these children all born at term), a negative association with WISC-IV FSIQ was detected. The first final multivariable model, not considering Bayley-III Cognitive

Table 2. Summary Statistics, Including Data for Term Controls With Available Data on Both Bayley-III (at 2.5 Years of Age) and WISC-IV Results (at 6.5 Years of Age; N = 162).

	Mean (95% CI)	*SD*	Median	Interquartile range	
Bayley-III					
Cognitive index	104.3 [102.7, 106.0]	10.6	100	95	110
Language index	111.7 [109.8, 113.5]	12.1	109	103	118
Motor index	108.0 [105.8, 110.2]	14.0	107	97	115
Cognitive–language	106.6 [104.6, 108.6]	12.6	106	100	114
Cognitive–language–motor	105.5 [103.6, 107.5]	12.6	106	99	113
WISC-IV					
Verbal Comprehension Index	103.7 [102.0, 105.4]	10.9	102	96	110
Perceptual Reasoning Index	105.4 [103.5, 107.3]	12.3	106	98	112
Working Memory Index	90.8 [89.3, 92.4]	9.8	91	83	97
Processing Speed Index	96.3 [94.3, 98.2]	12.6	97	88	103
General Ability Index	105.8 [104.1, 107.5]	11.2	105	98	111
WISC-IV Full-Scale IQ	100.3 [98.7, 102.0]	10.5	100	94	108

Note. CI = confidence interval; WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; Bayley-III = Bayley Scales of Infant and Toddler Development—Third Edition; IQ = intelligence quotient.

Table 3. Correlations Between Bayley-III and WISC-IV Index Scores.

	Bayley-III scores at 2.5 years				
WISC-IV at 6.5 years (rows)	Cognitive	Language	Motor	Cognitive–language[a]	Cognitive–language–motor[b]
Verbal Comprehension Index	0.32	0.25	0.17	0.26	0.17
Perceptual Reasoning Index	0.39	0.25	0.27	0.23	0.23
Working Memory Index	0.16	0.32	0.13	0.27	0.23
Processing Speed Index	0.18	0.07	0.21	−0.02	0.06
General Ability Index	0.42	0.30	0.26	0.29	0.24
WISC-IV Full-Scale IQ	0.41	0.32	0.29	0.27	0.25

Note. Values represent Pearson's rho. WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; Bayley-III = Bayley Scales of Infant and Toddler Development—Third Edition; IQ = intelligence quotient.
[a]The mean of Cognitive Index and Language Index scores.
[b]The mean of Cognitive Index, Language Index, and Motor Index scores.

Index scores, included factors for which the univariate p values < .2 as well as the p value in the multivariable setting was < .2. The first restricted multivariable model included maternal age (showing a weak nonsignificant positive association with increasing WISC-IV FSIQ scores), maternal educational level (a strong and significant positive association), gestational age (a negative association in this gestational age interval), and gender (significantly lower WISC-IV FSIQ scores in males compared with females). Together, those factors could explain 15% of the WISC-IV FSIQ score variation in the study group. The second final multivariate model included the Bayley-III Cognitive Index score. When information on Bayley-III Cognitive Index score was entered, the p value for child's gender was p = .88, and thus, gender was excluded from the model. The final model, including maternal age and education, gestational age, and Bayley-III Cognitive Index scores could explain 24% of the total WISC-IV FSIQ variation. Besides the Bayley-III Cognitive Index score, maternal educational level was the only significant predictor in the multivariable setting. The rightmost column shows the change of the overall prediction

