Published in final edited form as:
Res Dev Disabil. 2013 May ; 34(5): 1721–1730. doi:10.1016/j.ridd.2013.02.010.

Assessment and Stability of Early Learning Abilities in Preterm and Full-term Infants Across the First Two Years of Life

Michele A. Lobo, PT, PhD and
University of Delaware, Physical Therapy Department, Newark, DE 19716

James C. Galloway, PT, PhD
University of Delaware, Physical Therapy & Psychology Departments, Newark, DE 19716

Abstract

Infants born preterm have increased risk for learning disabilities yet we lack assessments to successfully detect these disabilities in early life. We followed 23 full-term and 29 preterm infants from birth through 24 months to assess for differences in and stability of learning abilities across time. Measures included the Bayley-III cognitive subscale, the mobile paradigm assessment, and a means-end learning assessment. Preterm infants had poorer performance on measures of cognition and learning across the first two years of life. Learning performance at 3–4 months was consistent with learning performance at 12–24 months of age. At 3–4 months, the mobile paradigm had better sensitivity and predictive values for predicting 24-month cognitive delays on the Bayley-III than did the Bayley-III itself. At 12–18 months, the means-end learning assessment had better sensitivity than the Bayley-III for identifying 24-month cognitive delays on the Bayley-III. The results suggest that: (1) infants born preterm may demonstrate learning differences as early as the first few months of life, (2) learning differences identified in the first months of life are likely to persist throughout the second year of life, and (3) learning assessments that measure how infants and toddlers use their typical behaviors to problem-solve to control external events may be more effective than traditional standardized assessment tools for detecting early learning delays.

Keywords

learning; preterm; assessment; early intervention; learning disability; developmental delay; cognitive delay

1. Introduction

An ongoing challenge for early intervention providers is the accurate identification of learning delays in the first two years of life. It is important to identify early learning delays because intervention provided in the first years of life can facilitate cognitive advancements (Orton, Spittle, Doyle, Anderson, & Boyd, 2009). Interventions should be provided as early as possible so there is a growing need for better early learning assessments (McManus, Carle, & Poehlmann, 2012; Nordhov, Ronning, Ulvund, Dahl, & Kaaresen, 2012). Current identification rates of delays are much lower than actual prevalence rates (Disabilities, 2006). Thus, the challenge of accurate identification remains. In this paper, we discuss the

---

challenges of using standardized normative assessment tools, the current standard of practice, to identify learning delays in the first years of life. We then introduce two early learning assessments from the developmental psychology literature: a 'mobile paradigm' assessment and a 'means end' learning assessment. Both have the potential to better identify early learning differences than the current standardized tests (Lobo & Galloway, 2008; Rovee & Rovee, 1969).

Infants born preterm would benefit from better early learning assessments. These infants have increased risk for cognitive, perceptual-motor, and behavioral problems (Jongbloed-Pereboom, Janssen, Steenbergen, & Nijhuis-van der Sanden, 2012). For example, they are more likely to have poorer academic performance in spelling, writing, language comprehension, mathematics, and physical education (Pritchard et al., 2009; Taylor et al., 2011). Although these infants as a group present with higher risk for future delays, many infants within the group will develop typically. Interventionists are therefore challenged with balancing the goals of: (1) not overburdening family and government resources with intense services for infants at risk who *do not* demonstrate delays, and (2) providing appropriate services as early as possible to infants who are at risk and *do* demonstrate delays. Assessment tools that accurately classify learning ability in very early development are critical to assist interventionists in balancing these two goals.

The assessments typically used to determine eligibility for early intervention have been standardized assessment tools that provide measures of performance relative to normative data sets. These tools typically involve scoring discrete tasks presented by an examiner using standardized equipment and specific instructions in a controlled environment. These assessments have a long history and can be useful in predicting future ability (Ramey, Campbell, & Nicholso.Je, 1973). For instance, one of the most prominent normative developmental assessments, the Bayley Scales of Development (three editions: Bayley, Bayley-II, Bayley-III), has been used since 1969. The Bayley has good psychometric properties when assessing children from the general population (Koseck, 1999). In effect, the Bayley has become the gold standard to which emerging assessments are graded and a diagnosis of delay is determined in research and clinical settings (Harris, Backman, & Mayson, 2010; Lung, Chen, & Shu, 2012; Vincer et al., 2005). For these reasons, one way we chose to measure cognition in this study was with the cognitive subscale of the most recent version of the Bayley, the Bayley-III.

