The Predictive Value of Developmental Assessments at 1 and 2 for Intelligence Quotients at 6

Jessica B. Girault[a,b], **Benjamin W. Langworthy**[c], **Barbara D. Goldman**[d,e], **Rebecca L. Stephens**[a], **Emil Cornea**[a], **J. Steven Reznick**[e,*], **Jason Fine**[c], and **John H. Gilmore**[1,a]

[a]Department of Psychiatry, Campus Box #7160, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[b]Center for Developmental Science, Campus Box # 8115, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[c]Department of Biostatistics, Campus Box # 7400, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[d]Frank Porter Graham Child Development Institute, Campus Box # 8180, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[e]Department of Psychology and Neuroscience, Campus Box # 3270, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

Abstract

Intelligence is an important individual difference factor related to mental health, academic achievement, and life success, yet there is a lack of research into its early cognitive predictors. This study investigated the predictive value of infant developmental assessment scores for school-age intelligence in a large, heterogeneous sample of single- and twin-born subjects (N = 521). We found that Early Learning Composite (ELC) scores from the Mullen Scales of Early Learning have similar predictive power to that of other infant tests. ELC scores at age 2 were predictive of Stanford-Binet abbreviated intelligence (ABIQ) scores at age 6 (*r* = 0.46) even after controlling for sex, gestation number, and parental education. ELC scores at age 1 were less predictive of 6-year ABIQ scores (*r* = 0.17). When the sample was split to test robustness of findings, we found that results from the full sample replicated in a subset of children born at 32 weeks gestation without birth complications (n = 405), though infant cognitive scores did not predict IQ in a subset born very prematurely or with birth complications (n = 116). Scores at age 2 in twins and singletons showed similar predictive ability for scores at age 6, though twins had particularly high correlations between ELC at age 1 and ABIQ at age 6.

[1]*Corresponding Author*: John H. Gilmore, University of North Carolina at Chapel Hill, Campus Box # 7160, Chapel Hill, NC 27599-7160, *Telephone*: 919-445-0209, *Fax*: 919-445-0234, john_gilmore@med.unc.edu.
[*]Dr. J. Steven Reznick is recognized posthumously for his contributions.

Introduction

Decades of research have revealed that intelligence is related to mental health, academic achievement, occupational status, life success, and longevity (Deary, Pattie, & Starr, 2013; Gottfredson, 1997; Keyes, Platt, Kaufman, & McLaughlin, 2016; Whalley & Deary, 2001). Twin and family studies find that the continuity of intelligence across the lifespan is driven largely by genetic factors, though environmental influences are notable during childhood (Bartels, Rietveld, Van Baal, & Boomsma, 2002; Bishop et al., 2003; Brant et al., 2013). Intelligence is also a marker of brain development and functioning, including trajectories of structural maturation across the lifespan (Schnack et al., 2015; Shaw et al., 2006) and patterns of functional brain activation (Gray, Chabris, & Braver, 2003) differing based on cognitive ability. Genome-wide association studies show that genes linked to brain development are markers of individual differences in cognitive ability (Davies et al., 2016), and that genetic correlations between intelligence in childhood and old age are high (Deary et al., 2012). This body of research highlights that intelligence is dynamically influenced by biological and environmental processes that contribute to unique developmental trajectories.

Much work has been done to understand the continuity and stability of intelligence across the lifespan, and it has been found that school-age intelligence quotients (IQs) are fairly stable predictors of adult ability (Bradway & Thompson, 1962; Deary et al., 2013; Deary, Whiteman, Starr, Whalley, & Fox, 2004; McCall, 1977). However, studies in younger children and infants have been less conclusive. In a sample of roughly fifty children, the Berkley Growth Study revealed that infant test scores (averaged between ages 10, 11, and 12 months) modestly correlated with school age scores (averaged between ages 5, 6, and 7 using different assessments; r = 0.20), while scores averaged between ages 18, 21, and 24 months correlated highly (r = 0.50) with school-age scores (Bayley, 1949). In a 1972 review (McCall, Hogarty, & Hurlburt), data were combined from four studies (including the Berkley Growth Study) using different cognitive tests; the median correlation reported between 19–30 month test scores and 5–7 year scores (r = 0.41) was similar to those observed by Bayley and colleagues (1949), while the correlation between school-age scores and scores from 7–12 month-olds was notably smaller (r = 0.06). In general, it was found that the later a test is given during infancy and toddlerhood, the better its predictive ability for subsequent outcomes (McCall et al., 1972).

Recent studies of the predictive value of such assessments focus almost exclusively on at-risk populations such as premature and very-low-birth-weight cohorts (Bode, D'Eugenio, Mettelman, & Gross, 2014; Hack et al., 2005; Leversen et al., 2012; Potharst et al., 2012; Soysal et al., 2014). Results from these studies provide conflicting evidence about the predictability of early tests for subsequent performance, which may be due to the unique characteristics of these at-risk populations, where some children overcome early deficits while others remain on a delayed trajectory. For example, infant scores from very premature children (Bode et al., 2014), those with neurological impairments (Hack et al., 2005) or perinatal complications (Potharst et al., 2012) were more highly correlated with their subsequent school-age performance, whereas infant scores showed limited predictive value for premature children without major impairments (Leversen et al., 2012).

Other recently published work reporting correlations between infant and school-age cognitive scores include large-scale twin and family studies. In a sample of over 1,000 twins and biological and adopted siblings, Bishop and colleagues (2003) found that infant scores at ages 1 and 2 correlated with principle components derived from cognitive tests at age 7 (r = 0.18 and 0.37, respectively; related participants included in correlations). Another study of 14,000 twins in the UK found that parent reports of 2-year-olds' cognitive ability was correlated with phone-administered portions of cognitive tests at age 7 (r = 0.23) (von Stumm, Gale, Batty, & Deary, 2009). It is important to note that determining the predictive ability of infant cognition for subsequent intelligence scores was not the primary purpose of either of those studies.