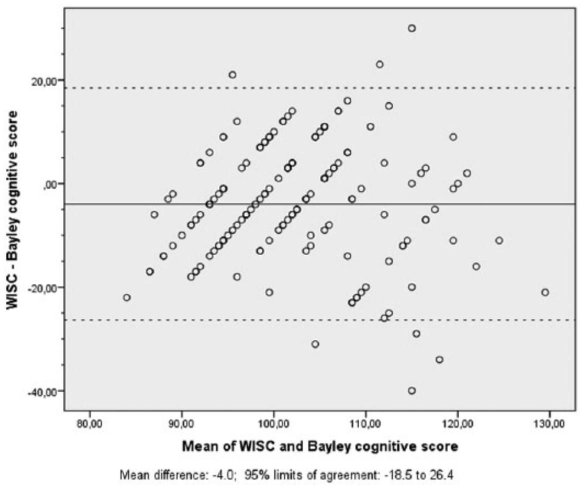

Figure 1. Bland–Altman plot illustrating the individual differences between WISC-IV Full-Scale IQ and Bayley Cognitive Index score.
Note. WISC-IV = Wechsler Intelligence Scale for Children —Fourth Edition; IQ = Intelligence Quotient.

capacity of the model if the factor in question was included. The table shows that, when Bayley-III Cognitive Index score was added to the model, the percentage of explained variance increased from 12% to 24%. Thus, Bayley-III Cognitive Index score was, by far, the most important predictor for WISC-IV FSIQ at 6.5 years of age. The comprehensive WISC-IV FSIQ measure is commonly used in clinical and research settings, although evaluators should consider using the GAI when there are obvious discrepancies between indices (Wechsler, 2003). In our study group, the WISC-IV WMI mean score is remarkably low. The same procedure (as described above) was therefore performed, using the GAI as dependent variable (Table 5). The final model explained 26% of the total GAI score variation. Besides the Bayley-III Cognitive Index score, maternal age and paternal educational level were significant predictors in the multivariable setting. Similar to the WISC-IV FSIQ, the Bayley-III Cognitive Index score was the most important predictor for GAI at 6.5 years of age.

Discussion

The present study investigated the relationship between developmental level at early preschool age and IQ at school age operationalized through a Bayley-III examination at 2.5 years and a WISC-IV assessment at 6.5 years. Contrary to several studies, we were able to analyze this relationship in a cohort of healthy children, born at term.

Our results showed generally weak to moderate associations between Bayley-III scores and WISC-IV scores. A relative strength was recognized between the measures of cognitive ability at 2.5 years and WISC-IV GAI at 6.5 years. Cognitive ability at 2.5 years was the variable with the strongest association with FSIQ at 6.5 years. Our results showed definitely weaker associations between Bayley-III scores and IQ than findings by Bode et al. (2014) and Bayley (2006). It should be noted that they investigated the associations between Bayley-III scores and IQ dimensions within a shorter time interval than we did, using other instruments, although the WPPSI-III, used in the study by Bayley (2006), is clearly related to the WISC-IV. This may explain the differences between our findings. Analysis of agreement showed a mean difference of four units between the Bayley-III Cognitive Index scores and the WISC-IV FSIQ scores. Nevertheless, the

Table 4. Prediction Model for WISC-IV Full-Scale IQ.

	Univariate models			Multivariate model[a], not including Bayley-III scores			Multivariate model[a], including Bayley-III scores			R^2
	Slope	95% CI	*p* value	Slope	95% CI	*p* value	Slope	95% CI	*p* value	R^2 full model—Model without the specified variable
Maternal age per 1 year increment	0.27	[−0.08, 0.62]	.13	0.26	[−0.08, 0.60]	.13	0.23	[−0.9, 0.55]	.16	0.01
Primiparity vs. multiparity	2.0	[−1.22, 5.20]	.22							
Maternal country of birth Non-Nordic vs. Nordic	0.88	[−8.57, 10.33]	.85							
Maternal smoking vs. nonsmoking	3.00	[−4.95, 10.94]	.46							
Maternal education per one-level increment	2.36	[1.18, 3.55]	< .001	2.16	[0.98, 3.34]	< .001	1.83	[0.71, 2.95]	.002	0.05
Paternal education per one-level increment	1.53	[0.37, 2.69]	.01							
Birth weight per 100 g increment	−0.02	[−0.05, 0.01]	.26							
Gestational age per 1 week increment	−1.39	[−2.78, −0.01]	.049	−1.44	[−2.78, −0.11]	.03	−0.97	[−2.24, 0.31]	.14	0.01
Male gender vs. female	−2.05	[−5.31, 1.21]	.22	−4.66	[−9.27, −0.06]	.047				
Bayley-III Cognitive Index score per one-unit increment	0.41	[0.27, 0.55]	< .001	—	—		0.35	[0.21, 0.49]	< .001	0.12
					Full model R^2 = .15			Full model R^2 = .24		