Although standardized normative assessments can be useful for some populations, current findings suggest they have key limitations when used with very young populations at risk for delays. First, they are often good at identifying typical development but are poor at identifying atypical development. For instance, the Bayley Infant Neurodevelopmental Screener had low sensitivity (0–40%) but higher specificity (66–100%) for identifying whether infants would show delays at 24 months (Hess, Papas, & Black, 2004). When predicting 8-year cognitive abilities, both the Bayley-II and the Fagan Test of Intelligence at 1 year had poor sensitivity (16–32%) and poor positive predictive value (35–42%) but higher specificity (80–93%) and negative predictive value (78–79%) (McGrath, Wypij, Rappaport, Newburger, & Bellinger, 2004). Second, assessment scores often fluctuate in ways that do not likely reflect real change. For example, when the Bayley was administered on two occasions one week apart, the scores varied by more than one standard deviation in half of the cases (Horner, 1988). Third, the normative data used for assessments might not be an appropriate comparison across time even within the same culture due to the Flynn effect of gradually increasing scores across time (Vohr et al., 2012). This creates the need for updates to standardized tests that can be associated with significant changes in identification rates. For example, a recent report found that the Bayley-III identifies delays in significantly fewer infants born preterm than does the Bayley-II. Consequently, the authors recommended interpreting Bayley-III results with caution and enrolling all infants born preterm with extremely low birth weight for early intervention services (Flynn, 1987). Overall, these limitations suggest that traditional standardized assessments cannot be relied upon in isolation to accurately identify early delays in populations at risk.

Given the desire to provide intervention as early as necessary and the limitations of existing assessments, there is an urgent need for the development of better assessments to identify learning disabilities in the first two years of life. Moreover, there is a need to understand if and how early learning abilities relate to future learning abilities, particularly in high-risk populations. The purpose of this study was to begin to address these needs by determining: (1) whether infants born preterm and at risk for cognitive delays show differences in cognition and learning early in development, (2) whether learning abilities are stable across the first two years of life, and (3) whether exploratory, play-based assessments supported by developmental psychology literature can successfully identify early learning delays. To make these determinations, we followed infants born full-term and preterm from birth through 2 years. We measured their cognitive abilities using the Bayley-III and two learning assessments, the mobile paradigm assessment and a means-end learning assessment, that test whether infants can adapt their ongoing behaviors to problem-solve how to control an external event. We then looked at cognitive abilities relative to birth history and across time on these varied measures. We chose to use the mobile paradigm and means-end learning assessments because they have been used to assess learning abilities in developmental research for decades and they have the potential for immediate use by early intervention professionals given they are play-based, engaging, easy to administer, and inexpensive (Matthews, Ellis, & Nelson, 1996; Rovee & Rovee, 1969).

2. Methods

2.1. Participants

Fifty-two infants participated in this study. Twenty-three were born full-term and without any known medical diagnoses or delays. They were recruited from the local hospital and community. Twenty-nine were born preterm at less than 32 weeks of gestational age and were at increased risk for future delays and learning disabilities. These infants were recruited from a local neonatal intensive care unit. Please see Table 1 for more information about the infants in each group. Parents of all participants provided informed consent.

2.2. Data collection

This was a longitudinal design using repeated measures to track preterm infants born with increased risk of delays and to compare them to a control group of infants born full-term without any known biological risk. A trained pediatric physical therapist visited participants in their homes at 3, 4, 12, 18, and 24 months of age, using corrected ages for infants born preterm. At the 3- and 4- month visits, participants were assessed using the mobile paradigm assessment and the cognitive subscale of the Bayley-III. At the 12-, 18-, and 24-month visits, participants were assessed using the means-end learning assessment and the cognitive subscale of the Bayley-III. All assessments were video recorded using frontal and side views and coded by trained individuals blind to participants' group assignments and level of risk. Videos were temporally synchronized using Final Cut Pro software and were coded using MacSHAPA coding software. For the mobile paradigm and means-end coding assessments, coders were first trained until they achieved 85% inter-rater reliability with a primary coder for each assessment. Then each coder re-coded 20% of their visits as they progressed to ensure that their inter- and intra-rater reliabilities remained above 85%. Reliability was calculated using the equation [Agreed/(Agreed + Disagreed)] *100. Five research assistants completed all mobile paradigm assessment coding (inter-rater reliabilities 88.25–93.25%;

intra-rater reliabilities 87.33–93.58%). Four research assistants completed all mean-end learning assessment coding (inter-rater reliabilities 85.47–88.48%; intra-rater reliabilities 90.35–97.11%). All Bayley III scoring was completed by two individuals with advanced degrees and knowledge about child development who maintained >90% intra- and inter-rater reliabilities across 20% of their scored visits (inter-rater reliability 97.72%; intra-rater reliabilities 98.1 and 98.6%).