The generalizability of much of the previous work is limited by small sample sizes (Bayley, 1949; Fagan, Holland, & Wheeler, 2007; McCall, 1977), focus on special populations (Bode et al., 2014; Hack et al., 2005; Leversen et al., 2012; Potharst et al., 2012; Soysal et al., 2014), or lack of participant diversity (Bishop et al., 2003; Sutcliffe, Soo, & Barnes, 2010). Results from twin-only studies, while large-scale, may also be difficult to generalize to other populations given that twins have lower IQs in childhood (Bishop et al., 2003; Ronalds, De Stavola, & Leon, 2005), and potentially different cognitive developmental trajectories than single-born children. Therefore, it remains unknown how well the correlations between infant and school-age intelligence reported in the literature generalize across more diverse samples.

The goal of the present study is to investigate the predictive value of cognitive assessments at 1 and 2 years of age for subsequent IQ at age 6 in a relatively large, heterogeneous, longitudinal sample of single- and twin-born children. This study is novel in several respects. First, it is one of the largest studies of the predictive ability of infant cognitive scores for school-age intelligence to date, with 521 subjects in the sample. Second, results are derived from a sample that is generally representative of the U.S. population (US Census, 2016a), whereas many previous studies were conducted in predominantly Caucasian-only samples, or those with less than 10% of participants from other racial or ethnic groups. Finally, to our knowledge, this is the first study to test the predictive ability of the Early Learning Composite (ELC) from the Mullen Scales of Early Learning (MSEL) (Mullen, 1995) for school-age intelligence scores in a healthy sample, despite its use in several longitudinal studies of development in the context of brain-behavior relations and its widespread use in autism spectrum disorders research (Deoni et al., 2014; Gilmore et al., 2007; Lee et al., 2017; Wolff et al., 2012). We expected ELC scores to show similar correlations with school-age intelligence scores as those reported using other infant tests, with scores at age 2 being a stronger predictor of IQ at age 6 than measures at age 1. In order to test the robustness of our findings and compare our results with those previously published, we also ran sensitivity analyses subdividing the sample into subsets with and without birth complications (prematurity and/or perinatal complications), and split by gestation number into twins and singletons. We expected that our results would be similar between the full sample and the subset without birth complications, but hypothesized that the premature subset may show a different trend based on previously reported inconsistencies in the literature with this at-risk group. We also expected similar predictive patterns between early cognition and later IQ in twins and singletons given the similarity in

effect sizes reported across samples in the literature. Finally, we explored the effects of demographic factors on infant and school-age cognitive scores, expecting that variables related to socioeconomic status (SES) and perinatal characteristics would be both predictive of and related to individual differences in ability.

Methods

Participants

Participants were part of the XXX Study of early childhood brain development in singletons and twins (XXX; XXX). Pregnant women were recruited during the second trimester of pregnancy at the Prenatal Diagnostic Clinics of the XXX Medical Center by flyers and study staff. Mothers were excluded from the current study for pregnancy complications (major illness, using illegal drugs, or severe infection), or a diagnosis of a major psychiatric disorder. All offspring participants, born between 2003 and 2014, underwent cognitive testing at ages 1, 2, and 6 years. We retrospectively identified 521 children with at least cognitive test scores from at least two ages, no major medical issues, and no psychiatric diagnoses up to age 6. We chose to exclude subjects on the basis of maternal and child psychiatric diagnoses as we have a substantial enrichment of this population in our total subject pool due to recruiting mothers with psychiatric illness as part of other lines of research in the lab. Our sample is generally representative of the local area (US Census, 2016b) and the U.S. population (US Census, 2016a) in terms of race and ethnicity, though our sample over-represents African Americans in both regards (12.9% of local population, 13.3% of national population, 21.3% of our sample), and under-represents Asians (5.7% of national population, 1.5% of our sample) and American Indians (1.3% of national population, 0.4% of our sample), compared to current national statistics. Hispanics are underrepresented in these data (8.4% of national population, 4.8% of our sample) because some children could not undergo cognitive testing in English. Table 1 outlines the demographic characteristics of the entire sample. Informed written consent and parental permission were obtained for all participants and all study protocols were approved by the XXX Institutional Review Boards of XXX and XXX.

In sensitivity analyses testing the robustness of our results, we subdivided the sample into subsets with and without birth complications and split by gestation number into twins and singletons. Those with birth complications (n = 116, 22% of entire sample) included all subjects born at <32 weeks gestation and spending >24 hours in the neonatal intensive care unit (NICU). Twin versus singleton analyses were only conducted on subjects without birth complications (n = 405, 78% of entire sample) to avoid an over-representation of very premature subjects in the twin sample. We compared a sample of 175 twins to 230 singletons. For details on demographics for the subsets, see Supplement S1.A.

Cognitive Assessments

Cognitive ability was assessed in the Infant and Child Assessment lab at the XXX. Experienced testers were trained and supervised by a developmental psychologist with extensive assessment experience. At ages 1 and 2 years, we used the Mullen Scales of Early Learning (MSEL). At age 1, infants were assessed while being held in the lap of a parent,

guardian, relative, or, rarely, study staff in the case of twins if only one parent or relative accompanied the family. At age 2, children were seated on their own during testing, with a parent, guardian, or relative present in the room. Performance on the four MSEL cognitive Scales (Visual Reception, Fine Motor, Expressive and Receptive Language) are conventionally combined into an Early Learning Composite (ELC) standard score (range: 49–155, M =100, SD =15). The ELC has high internal consistency (median = 0.91) and reliability (median = 0.84 for the cognitive scales during these testing ages), and principal factor loadings of the scales lend support for the construct validity of the ELC as a general measure of cognitive ability (Mullen, 1995).