Note. Slope estimates (with 95% CI), *p* values, and R^2 values were obtained from univariate and multivariate linear regression analyses. CI = confidence interval; WISC-IV = Wechsler Intelligence Scale for Children—Fourth Edition; IQ = intelligence quotient; Bayley-III = Bayley Scales of Infant and Toddler Development—Third Edition.
[a]The model includes the factors listed in the column and are factors with *p* values < .20 in the multivariate model.

Table 5. Prediction Model for General Ability Index (GAI).

	Univariate models			Multivariate model[a], not including Bayley-III scores			Multivariate model[a], including Bayley-III scores			R^2
	Slope	95% CI	*p* value	Slope	95% CI	*p* value	Slope	95% CI	*p* value	R^2 full model—Model without the specified variable
Maternal age per 1-year increment	0.38	[0.01, 0.76]	.04	0.41	[0.02, 0.80]	.04	0.41	[0.06, 0.76]	.02	0.03
Primiparity vs. multiparity	1.8	[−1.65, 5.29]	.30	2.53	[−0.99, 6.05]	.16	2.8	[−0.39, 6.00]	.09	0.02
Maternal country of birth Non-Nordic vs. Nordic	2.49	[−7.54, 12.53]	.62							
Maternal smoking vs. nonsmoking	4.03	[−4.40, 12.47]	.35							
Maternal education per one-level increment	2.14	[0.87, 3.42]	.001	1.26	[−0.13, 2.65]	.08				
Paternal education per one-level increment	2.12	[0.91, 3.33]	< .001	1.40	[0.09, 2.70]	.04	1.68	[0.58, 2.79]	.003	0.05
Birth weight per 100 g increment	−0.01	[−0.05, 0.03]	.55							
Gestational age per 1 week increment	−0.88	[−2.36, −0.61]	.24	−1.04	[−2.47, −0.40]	.16				
Male gender vs. female	0.58	[−2.90, 4.06]	.74							
Bayley-III Cognitive Index score per one-unit increment	0.44	[0.29, 0.59]	< .001	—	—		0.42	[0.27, 0.56]	< .001	0.16
				Full model R^2 = .13			Full model R^2 = .26			

Note. Slope estimates (with 95% CI), *p* values, and R^2 values were obtained from univariate and multivariate linear regression analyses.CI = confidence interval; Bayley-III = Bayley Scales of Infant and Toddler Development—Third Edition.

[a]The model includes the factors listed in the column and factors with *p* values < .20 in the multivariate model.

agreement interval, when comparing WISC-IV scores with Bayley-III scores, was −18 units to 26 units for 95% of the cases. The interval indicates poor agreement between the methods. Multivariate analyses recognized the Bayley-III Cognitive Index scores as the most important factor for explaining WISC-IV FSIQ and GAI score variability on group level, in comparison with demographic factors and child's gestational age. Because of the well-known impact of socioeconomic status on cognitive ability, it was expected that parental educational level would contribute significantly as well. The best prediction model (including Bayley-III Cognitive Index scores) explains only 24% of the later FSIQ variance and 26% of the later GAI variance, respectively. This is not necessarily a failure of the predictive models or the instrument construction per se. Instead, this may prove the impact of all the unaccounted for and intervening developmental variables over the course of these 4 critical years. Given how environmental experiences can substantially affect cognitive trajectories in young children, weak associations can be expected regardless.