2.3. Assessments

2.3.1. Mobile paradigm assessment—Infants were provided one opportunity at 3 months and another at 4 months to demonstrate learning in this assessment. At each assessment, the infant was placed supine on the back in his/her crib with the right ankle tethered to an overhead mobile stand using a soft, fleece-lined ribbon (Figure 1A). In cases where families did not have cribs or their cribs did not fit our mobile stands, we used a mock crib we created out of plastic piping (Figure 1B). For the first 2 minutes (baseline), the mobile hung stationary above the infant from a stand other than the one with the tether. This period revealed how infants behaved before they had the opportunity to learn in the assessment, or when their kicks did not result in movement of the mobile. For the next 6 minutes (acquisition), the mobile hung from the stand to which infants were tethered. This period showed how infants behaved when they were provided the opportunity to learn that their kicks could elicit movement of the mobile. The final 2-minute period (extinction) mirrored the baseline, removing infants' abilities to control movement of the mobile.

2.3.2. Means-end learning assessment—The means-end learning assessment was performed at 12, 18, and 24 months (Lobo & Galloway, 2008; Munakata, Bauer, Stackhouse, Landgraf, & Huddleston, 2002). It required infants to activate lights and sound on a distant toy by simultaneously pressing two push-button switches (Figure 2). Participants were seated at a table with the switches directly in front of them and the toy about a foot beyond their reach. The toy was activated only so long as participants simultaneously held down both switches. In the first minute (baseline), simultaneous presses of the switches activated a light out of view of participants. This allowed coders to determine the amount of time participants held down both switches when they were not tied to another known event. For the next 3 minutes (acquisition), participants were provided the opportunity to learn to use their arm behaviors to activate the toy since simultaneous pressing of the switches would now activate the toy. The final minute (extinction) mirrored the baseline, with the switches again activating the light out of view rather than the toy.

2.3.3. Cognitive subscale of the Bayley-III—The cognitive subscale of the Bayley-III was administered at every visit. The Bayley-III is a norm-referenced assessment for children between 1 and 42 months of age (Bayley, 2006). It is commonly used in research and clinical practice to monitor development and to detect delays and is often considered the "gold standard" in early assessment tools (Vincer et al., 2005).

2.4. Variables

2.4.1. Mobile paradigm assessment—Throughout the mobile paradigm assessment, coders quantified kicking, visual attention, and affect. A kick occurred when infants moved the tethered leg into a position of hip and knee extension, tugging on the mobile stand. Visual attention to the mobile occurred when infants looked at the mobile for more than one second. Affect was coded using the AFFEX M system using facial expressions, postures, and sound production to rate affect (Izard, Dougherty, & Hembree, 1989). We combined interested, happy, and neutral affect under the umbrella of positive or neutral affect and sad, angry, and fearful affect under the umbrella of negative affect.

Learning in this assessment required that during at least one of the three 2-minute periods of the acquisition phase: 1) infants increased their kick rate to more than 1.5 times their baseline rate; 2) infants looked at the mobile more than two-thirds of the time; and 3) infants were in a positive or neutral affective state more than two-thirds of the time (Haley, Grunau, Oberlander, & Weinberg, 2008; Heathcock, Bhat, Lobo, & Galloway, 2004).

2.4.2. Means-end learning assessment—Throughout the means-end learning assessment, coders quantified light or toy activations, visual attention, and affect. Toy activation occurred during acquisition when the lights and music from the toy were active. Light activation occurred during baseline and extinction when the light was on. Visual attention to the toy occurred when the participant's eyes were directed towards the toy for more than one second. Visual attention to the switches occurred when the participant's eyes were directed towards the switches for more than one second. Affect was quantified as described in section 2.4.1.

Learning occurred when during acquisition: 1) participants had a normalized toy activation rate more than 1.5 times their baseline light activation rate; 2) participants looked at the toy more than 40% of the time; 3) participants were in a positive or neutral affective state more than two-thirds of the time (Lobo & Galloway, 2008).