The MSEL was used in this prospective study of brain development specifically because of its potential to capture uneven development in different cognitive abilities (Askhoomoff et al., 2006; De Giacomo & Fombonne, 1998; Filipek et al., 1999). Compared to the commonly used second edition of the Bayley Scales of Infant Development (Bayley II; (Bayley, 1993)), which was the version available at the start of this longitudinal study, the MSEL has the advantage of providing standardized T-scores that factor in age at testing for each of the scales, as well as age equivalent independent measures of gross and fine motor, visual reception, and expressive and receptive language scores. In contrast, the Bayley II generated a Mental Developmental Index (MDI) which assessed cognition through evaluating sensory perception, knowledge, memory, problem solving, and early language that could not be decomposed to probe specific cognitive versus language deficits (Lowe et al., 2012). Due to the fact that, as part of the larger study of brain development, we were collecting data on a heterogeneous population including infants born to mothers with diagnosed psychiatric illness, we wanted to ensure the ability to test specific deficits in distinct developmental domains (i.e. language vs. motor). Importantly, however, the ELC standard score derived from the fine motor, visual reception, expressive, and receptive language scales is highly correlated with the Bayley MDI (r = 0.70, n = 103 between 6 and 15 months of age), according to a study presented in the MSEL technical manual (Mullen, 1995).

Intelligence at age 6 was assessed in the same Infant and Child Assessment lab by experienced testers, supervised by the same developmental psychologist, using the 5th Edition of the Stanford-Binet Intelligence Scales (SB5;(Roid, 2003)). At age 6, children were typically tested alone while a parent was present directly outside the room on the other side of one-way glass, but parents were given the option to sit in the room as the SB5 was administered. The outcome used in this analysis is the abbreviated IQ (ABIQ) measure (range: 50–150, M =100, SD =15) derived from scores on the verbal knowledge and non-verbal fluid reasoning tasks reflecting the child's lexical knowledge and ability to solve problems. These two tests serve as the "routing" tests, which are used to determine the entry level for subsequent tests of verbal and non-verbal abilities. The ABIQ has an internal consistency of 0.91 and a test-retest reliability of 0.84, and correlates highly with the full-scale IQ, which can only be derived from significantly longer testing sessions (Roid, 2003).

A total of 509 1-year ELC scores (ELC1), 499 2-year ELC scores (ELC2), and 275 6-year ABIQ scores (ABIQ6) were used in this study. All included participants had at least two test scores, 487 had both ELC scores, 263 had ELC1 and ABIQ6, 253 had ELC2 and ABIQ6,

and 241 had all three cognitive assessment scores. Our participants, on average, performed slightly better on the MSEL and SB5 than the normalization samples (Table 1).

Statistical Analysis

The relation between ELC scores and ABIQ6 was estimated using Generalized Estimating Equations (GEE) treating each family (twins and siblings) as a cluster, accounting for possible correlations in observational data from twins and siblings. GEE estimates allow for consistent estimates of the relationship between ELC and ABIQ6 even if there is correlation within families (twins and siblings). Unlike other methods that can account for such correlation, like mixed effects models, GEE estimates are consistent even if the underlying correlation structure between families is unknown or misspecified. Using methods similar to Yan and Fine (2004) which allow for modelling the effects of covariates on the correlation parameters, we were able to estimate correlations between infant cognitive scores for the same participant over time and for scores between twins and siblings. These analyses permitted covariates in the model for predicting ABIQ6 scores, with the best fitting model selected using quasi-Akaike's Information Criterion (QIC, Pan, 2004). Potential variables included sex, gestation number (twin or singleton), gestational age at birth (days), maternal and paternal education (years), chronological age at the time of the assessment administration (days), and the number of months since start of assessment collection in the study (to account for possible drift in cognitive testing administration due to changes in personnel over the 10-year study period), as well as the interaction between all the variables and cognitive scores. Initially models were run including only ELC scores as explanatory variables. QIC was used to determine whether the model with only ELC1, only ELC2, or both was best. Next, in addition to the ELC scores, covariates mentioned above were added to the model. The final model selected through QIC to predict ABIQ6 included ELC2, sex, age at SB5 testing, paternal education, gestation number, and months since start of SB testing. Additionally, models were run using the same approach to estimate the relation between ELC1 and ELC2 scores and demographic variables. We also used the GEE model to calculate correlations between ELC1, ELC2, and ABIQ6 scores so that we can estimate the variance in later ABIQ explained by early cognitive performance, allowing for comparison to previously published works.

Some of our infants were lost to follow up or were not old enough to have taken the SB5 at the time of data analysis. Of those old enough to have taken the SB5 at age 6, 32% of subjects did not take the test. We investigated the missingness using a binomial GEE in which the outcome variable was a binary indicator variable for whether or not the child had an ABIQ6 score. Potential explanatory variables were ELC1, ELC2, calendar year and month for taking the ELC1 test, maternal and paternal education, gestation number and gestational age at birth. An independent working correlation matrix was used with each family treated as a cluster. The final model was chosen using QIC and is reported in Table 2. We found that increased paternal education resulted in reduced likelihood of follow-up 6-year SB5 assessments, possibly related to changing paternal employment locations (and thus a family relocation) in the 4 to 5 years following the earlier assessments. Given the recruitment area and proximity to the University, it is possible that fathers may have completed graduate degrees, internships, or residencies at the University and relocated

afterwards. The data showed trending significance for taking the MSEL at age 1 later in the study increasing the odds of a follow up, while conversely suggesting that higher ELC1 scores resulted in a decreased likelihood of 6-year follow-up. Since participants were not lost to follow up at random, a linear mixed model was employed as a sensitivity check because they are valid under a weaker missing at random assumption and provide a check the GEE findings. Results were highly similar between the two models and only the GEE is reported here (see Supplement S1.B for linear mixed model results).

Results

Prediction of 6-year ABIQ

Models using only ELC scores as predictors of ABIQ6 scores revealed that a one-point increase in ELC1 and ELC2 predicts an increase in ABIQ6 of 0.16 points (SE = 0.06, p = 0.01), and 0.41 points (SE = 0.06, p = <0.001), respectively (Table 3). Uncorrected scatterplots of these data can be seen in Figure 1B–C. When both ELC scores were in the model together, ELC1 was not predictive of ABIQ6, while a one-unit increase in ELC2 increased the expected ABIQ6 by 0.40 points (SE = 0.06, p = <0.001). This demonstrates that after controlling for ELC2, the additional knowledge of ELC1 does not significantly contribute to the prediction of ABIQ6. Correlations calculated between ELC1, ELC2 and ABIQ6 based on the GEE model reveal that ELC1 scores account for 2.8% of the variance in ABIQ6 (r = 0.169, SE = 0.066), while ELC2 scores account for 21.3% of the variance in ABIQ6 (r = 0.461, SE = 0.061). It should be noted that the GEE *estimates* are model coefficients and should be interpreted such that unit-wise increases in each predictor variable result in a unit-wise change in the response variable, while GEE *correlations* are measures of the strength of a linear association between predictor and response variables that can be interpreted similarly to Pearson's correlations. The similarity in magnitude between the GEE estimates and correlations is coincidental, as they compare different associations between scores.