One of the main goals of the Bayley-III assessment is early identification of severe and chronic deficits. Studies have shown fairly strong associations between cognitive assessment during toddlerhood and school-age IQ in groups of children born preterm and with serious neonatal illnesses (e.g., Doyle & Casalaz, 2001; Voss et al., 2007), indicating that Bayley-III accomplishes this goal in clinical settings. In contrast to previous study by Bode et al. (2014), we could not observe strong relationships between Bayley-III and IQ scores. As their study was performed on children born very preterm, a more robust association would be expected. We investigated method agreement by analyses on individual levels as well, instead of correlational analyses only. Our results rather indicated a weak association between Bayley-III and WISC-IV scores in normal populations, confirming findings from previous studies. Therefore, the question of issue becomes how functions measured by infant development tests are actually related to the factors incorporated in the IQ construct. Previous studies have investigated specific underlying cognitive functions that could influence a child's intelligence level. Prediction of IQ at early school age from infant habituation, novelty preference, and reaction time has been proven higher than infant test scores (Domsch et al., 2009; McCall & Carriger, 1993). These skills are basic and central in cognitive development and, consequently, they may be more stable and less dependent on item construction. It could be argued that if tasks that reflect young children's information processing replaced standardized developmental tests, predictions would improve in normal populations. Nevertheless, our findings that the Bayley-III Cognitive Index score had the strongest association with later IQ (both GAI and FSIQ) support the relative importance of cognitive skills as predictors of later IQ. Hence, it may reflect the central status of information processing abilities, processing speed, and problem-solving skills measured in the Cognitive Index of the revised Bayley-III. Furthermore, associations between the Bayley-III Language Index and the WISC-IV FSIQ were lower than findings by Bode et al. (2014) and Bayley (2006). Potential reasons, related to differences in study design, were discussed above. Nevertheless, weak correlations between the Bayley-III Language Index and the WISC-IV VCI were rather surprising. The Bayley-III Language Index includes measurements of the child's ability to interact with others (Bayley, 2006). This means not only word comprehension and productive vocabulary, but social–emotional aspects of communication as well (e.g., the ability to engage in joint attention). Hence, the Bayley-III Language Index examines communicative developmental functions in a broader sense than do, for example, the WISC-IV VCI. Previous studies propose that socioeconomic factors may have particularly strong effects on language skills (Hoff, 2003; von Stumm & Plomin, 2015).

Both classic developmental theorists such as Jean Piaget (1896-1980; cited in Ellingsen, 2016) have proposed the notion that motor and cognitive development are closely interrelated processes. This relationship has been explored in more recent, clinical studies as well. For example, Burns, O'Callaghan, McDonell, and Rogers (2004) found that preschoolers born extremely preterm, who had developed motor deficiencies, scored lower than at-term peers on general IQ measures, independently of medical and social factors. They conclude that when a child's ability to explore the world is limited, then his or her opportunities to learn about it are limited as well. In our study, performances by generally healthy children, without deficiencies, were examined. Hence, we would expect a stronger relationship between the Bayley-III Motor Index scores and the WISC-IV FSIQ in a clinical group, for example, the EXPRESS sample of children born < 27 gestational weeks (mentioned in the "Method" section).

There were some limitations to the present study. The parental average education level is considered high as 50% of the mothers, and 40% of the fathers have a postsecondary education of ⩾ 2 years. This reflects the general academic achievements of Swedish men and women at 25 to 34 years of age (Statistics Sweden, 2015). However, 3.7% of the mothers and 4.3% of the fathers in our cohort had compulsory education only, compared with approximately 10% of the Swedish working age population in general. This may raise the question of generalizability to both other Swedish and transnational contexts.