2.4.3. Cognitive subscale of the Bayley-III—Experimenters scored each item of the cognitive subscale of the Bayley-III according to the manual. The basal level occurred when the participant achieved 1's for the first three items for an age group, meaning these behaviors were observed. The ceiling level occurred when the participant received five consecutive 0's on items, meaning these behaviors were not observed or were not performed correctly. Between the basal and ceiling levels, participants received 1's for behaviors observed and performed correctly and 0's for tested behaviors not observed or performed incorrectly. A total raw score was calculated by summing the scores. This raw score was converted to a normalized scaled score using Table A.1 of the Bayley III Administration Manual. The scaled scores range from 1–19, with a value of 10 representing the mean, lower scores representing poorer performance, and a value of 3 representing one standard deviation. Cognitive delay was defined as a scaled score more than 1.5 standard deviations below the mean (< 5.5) because this is a common cut-off for early intervention eligibility in many states within the USA.

2.5. Data analyses

2.5.1. Differences in learning and cognition in relation to gestational age at birth—We determined for each participant whether s/he demonstrated learning in the mobile paradigm assessment at 3 or 4 months and in the means-end learning assessment at 12, 18, or 24 months. If the participant demonstrated learning at any time point, s/he was recorded as having learned in that assessment. We summarized the data this way because participants demonstrated different developmental abilities in relation to one another and across time, so we could not expect the assessments to provide the appropriate challenge for mastery motivation at every time point for every individual (Redding, Morgan, & Harmon, 1988; Ruskin, Mundy, Kasari, & Sigman, 1994). At some visits for some children, the assessments would likely prove too challenging to elicit persistent and successful problem solving. At other visits for other children, especially for those who previously mastered the assessment, the assessments likely did not present sufficient challenge and interest to elicit engagement.

We then looked at the Bayley-III cognitive scores for each participant and determined if the highest score at 3 and 4 months was delayed, if the highest score at 12 and 18 months of age was delayed, and if the score at 24 months was delayed. We took the best of the scores from each time period in order to capture optimal performance.

We used Mann-Whitney tests to determine if infants born full-term and preterm differed in: 1) their learning abilities at 3–4 months in the mobile paradigm assessment; 2) their cognition at 3–4 month measured by the Bayley-III; 3) their learning abilities at 12–24 months in the means-end learning assessment; 4) their cognition at 12–18 months measured by the Bayley-III; and 5) their cognition at the end of the study (24 months) measured by the Bayley-III. Significance was set at ≤ .05. We used the z scores generated in the analyses to calculate effect sizes (*r*).

2.5.2. Relationship between performance during the mobile paradigm assessment at 3–4 months of age and the means-end learning assessment at 12–24 months of age—To determine if participants showed stability in their ability to learn in the mobile paradigm assessment at 3 or 4 months and in the means-end learning assessment at 12, 18, or 24 months of age, we used a paired samples McNemar test to compare whether participants ever demonstrated learning in the two assessments. Significance was set at ≤ .05.

2.5.3. Relationships between early measures of learning and cognitive outcome at 24 months of age—Using delay at 24 months on the cognitive subscale of the Bayley-III as the outcome measure, we looked at the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)[1] of: 1) learning performance at 3–4 months in the mobile paradigm assessment; 2) cognitive ability at 3–4 months measured by the Bayley-III; 3) learning performance at 12–18 months in the means-end learning assessment; and 4) cognitive ability at 12–18 months measured by the Bayley-III.

3. Results

3.1. Learning and cognitive performance of infants born full-term versus infants born preterm

Infants born preterm showed consistently poorer cognitive abilities across the timeline of the study compared to infants born full-term. At 3–4 months of age, there was a small to medium effect of birth status on infants' abilities to learn in the mobile paradigm assessment ($U = 90.5$, $z = -1.58$, $p = .06$, $r = -.28$). A smaller proportion of infants born preterm demonstrated learning in this assessment than did full-term infants (84.6% of full-term infants learned, 57.9% of preterm infants learned). At the same age, there was a small to medium effect of birth status on infants' cognitive abilities measured via the Bayley-III ($U = 276$, $z = -2.07$, $p = .02$, $r = -.29$). None of the full-term infants and 17.2% of the preterm infants were identified as having cognitive delays.