Results from the full model (Table 4) estimated a one-point increase in ELC2 predicted an increase in ABIQ6 of 0.28 (SE = 0.06, p = <0.001), when holding all other covariates constant. A one-day increase in age at 6-year testing led to an increase in expected ABIQ6 of 0.04 points (SE = 0.02, p = <0.001). Date of the 6-year assessment was not significantly related to ABIQ6 (SE = 0.04, p = 0.06). Every additional year of paternal education accounted for an increase of 1.18 points in offspring ABIQ6 (SE = 0.25, p = <0.001). In a separate model, we replaced paternal with maternal education, and results were highly similar (Supplement S1.C). This was expected given the strong correlations between maternal and paternal education (r =0.67), and their correlations with household income (r = 0.49 and r = 0.42, respectively) in our sample. There was a trend for males to score 2.77 points lower than females at age 6, though it did not reach statistical significance (SE = 0.02, p = 0.08). The strongest predictor of ABIQ6 was gestation number; when controlling for all other covariates, twins scored 6.11 points lower than singletons (SE = 1.61, p = <0.001). Gestational age at birth was not selected in the model.

In a set of sensitivity analyses, we tested the robustness of the predictive value of ELC scores for ABIQ6. We found that results from the full sample were in line with those found

in a subset of participants without birth complications (n = 405;   32 weeks gestation,   24 hours in the neonatal intensive care unit (NICU)), though they did not replicate in a subset of children born very prematurely or with birth complications (n = 116; <32 weeks gestation, >24hrs in NICU). When comparing the predictive ability of infant cognitive scores for ABIQ6 between twins (n = 157) and singletons (n = 230), we found that predictions were stronger among twins, particularly from age 1 to age 6. See Table 5 for a summary of results across samples.

Infant Cognitive Scores

We estimated correlations between ELC1 and ELC2 scores to be 0.30 (SE = 0.05, p = <0.001) and found that correlations between scores of twins (r = 0.70, SE = 0.06, p = <0.001) and siblings (r = 0.41, SE = 0.22, p = 0.06) taken at the same age were higher than those for the same child over time (results verified with Pearson correlations; Table 6). Scatterplots showing unadjusted associations between ELC1 and ELC2 are shown in Figure 1A. The model investigating the relation between infant scores and other demographic variables (Table 7) revealed that at age 1, twins did not score significantly lower than singletons (p = 0.39), but by age 2, twins scored 11.47 points lower than single-born children, when all other variables in the model are held constant (se = 2.46, p = 1.60E04). Every additional year of paternal education predicted an increase of 1.61 points in ELC2 (se = 0.25, p = 1.20E-04), but a decrease of 0.53 points in ELC1 (se=0.19, p=7.01E-03). Holding all other variables constant, each additional day of age led to a decrease in expected ELC1 of 0.16 points (se=0.5, p=1.02E-03), and a 0.14-point increase in ELC2 (se= 0.06, p=0.02). Finally, each additional month after the start of data collection led to an expected increase in ELC1 scores of 0.15 points (se=0.02, p= 6.50E-12), but did not significantly impact ELC2 scores (p=0.12), holding all other variables constant.

Discussion

In the present study, we show that scores on cognitive assessments at age 2 are significant predictors of intelligence scores at age 6 in a large, heterogeneous sample. As expected, scores at age 1 were far less predictive. These associations are of similar magnitude to those published in the literature on typically developing samples and twins using other infant developmental assessments (Bayley, 1949; Bishop et al., 2003; McCall, 1977), where correlations between scores taken around age 2 and age 6 ranged from 0.37 to 0.50 compared to our finding of a correlation of 0.46. This suggests that the MSEL has similar predictive power to other infant tests. Importantly, we also show that the relation between infant and school-age ability vary based on individual difference factors including prematurity, birth complications resulting in extended NICU stay, and gestation number. Together, these results extend our understanding of the predictive value of infant cognitive tests for later intelligence by informing us of the extent to which such predictions have the ability to generalize to more diverse populations.

The low predictive ability of cognitive tests at age 1, which accounted for less than 3% of the variance in 6-year cognitive performance, may be related to the large dependence of many tests, including the MSEL, on language comprehension, which is limited at this age,

and items that involve maternal report. Mothers differ: some are able to readily provide a list of words their children know and understand, while others are less prepared to present such a list, but may provide additional information over the course of the assessment. Infants and toddlers also differ dramatically in their comfort with the testing environment. Some infants are comfortable from beginning of the testing session, in a new place, with new people, to use the words that they know, or to respond by copying what the tester had demonstrated, while others take significantly more time to acclimate to the testing environment.

In addition, the absence of strong prediction may also be related to the generally inevitable lack of methodological continuity in the testing of various constructs across early childhood, given the dramatic changes in skill levels in multiple domains between infants and 5-to 7-year-old children, and the differences in the test items given at these very different ages. As skills and test items become more similar, one would expect increasing concordance; this is certainly a developmental issue, as the repertoire that is available to the infant changes dramatically over this overall time frame. This may be partly reflected in our finding that correlations between infant scores from family members taken at the same age are stronger than the correlations between scores for the same child at ages 1 and 2.

Finally, some developmentalists, such as Piaget, would argue that a discontinuous shift in cognitive processing occurs between the first and second year of life such that younger infants are limited to more sensorimotor based forms of cognition that later shift to representational thinking by age 2 which is more consistent with cognition in adults (Müller, 2009). This could account for the lack of correlation seen between the scores at ages 1 and 6, though more recent work would suggest that cognitive development is more continuous than previously thought, such that even very young infants possess at least a very rudimentary conceptual system (Mandler 2007; Moore & Meltzoff, 2004). Elements of such a rudimentary conceptual system may be demonstrated in clever research designs, but may not be present in many of the items in traditional assessments.