The WISC-IV WMI mean score is noticeably low, for reasons we have not been able to identify. As expected, the WISC-IV GAI mean was higher than the WISC-IV FSIQ mean. Yet again, they were equally correlated to the Bayley-III Cognitive Index score. As Sweden is a relatively small linguistic area, the possibility to establish Swedish WISC-IV norms is delimited. Instead, the WISC-IV scores in our study have been evaluated against British norms. Experiences from previous Wechsler scales have shown that mean value differences between British and Swedish versions are small. As an effect of the cultural adaptation of the British version to European circumstances, the content in British and Swedish versions differ only to a small degree (Tideman, 2007; Wechsler, 2007).

Conclusion

We investigated the agreement between Bayley-III measurements at 2.5 years and WISC-IV measurements at 6.5 years in a population of children born at term to highly educated parents. We conclude that the Bayley-III Cognitive Index score specifically may be of relative importance to predict early school-age IQ. The Bayley-III Cognitive Index score was the best explanatory factor of IQ score variability, compared with important social-demographic factors. Results may reflect the central components of the Bayley-III Cognitive Index and their relevance to later IQ. Nevertheless, on an individual level, the Bayley-III Cognitive Index score was demonstrated as being an insufficient predictor of the later WISC-IV FSIQ score. The total regression models, including the Bayley-III Cognitive Index score, modestly contributed to the WISV-IV FSIQ and GAI score variance. We conclude that performance on the Bayley-III should primarily be regarded as a depiction of a child's current functioning level. From a clinical point of view, this picture provides a useful basis from which early intervention plans can be established. However, it does not enable long-standing, reliable prognosis about intelligence in typically developing children.

Acknowledgments

The authors acknowledge the contributions from the members of the Extremely Preterm Infants in Sweden Study (EXPRESS) Group.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Anderson, P. J., De Luca, C. R., Hutchinson, E., Roberts, G., & Doyle, L. W. (2010). Underestimation of developmental delay by the new Bayley-III scale. *Archives of Pediatric & Adolescent Medicine*, *164*, 352-356. doi:10.1001/archpediatrics.2010.20

Bayley, N. (1969). *The Bayley Scales of infant development*. San Antonio, TX: The Psychological Corporation.

Bayley, N. (1993). *The Bayley Scales of infant development-II*. San Antonio, TX: The Psychological Corporation.

Bayley, N. (2006). *Bayley Scales of infant and toddler development technical manual* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Bayley, N. (2009). *Bayley Scales of infant and toddler development* (3rd ed., *Svenskt manual supplement*). Stockholm, Sweden: Pearson.

Bode, M. M., D'Eugenio, D. B., Mettelman, B. B., & Gross, S. J. (2014). Predictive validity of the Bayley, Third Edition at 2 years for intelligence quotient at 4 years in preterm infants. *Journal of Developmental and Behavioral Pediatrics*, *35*, 570-575. doi:10.1097/DBP.0000000000000110

Burns, Y., O'Callaghan, M., McDonell, B., & Rogers, Y. (2004). Movement and motor development in ELBW infants at 1 year is related to cognitive and motor abilities at 4 years. *Early Human Development*, *80*(1), 19-29. doi:10.1016/j.earlhumdev.2004.05.003

Deary, I. J., & Brett, C. E. (2015). Predicting and retrodicting intelligence between childhood and old age in the 6-day sample of the Scottish Mental Survey 1947. *Intelligence*, *50*, 1-9. doi:10.1016/j.intell.2015.02.002

Domsch, H., Lohaus, A., & Thomas, H. (2009). Prediction of childhood cognitive abilities from a set of early indicators of information processing capabilities. *Infant Behavior and Development*, *32*, 91-102. doi:10.1016/j.infbeh.2008.10.006

Doyle, L. W., & Casalaz, D. (2001). Outcome at 14 years of extremely low birth weight infants: A regional study. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, *85*, 159-164. doi:10.1136/Fn.85.3.F159