Infants born preterm showed cognitive differences relative to infants born full-term throughout the second year of the study as well. There was a medium size effect of birth status on participants' abilities to learn in the means-end learning assessment from 12–24

---

[1] *Sensitivity* measured the proportion of participants with delays at 24 months that were correctly identified earlier as having impaired cognitive abilities. It was calculated using the equation: [Number of True Positives/(Number of True Positives + Number of False Negatives)]. *Specificity* measured the proportion of participants with typical development at 24 months who were correctly identified earlier as having typical cognitive abilities. It was calculated using the equation: [Number of True Negatives/(Number of False Positives/Number of True Negatives)]. *Positive predictive value (PPV)* was the proportion of participants identified as having early cognitive impairments who actually had delays at 24 months of age. It was calculated using the equation: [Number of True Positives/(Number of True Positives + Number of False Positives)]. *Negative predictive value (NPV)* was the proportion of participants identified as having typical early cognitive abilities who actually had typical cognitive development at 24 months of age. It was calculated using the equation: [Number of True Negatives/(Number of True Negatives/Number of False Negatives)].

months of age ($U = 212$, $z = -2.69$, $p = .004$, $r = -.38$). 91.3% of the full-term infants demonstrated the ability to learn in this assessment compared to just 57.1% of the preterm infants. While there was only a small effect of birth status on cognition measured by the Bayley-III at 12–18 months of age, there was a medium effect at 24 months (12–18 months: $U = 276$, $z = -1.63$, $p = .052$, $r = -.23$; 24 months: $U = 230$, $z = -2.91$, $p = .002$, $r = -.40$). None of the full-term infants were identified as having cognitive delays at 12–18 months or at 24 months. For the preterm infants, 11.1% were identified as having cognitive delays at 12–18 months and 31.0% were identified at 24 months.

3.2. Consistency of learning ability across time

Participants demonstrated similar learning abilities during the mobile paradigm assessment at 3–4 months of age and the means-end learning assessment from 12–24 months of age ($p = .69$, reject the null hypothesis that learning ability was different in the two assessments). Learning performance was stable across the two assessments and age ranges for 80% of the participants.

3.3. Ability of the Bayley-III, mobile paradigm, and means-end learning assessments to detect early learning delays

The results suggest that exploratory, play-based learning assessments may be helpful in detecting cognitive delays in early development. In relation to 24-month outcome on the Bayley-III cognitive subscale, both the mobile paradigm assessment at 3–4 months of age and the means-end learning assessment at 12–18 months of age had much better sensitivities but lower specificities than the Bayley-III cognitive subscale (Table 2). All assessments had relatively high specificities and negative predictive values, suggesting that they are all successful at identifying participants with typical development. The sensitivity and PPV of the Bayley-III were quite low at 3–4 months of age. At 12–18 months of age, the PPV of the Bayley-III was perfect but the sensitivity remained low. The PPV of the mobile paradigm and means-end learning assessments were moderate and the sensitivities were moderate to high.

4. Discussion

4.1. Summary of the findings

Our results reveal several important findings about early learning in preterm infants. First, they provide further support for the idea that infants born preterm are at greater risk for learning disabilities. Preterm infants in this study had poorer performance on measures of learning and cognition across the first two years of life. Second, the results suggest that preterm infants may demonstrate these cognitive differences very early. Preterm infants in this study had poorer performance than full-term infants in the mobile paradigm assessment and on the Bayley-III cognitive subscale at just 3–4 months of age. Third, the results suggest that learning differences identified in infancy are likely to persist throughout the second year of life. Learning performance in the mobile paradigm assessment at 3–4 months was consistent with learning performance in the means-end learning paradigm at 12–24 months. Finally, the results suggest that learning assessments measuring how infants and toddlers use their typical movements to explore and problem-solve to control external events may be more sensitive for detecting early delays than more traditional standardized assessment tools. Both the mobile paradigm and means-end learning assessments were more sensitive than the Bayley-III for predicting 24-month cognitive delay. Below we discuss the main implications and significance of these findings.