It is important to highlight that even though scores at age 2 are better predictors than those obtained at age 1, every ELC2 point only accounted for a predicted increase of 0.41 points in ABIQ6 scores, accounting for roughly 21% of the variance in 6-year scores, leaving a large portion of variability unexplained. This is in line with previous research concluding that early cognitive scores should not be used alone to identify infants at-risk for future poor performance (Colombo, 1993). Also of note, we observed an increase in cognitive scores from the beginning of the study to the end of assessment collection for ELC1, but not ELC2 or ABIQ6. This may be related to changes in personnel, an increase in their experience and training, or changes in the larger community environment or subject population over time. Because we noticed this trend, we controlled for months since the start of data collection, which is approximately a decade long; however, this might be considered as a limitation of our study.

In our sample, we observed that twins score more than 6 points lower on the ABIQ6 than singletons. These findings are consistent with previous reports (Ronalds et al., 2005), but often go undiscussed in heritability studies of cognition (Bishop et al., 2003; Stumm & Plomin, 2015). Importantly, we found that the predictive ability of scores at ages 1 and 2 for

subsequent school-age IQ were notably higher for twins compared to singletons, with the ELC1 being nearly three times as predictive in twins. This may be due to differences in demographic characteristics between families of twins and singletons in our sample, however Ronalds and colleagues (2005) found that twins have lower IQ scores at ages 7 and 9 than singleton children in the same family. The lower intelligence scores of twins may reflect reduced fetal growth and shorter gestation, though we excluded participants that were born very prematurely and spent more than a day in the NICU in our sensitivity analyses. Additionally, none of our models selected gestational age at birth as a significant factor predicting infant or school-age cognitive scores. However, it is important to note that MSEL scores were adjusted for gestational age at birth, and thus hinder our ability to understand the impact of gestational age on scores at ages 1 and 2. The notable increase in predictive ability of infant scores for 6-year IQ scores in twins remains puzzling, but could potentially be related to twin-twin interactions or twin-specific parenting styles (Rutter & Redshaw, 1991) that could shape the child's learning environment across development.

Another important factor contributing to cognitive scores was parental education, which may be a proxy of SES effects, as paternal and maternal education in this sample are highly correlated with each other and with household income. SES has been found to be significantly associated with the development of intelligence in a large twin study, with those from low SES families scoring approximately 6 points lower on IQ tests at age 2 than those from high SES backgrounds – an effect that nearly tripled by age 16 (Stumm & Plomin, 2015). The effects of parental education were smaller in our sample, with every additional year of either maternal or paternal education contributing to an increase of roughly 1 point in children's 6-year scores.

Our sample contained a subset of participants born prematurely or with birth complications resulting in a stay in the NICU. When we included this at-risk group in our main analysis, it did not change the findings. However, when analyzed alone, we observed particularly low, non-significant correlations between infant cognitive scores and school-age IQ. This sample only included participants without any major medical issues or psychiatric disorders up to age 6, and thus we may have analyzed a potentially "resilient" group. These findings echo those showing that premature children without neurological abnormalities have the lowest predictions from infancy to later outcomes (Hack et al., 2005), presumably because there is more variability in outcomes. Alternatively, the predictive value of infant tests in at-risk groups may be inherently lower because other factors, such as access to resources and postnatal care, are more important or deterministic than early test scores in predicting later outcomes.

Conclusions & Future Directions

Our study revealed important information about the predictive ability of infant cognitive scores for school-age IQ; namely that by age 2, infant cognitive ability is a fairly strong predictor of outcomes 4 years later, across a period marked by tremendous cognitive gains (Kagan, Herschkowitz, & Herschkowitz, 2005). These results would suggest that the foundations of later intelligence are largely in place by age 2, which is in line with work illustrating the heightened plasticity of the first two postnatal years for both cognitive and

brain development (Gilmore et al., 2012; Lyall et al., 2015; Nelson et al., 2007). Importantly, this work is also in agreement with the large body of research highlighting the long-lasting impact of early life experience on subsequent development (Lupien, McEwen, Gunnar, & Heim, 2009; Sonuga-Barke et al., 2017). Taken together, these results emphasize that this period of early childhood, particularly before age 2, is one that deserves additional study from developmental science and intervention-based perspectives. Interestingly, individual difference factors relating to cognition in this study, namely paternal education, have also been linked to infant brain structure (Knickmeyer et al., 2016), highlighting the need for future studies of the potential mechanisms by which brain-cognition relations emerge across ontogeny and may be influenced by sociodemographic factors. Finally, studies that focus on identifying measures of cognitive continuity across early development will be key to understanding how infant abilities may form the basis of later intelligence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Akshoomoff N. Use of the Mullen Scales of Early Learning for the assessment of young children with Autism Spectrum Disorders. Child Neuropsychology: a Journal on Normal and Abnormal Development in Childhood and Adolescence. 2006; 12(4–5):269–277. DOI: 10.1080/09297040500473714 [PubMed: 16911972]

Bartels M, Rietveld MJH, Van Baal GCM, Boomsma DI. Genetic and environmental influences on the development of intelligence. Behavior Genetics. 2002; 32(4):237–249. [PubMed: 12211623]

Bayley N. Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years. The Pedagogical Seminary and Journal of Genetic. 1949; 75(2):165–196. DOI: 10.1080/08856559.1949.10533516

Bayley N. Bayley Scales of Infant Development Manual. 2. San Antonio, TX: The Psychological Corporation; 1993.