Ellingsen, K. M. (2016). Standardized assessment of cognitive development: Instruments and issues. In A. Garro (Ed.), *Early childhood assessment in school and clinical child psychology*. New York, NY: Springer. doi:10.1007/978-1-4939-6349-2_2

Fellman, V., Hellström-Westas, L., Norman, M., Westgren, M., Källén, K., Lagercrantz, H., . . . Wennergren, M. (2009). One-year survival of extremely preterm infants after active perinatal care in Sweden. *Journal of the American Medical Association*, *301*, 2225-2233. doi:10.1001/jama.2009.771

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica (Zagreb)*, *25*, 141-151. doi:10.11613/BM.2015.015

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368-1378. doi:10.1111/1467-8624.00612

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253-270. doi:10.1037/h0023816

Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: Which cut-off should be used? *Pediatric Research*, *75*, 670-674. doi:10.1038/pr.2014.10

Luttikhuizen dos Santos, E. S., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: A meta-analysis. *Early Human Development*, *89*, 487-496. doi:10.1016/j.earlhumdev.2013.03.008

McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, *64*, 57-79. doi:10.2307/1131437

McGrath, E., Wypij, D., Rappaport, L. A., Newburger, J. W., & Bellinger, D. C. (2004). Prediction of IQ and achievement at age 8 years from neurodevelopmental status at age 1 year in children with D-transposition of the great arteries. *Pediatrics*, *114*(5), e572-576. doi:10.1542/peds.2003-0983-L

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77-101. doi:10.1037/0003-066X.51.2.77

Serenius, F., Ewald, U., Farooqi, A., Fellman, V., Hafström, M., Hellgren, K., . . . Källén, K. (2016). Neurodevelopmental outcomes among extremely preterm infants 6.5 years after active perinatal care in Sweden. *JAMA Pediatrics*, *170*, 954-963. doi:10.1001/jamapediatrics.2016.1210

Serenius, F., Källén, K., Blennow, M., Ewald, U., Fellman, V., Holmström, G., . . . Strömberg, B. (2013). Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in Sweden. *Journal of the American Medical Association*, *309*, 1810-1820.

Spencer-Smith, M. M., Spittle, A. J., Lee, K. J., Doyle, L. W., & Anderson, P. J. (2015). Bayley-III cognitive and language scales in preterm children. *Pediatrics*, *135*(5), e1258-1265. doi:10.1542/peds.2014-3039

Statistics Sweden. (2015). *Educational attainment of the population 2014*. Retrieved from https://www.scb.se/en/finding-statistics/statistics-by-subject-area/education-and-research/education-of-the-population/educational-attainment-of-the-population/

Tideman, E. (2007). Hur ska WISC-IV användas? In *WISC-IV Svensk Version Manual Del I* (pp. 91-138). Stockholm, Sweden: NCS Pearson.

Tucker-Drob, E. M., Briley, D. A., & Harden, K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science*, *22*, 349-355. doi:10.1177/0963721413485087

Vohr, B. R., Stephens, B. E., Higgins, R. D., Bann, C. M., Hintz, S. R., Das, A., . . . Fuller, J. (2012). Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. *The Journal of Pediatrics*, *161*, 222-228. doi:10.1016/j.jpeds.2012.01.057

von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence*, *48*, 30-36. doi:10.1016/j.intell.2014.10.002

Voss, W., Neubauer, A.-P., Wachtendorf, M., Verhey, J. F., & Kattner, E. (2007). Neurodevelopmental outcome in extremely low birth weight infants: What is the minimum age for reliable developmental prognosis? *Acta Paediatrica*, *96*, 342-347. doi:10.1111/j.1651-2227.2006.00130.x

Wechsler, D. (2003). *Wechsler Intelligence Scale for children* (4th ed.). San Antonio, TX: NCS Pearson.

Wechsler, D. (2007). *Wechsler Intelligence Scale for Children, Svensk version* (4th ed.). Stockholm: NCS Pearson.