4.2. Infants born preterm demonstrated learning differences early and consistently throughout the first two years of life relative to infants born full-term

The first purpose of this study was to determine whether infants born preterm and at risk for cognitive delays showed differences in cognition and learning early in their development. Preterm infants performed poorer on all measures of cognition across the first two years of life compared to full-term infants. This supports the tenet that infants born preterm are at greater risk for learning disabilities and suggests that learning differences exist and can be identified in this population even in the first months of life (Jongbloed-Pereboom et al., 2012). Only a handful of studies have identified learning differences in very young infants and children born preterm. The mobile paradigm has previously been shown to successfully identify poorer learning abilities in young infants born preterm (Gekoski, Fagen, & Pearlman, 1984; Heathcock et al., 2004). For instance, infants born full-term learned the mobile paradigm relationship within one session while infants born preterm did not show learning across 12 sessions (Heathcock et al., 2004). Infants born preterm show less learning as well as less time looking at the mobile and poorer memory than full-term infants in this assessment (Gekoski et al., 1984; Haley et al., 2008). Our study confirms these findings that very young preterm infants perform poorly on the mobile paradigm. In addition, as with our findings, infants born preterm have been shown to have lower cognitive scores on the Bayley in the first half year of life (Crnic, Ragozin, Greenberg, Robinson, & Basham, 1983). Therefore, our results support the small body of literature suggesting that learning differences can be identified in infants born preterm in the first months and years of life.

4.3. Learning differences identified in the first months of life persisted throughout the second year of life

The second purpose of this study was to determine whether learning abilities are stable across the first two years of life. The results suggest that the very early differences in learning evident in infants born preterm are important to identify because they are likely to persist. Learning abilities in this study were relatively stable for full-term and preterm infants throughout the first two years of life. Whereas there are only a handful of articles demonstrating early learning differences for infants born preterm, there are no studies assessing the stability of early learning abilities for this population. For infants with typical development, we know that stability has been observed for some behaviors and abilities in early development. For instance, in a learning paradigm where arm movements elicited pleasing visual and auditory reinforcement, emotional responses, such as interest and enjoyment during learning and anger when the association was broken, remained consistent within infants across a two-month period (Sullivan, Lewis, & Alessandri, 1992). In addition, there is stability between measures of cognition in preschool and measures of cognition in early childhood for individuals with typical development (Yang, Jong, Hsu, & Lung, 2011).

For infants born at risk for delays, there are conflicting findings regarding the stability of early learning abilities. On the one hand, Bayley-II scores at 2 years of age were consistent with intelligence scores at 4 years of age for children born preterm (Sajaniemi, Hakamies-Blomqvist, Katainen, & von Wendt, 2001). On the other hand, when the Bayley-II was used to classify delay in infants born preterm at 6, 12, and 18 months of age, the classifications varied across time within individuals with 85% of the cases showing instability (Janssen et al., 2011). Furthermore, cognitive development on the Bayley-II at 2–3 years of age only predicted 44–57% of intelligence at 5 years (Potharst et al., 2012). The present study begins to inform us about stability of early learning and demonstrates that exploratory, play-based learning assessments can identify persistent learning disabilities very early in development.

4.4. Exploratory, play-based learning assessments were more successful for identifying cognitive delays

The third purpose of this study was to determine whether exploratory, play-based assessments supported by developmental psychology literature could successfully identify learning delays in the first months and years of life. The results suggest that these assessments may be more sensitive than standardized assessments for identifying early learning delays. These assessments do not look at performance of discrete behaviors in isolated social and environmental contexts but focus on observation of active exploration and problem solving in more typical social and environmental contexts. Play-based assessments offer a number of benefits including: (1) exploration in a natural play setting rather than performance of discrete behaviors within a sterile, highly-structured, unfamiliar setting, (2) opportunity to demonstrate a broad range of functional performance versus demonstration of only behaviors within one's test window, (3) flexible format that can vary based on a child's interests rather than a rigid, predetermined sequenced format, and (4) identification of strengths, needs, and interests of the child rather than simply standardized scores (Kelly-Vance, Ryalls, & Glover, 2002).

There has been a gradual shift toward the use of play-based assessments in early intervention assessment with preschoolers. For instance, the Transdisciplinary Play-Based Assessment has become more common for arena evaluations and assessments of preschoolers for early intervention in some states (Kelly-Vance et al., 2002; Myers, McBride, & Peterson, 1996). It involves 30-minutes of non-directed play followed by a period of play aimed at eliciting behaviors not yet observed. It can be used from infancy through 6 years of age. It can be completed more quickly than typical standardized assessments. And it results in findings congruent with developmental ratings (Myers et al., 1996).