Bishop EG, Cherny SS, Corley R, Plomin R, Defries JC, Hewitt JK. Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. Intelligence. 2003; 31(1):31–49. DOI: 10.1016/S0160-2896(02)00112-5

Bode MM, D'Eugenio DB, Mettelman BB, Gross SJ. 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: JDBP. 2014; 35(9):570–575. DOI: 10.1097/DBP.0000000000000110 [PubMed: 25370298]

Bradway KP, Thompson CW. Intelligence at adulthood: A twenty-five year follow-up. Journal of Educational Psychology. 1962; 53(1):1–14. DOI: 10.1037/h0045764

Brant AM, Munakata Y, Boomsma DI, Defries JC, Haworth CMA, Keller MC, et al. The nature and nurture of high IQ: an extended sensitive period for intellectual development. Psychological Science. 2013; 24(8):1487–1495. DOI: 10.1177/0956797612473119 [PubMed: 23818653]

Colombo J. Infant Cognition. Plomin R, editorVol. 5. Sage Publications; 1993.

Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Molecular Psychiatry. 2016; 21(6):758–767. DOI: 10.1038/mp.2016.45 [PubMed: 27046643]

De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. European Child & Adolescent Psychiatry. 1998; 7(3):131–136. DOI: 10.1007/s007870050058 [PubMed: 9826299]

Deary IJ, Pattie A, Starr JM. The stability of intelligence from age 11 to age 90 years: the Lothian birth cohort of 1921. Psychological Science. 2013; 24(12):2361–2368. DOI: 10.1177/0956797613486487 [PubMed: 24084038]

Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. Journal of Personality and Social Psychology. 2004; 86(1):130–147. DOI: 10.1037/0022-3514.86.1.130 [PubMed: 14717632]

Deary IJ, Yang J, Davies G, Harris SE, Tenesa A, Liewald D, et al. Genetic contributions to stability and change in intelligence from childhood to old age. Nature. 2012; 482(7384):212–215. DOI: 10.1038/nature10781 [PubMed: 22258510]

Deoni SCL, O'Muircheartaigh J, Elison JT, Walker L, Doernberg E, Waskiewicz N, et al. White matter maturation profiles through early childhood predict general cognitive ability. Brain Structure and Function. 2014; 221(2):1189–1203. DOI: 10.1007/s00429-014-0947-x [PubMed: 25432771]

Fagan JF, Holland CR, Wheeler K. The prediction, from infancy, of adult IQ and achievement. Intelligence. 2007; 35(3):225–231. DOI: 10.1016/j.intell.2006.07.007

Filipek PA, Accardo PJ, Baranek GT, Cook EH Jr, Dawson G, Gordon B, et al. The screening and diagnosis of autistic spectrum disorders. Journal of Autism and Developmental Disorders. 1999; 29(6):439–484. DOI: 10.1023/A:1021943802493 [PubMed: 10638459]

Gilmore JH, Lin W, Prastawa MW, Looney CB, Vetsa YSK, Knickmeyer RC, et al. Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. The Journal of Neuroscience: the Official Journal of the Society for Neuroscience. 2007; 27(6):1255–1260. [PubMed: 17287499]

Gilmore JH, Shi F, Woolson SL, Knickmeyer RC, Short SJ, Lin W, et al. Longitudinal development of cortical and subcortical gray matter from birth to 2 years. Cerebral Cortex (New York, N.Y.: 1991). 2012; 22(11):2478–2485. DOI: 10.1093/cercor/bhr327

Gottfredson LS. Why g matters: The complexity of everyday life. Intelligence. 1997; 24(1):79–132. DOI: 10.1016/S0160-2896(97)90014-3

Gray JR, Chabris CF, Braver TS. Neural mechanisms of general fluid intelligence. Nature Neuroscience. 2003; 6(3):316–322. DOI: 10.1038/nn1014 [PubMed: 12592404]

Hack M, Taylor HG, Drotar D, Schluchter M, Cartar L, Wilson-Costello D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. Pediatrics. 2005; 116(2):333–341. DOI: 10.1542/peds.2005-0173 [PubMed: 16061586]

Kagan J, Herschkowitz N, Herschkowitz E. A Young Mind in a Growing Brain. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.; 2005.

Keyes KM, Platt J, Kaufman AS, McLaughlin KA. Association of Fluid Intelligence and Psychiatric Disorders in a Population-Representative Sample of US Adolescents. JAMA Psychiatry. 2016.

Knickmeyer RC, Xia K, Lu Z, Ahn M, Jha SC, Zou F. , et al. Impact of Demographic and Obstetric Factors on Infant Brain Volumes: A Population Neuroscience Study. Cerebral Cortex. 2016.

Lee SJ, Steiner RJ, Yu Y, Short SJ, Neale MC, Styner MA, et al. Common and heritable components of white matter microstructure predict cognitive function at 1 and 2 y. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114(1):148–153. DOI: 10.1073/pnas.1604658114 [PubMed: 27994134]

Leversen KT, Sommerfelt K, Elgen IB, Eide GE, Irgens LM, Júlíusson PB, Markestad T. Prediction of outcome at 5 years from assessments at 2 years among extremely preterm children: a Norwegian

national cohort study. Acta Paediatrica (Oslo, Norway: 1992). 2012; 101(3):264–270. DOI: 10.1111/j.1651-2227.2011.02504.x

Lowe JR, Erickson SJ. paediatrica, R. S. A. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing?. Wiley Online Library. 2012. (n.d.).

Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nature Reviews Neuroscience. 2009; 10(6):434–445. DOI: 10.1038/nrn2639 [PubMed: 19401723]

Lyall AE, Shi F, Geng X, Woolson S, Li G, Wang L, et al. Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. Cerebral Cortex (New York, N.Y.: 1991). 2015; 25(8):2204–2212. DOI: 10.1093/cercor/bhu027

McCall RB. Childhood IQ's as Predictors of Adult Educational and Occupational Status. Science. 1977; 197(4302):482–483. DOI: 10.1126/science.197.4302.482 [PubMed: 17783247]

McCall RB, Hogarty PS, Hurlburt N. Transitions in infant sensorimotor development and the prediction of childhood IQ. American Psychologist. 1972; 27(8):728–748. DOI: 10.1037/h0033148 [PubMed: 4560389]

Mullen EM. Mullen Scales of Early Learning: AGS Edition. Circle Pines, MN: American Guidance Service, Inc.; 1995.

Nelson CA, Zeanah CH, Fox NA, Marshall PJ, Smyke AT, Guthrie D. Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. Science. 2007; 318(5858): 1937–1940. DOI: 10.1126/science.1143921 [PubMed: 18096809]

Pan W. Akaike's Information Criterion in Generalized Estimating Equations. Biometrics. 2004; 57(1): 120–125. DOI: 10.1111/j.0006-341X.2001.00120.x

Potharst ES, Houtzager BA, van Sonderen L, Tamminga P, Kok JH, Last BF, van Wassenaer AG. Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. Developmental Medicine & Child Neurology. 2012; 54(3):240–246. DOI: 10.1111/j.1469-8749.2011.04181.x [PubMed: 22188215]

Roid GH. Stanford-Binet intelligence scales (5th ed.). 5. Itasca, IL: Riverside; 2003.

Ronalds GA, De Stavola BL, Leon DA. The cognitive cost of being a twin: evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. Bmj. 2005; 331(7528):1306–0. DOI: 10.1136/bmj.38633.594387.3A [PubMed: 16299014]

Rutter M, Redshaw J. Annotation: Growing up as a Twin: Twin-Singleton Differences in Psychological Development. Journal of Child Psychology and Psychiatry. 1991; 32(6):885–895. DOI: 10.1111/j.1469-7610.1991.tb01916.x [PubMed: 1744192]

Schnack HG, van Haren NEM, Brouwer RM, Evans A, Durston S, Boomsma DI, et al. Changes in Thickness and Surface Area of the Human Cortex and Their Relationship with Intelligence. Cerebral Cortex. 2015; 25(6):1608–1617. DOI: 10.1093/cercor/bht357 [PubMed: 24408955]

Shaw P, Greenstein D, Lerch J, Clasen L, Lenroot R, Gogtay N, et al. Intellectual ability and cortical development in children and adolescents. Nature. 2006; 440(7084):676–679. DOI: 10.1038/nature04513 [PubMed: 16572172]

Sonuga-Barke EJS, Kennedy M, Kumsta R, Knights N, Golm D, Rutter M. , et al. Child-to-adult neurodevelopmental and mental health trajectories after early life deprivation: the young adult follow-up of the longitudinal English and Romanian Adoptees study. Lancet (London, England). 2017.

Soysal AS, Gucuyener K, Ergenekon E, Turan Ö, Koc E, Turkyılmaz C, et al. The prediction of later neurodevelopmental status of preterm infants at ages 7 to 10 years using the Bayley Infant Neurodevelopmental Screener. Journal of Child Neurology. 2014; 29(10):1349–1355. DOI: 10.1177/0883073813520495 [PubMed: 24563478]

Stumm von S, Plomin R. Socioeconomic status and the growth of intelligence from infancy through adolescence. Intelligence. 2015; 48:30–36. DOI: 10.1016/j.intell.2014.10.002 [PubMed: 26640306]

Stumm von S, Gale CR, Batty GD, Deary IJ. Childhood intelligence, locus of control and behaviour disturbance as determinants of intergenerational social mobility: British Cohort Study 1970. Intelligence. 2009; 37(4):329–340. DOI: 10.1016/j.intell.2009.04.002

Sutcliffe AG, Soo A, Barnes J. Predictive value of developmental testing in the second year for cognitive development at five years of age. Pediatric Reports. 2010; 2(2):e15.doi: 10.4081/pr. 2010.e15 [PubMed: 21589828]

U.S. Census Bureau. United States Quick Facts. 2016a. Retrieved February 23, 2017 from: https://www.census.gov/quickfacts/

U.S. Census Bureau. State & County Quick Facts. Orange County, NC: 2016b. Retrieved February 23, 2017 from: https://www.census.gov/quickfacts/table/PST045216/37135,00

Whalley LJ, Deary IJ. Longitudinal cohort study of childhood IQ and survival up to age 76. Bmj. 2001; 322(7290):819–819. DOI: 10.1136/bmj.322.7290.819 [PubMed: 11290633]

Wolff JJ, Gu H, Gerig G, Elison JT, Styner M, Gouttard S, et al. Differences in White Matter Fiber Tract Development Present From 6 to 24 Months in Infants With Autism. American Journal of Psychiatry. 2012; 169(6):589–600. [PubMed: 22362397]

Yan J, Fine J. Estimating equations for association structures. Statistics in Medicine. 2004; 23(6):859–74. discussion 875-7-879-80. DOI: 10.1002/sim.1650 [PubMed: 15027075]

Highlights

- Cognitive scores at age 2 predict IQ at age 6 ($r = 0.46$).

- Scores at age 1 are far less predictive of IQ at age 6.

- Twins have lower IQs that are better predicted by their infant scores.

- Infant cognitive scores did not predict IQ in a sample of premature subjects.

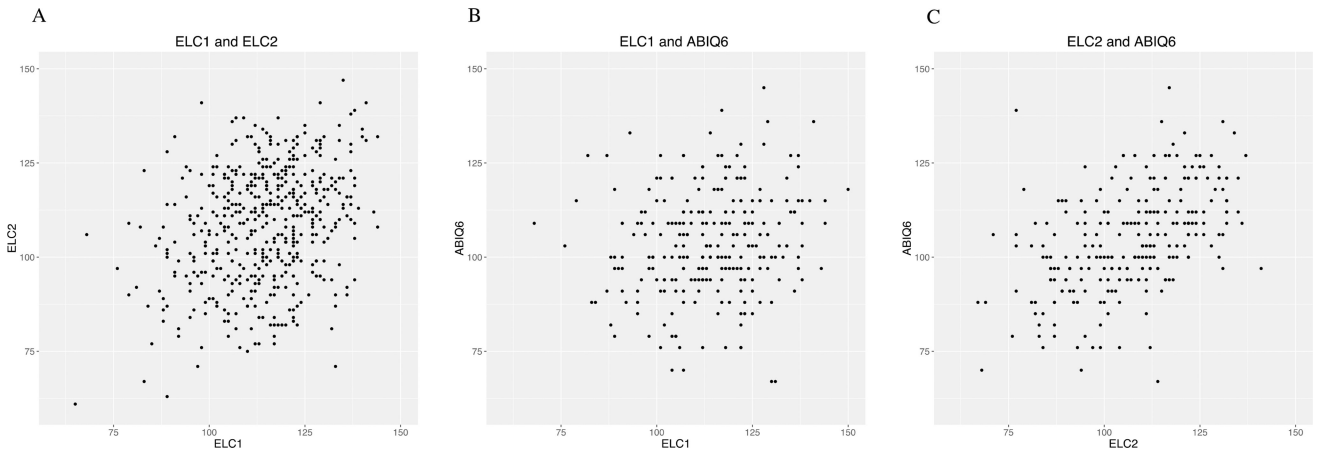- Parental education, sex, and gestation number contributed to predictions.

Figure 1. Relationships between ELC1, ELC2, and ABIQ6
Raw plots of the relationships between ELC1 and ELC2 (A), ELC1 and ABIQ6 (B), and ELC2 and ABIQ6 (C). *[editorial comment: 2-column fitting image]*

Table 1

Sample Characteristics

	Mean (Std Dev)/Percent
Sex (% Male)	50.10
Gestation (% Twins)	52.4
Gestational Age at Birth (Days)	260.58 (20.47)
Duration in NICU [*] (Days)	4.48 (12.86)
Days Since Birth ELC1	393.36 (26.86)
Days Since Birth ELC2	759.48 (30.25)
Days Since Birth ABIQ6	2230.84 (62.72)
ELC1	114.34 (13.45)
ELC2	108.16 (15.33)
ABIQ6	104.03 (14.14)
Maternal Education (Years)	15.90 (3.07)
Paternal Education (Years)	15.42 (3.29)
Total Household Income ($)	79,053.56 (57,440.40)
Maternal Ethnicity (%)	
White/Black/Asian/Indian	76.8 / 21.3 / 1.5 / 0.4
Hispanic	4.8
Paternal Ethnicity (%)	
White/Black/Asian/Indian/Unknown	70.4 / 24.2 / 3.3 / 0.6 / 1.5
Hispanic	5.4

[*]NICU = Neonatal Intensive Care Unit

Table 2

Model of Missing Data

Parameter	Estimate	Standard Error	P-Value
Intercept	4.24	1.37	2.00E-03
ELC1	−0.02	0.01	0.08
Paternal Education in Years	−0.13	0.04	1.70E-03
Months Since Start of 1yr MSEL Testing	0.01	0.01	0.06
Scale	1.01	0.09	
n=361, 264 clusters			

Author Manuscript

Author Manuscript

Table 3

ELC Scores as Predictors of ABIQ6

	ELC1			ELC2			Both		
	Estimate	Standard Error	P-Value	Estimate	Standard Error	P-Value	Estimate	Standard Error	P-Value
ELC 1	0.16	0.06	0.0125				0.03	0.06	0.58
ELC 2				0.41	0.06	3.00 E-12	0.40	0.06	1.70E-10
QIC	1249			1199			1201		

Author Manuscript

Author Manuscript

Table 4

Full GEE Model Predicting ABIQ6

	Parameter	Estimate	Standard Error	P-Value
Mean	Intercept	107.28	1.59	< 2E-16
	ELC2 (centered)	0.28	0.06	1.50E-06
	Sex (Male)	−2.77	1.59	0.08
	Age in Days (centered)	0.04	0.02	7.64E-03
	Paternal Education in Years (centered)	1.18	0.25	2.30E-06
	Gest Number (Twin)	−6.11	1.61	1.50E-04
	Months Since Start of SB5 Testing	0.07	0.04	0.06
Scale	Intercept	119	13.1	

n=235, 174 clusters. *Note*: males and twins are base variables for binary sex and gestation number covariates.

Table 5

The Predictive Value of ELC Scores for ABIQ6 Across Samples

Sample	N	ELC1			ELC2		
		Estimate	SE	P-Value	Estimate	SE	P-Value
Full Sample	521	**0.16**	**0.06**	**0.012**	**0.41**	**0.06**	**3.00E-12**
32wks, 24hr NICU	405	**0.17**	**0.08**	**0.030**	**0.43**	**0.06**	**6.10E-15**
<32wks, >24hr NICU	116	0.11	0.11	0.309	0.26	0.18	0.150
Twins	175	**0.31**	**0.11**	**0.006**	**0.45**	**0.90**	**<0.001**
Singletons	230	0.12	0.09	0.210	**0.39**	**0.07**	**9.00E-08**

Note: Results compiled from GEE models using only ELC1 or ELC2 as predictors. Bolded and highlighted results are significant.

Table 6

Within Subject and Within-Family Correlations of ELC Scores

	GEE Correlations			Pearson Correlations	
	r	Standard Error	P-value	*r*	P-value
Same Child Over Time	0.30	0.05	1.97E-09	0.30	1.00E-11
Twins Same Age	0.70	0.06	< 2.2E-16	0.78	< 2.2E-16
Siblings Same Age	0.41	0.22	0.06	0.45	1.51E-04

Table 7

Full GEE Model Predicting ELC Scores

Parameter	Estimate	Standard Error	P-Value
Intercept	200.56	27.25	1.80E-13
Year 2	−71.97	35.60	0.04
Sex (Male)	−2.21	1.18	0.06
Gest Number (Twin)	−1.79	2.09	0.39
Gestational Age at Birth	−0.08	0.05	0.09
Paternal Education in Years	−0.53	0.19	7.01E-03
Age in Days	−0.16	0.05	1.02E-03
Months Since Start of 1yr MSEL Testing	0.15	0.02	6.50E-12
Year 2*Gest Number (Twin)	−9.26	2.46	1.60E-04
Year 2*Paternal Education in Years	1.61	0.25	1.20E-10
Year 2*Age in Days	0.14	0.06	0.02
Year 2*Months Since Start of 2yr MSEL Testing	−0.05	0.03	0.12

Note: Because the reference group for year (i.e. year 1 or year 2) is year 1, the coefficients for effects of year 2 are calculated by adding the coefficients for the single term (i.e. Gestation Number (Twin)) plus the coefficient for the interaction term (i.e. Year 2 * Gest Number (Twin)).