Naturalistic play-based assessments are less common in early intervention with infants and toddlers. Perhaps this is because there is less understanding of what constitutes "play" for young children, particularly infants. Meanwhile, the most commonly used standardized assessments lack sensitivity in the first year and a half of life (Hess et al., 2004; McGrath et al., 2004). Assessments focused on observation of spontaneous and elicited exploratory play behaviors are significantly more successful at identifying early learning delays. For example, the mobile paradigm assessment has been consistent in demonstrating learning delays in preterm infants in the first months of life (Gekoski et al., 1984; Haley et al., 2008; Heathcock et al., 2004). Our results identify these same learning differences in preterm infants and suggest that they relate to future learning differences in the second year of life. Furthermore, observations and ratings of spontaneous exploratory play behaviors (general movements) in the first months of life may be a useful tool for identification of early delays. Observations of spontaneous exploratory play behaviors in preterm infants at birth and 3 months showed sensitivities of 83–100% in relation to cognitive delays at 2–3 years (Kodric, Sustersic, & Paro-Panjan, 2010). Assessments focused on these types of spontaneous play behaviors in isolation and in the context of learning may improve our ability to detect early learning delays.

4.3. Implications, limitations, and future research

The results have important implications for early assessment and intervention. First, they suggest that parents of infants born preterm should be educated on how to provide their infants with enhanced learning experiences as soon as possible. These infants should be monitored closely in the first months and years of life. At the first sign of delay, appropriate interventions should be initiated rather than monitoring development to determine if the delays are transient. In accordance with the recommendations from the National Joint

Committee on Learning Disabilities, we should provide enhanced learning opportunities and appropriate intervention services as soon as a learning disability is encountered (Gartland & Strosnider, 2007). We should not wait with the hope that infants and children will outgrow these delays. Second, the results demonstrate that assessments involving opportunities for active exploration and problem solving can be more sensitive for detecting delays in infants than traditional standardized assessments. They join other recent research to suggest we should rethink the way we assess for early developmental delays, focusing on observations of exploratory and functional behaviors, play-based learning assessments, and input from parents and caregivers (Gartland & Strosnider, 2007; Kodric et al., 2010; Vohr et al., 2012).

This study provides a novel view of early learning abilities in a group of typical and high-risk infants throughout the first two years of life, but it has some limitations. First, the sample size was only moderate. A larger sample would have improved power and allowed for subgroup analyses to determine whether these assessments are better suited for certain individuals. It would also have allowed for analyses in relation to not only biological risk but also environmental risk. Second, although the results suggest that the mobile paradigm assessment and the means-end learning assessment may be helpful in early intervention practice, their implementation at the current time is limited because they are labor intensive to perform and score in real time using the current technology. For instance, it is not possible for one person to observe an infant’s visual attention to the mobile and count the number of tethered leg kicks simultaneously. Researchers should focus on using emerging technology to make these assessments feasible for real world application. For instance, software could be developed for use in combination with the relatively inexpensive Kinect motion-sensing device or other movement tracking devices to create smart assessment tools that code and categorize infants’ exploratory behaviors and learning performance across multiple sessions and contexts for real-time assessment of delays (Dutta, 2012). This would allow early intervention providers to use assessments like the mobile paradigm and means-end assessments with real time analysis of performance via small, inexpensive, portable computerized systems.

In terms of intervention, infants and children at risk often perform fewer, less variable movements and behaviors and demonstrate a lower level of general physical activity than their typically developing peers. This can result in delays in the onset of behaviors from reaching through walking, exploration, and problem solving (Lloyd, Burghardt, Ulrich, & Angulo-Barroso, 2010; Piek & Gasson, 1999). Future research should aim to determine how adaptations of the mobile paradigm, means-end learning tasks, and similar play-based exploratory learning activities can be used to motivate movement, play, and exploration in order to advance early development (Heathcock, Lobo, & Galloway, 2008; Lobo & Galloway, 2008; Oberg et al., 2012).

Acknowledgments

Funding Source

The authors acknowledge the following funding source for supporting all aspects of this research study: National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health & Human Development, 1R01HD051748 (Motor learning and coordination in high-risk infants. Principal investigator: Cole Galloway, PT, PhD, August 2007-June 2012).

References

Bayley, N. Bayley Scales of Infant Development. 3rd edition ed.. San Antonio, TX: The Psychological Corporation; 2006.

Crnic KA, Ragozin AS, Greenberg MT, Robinson NM, Basham RB. Social interaction and developmental competence of preterm and full-term infants during the 1st year of life. Child Development. 1983; 54(5):1199–1210. [PubMed: 6354627]

Disabilities, C. o. C. W. Identifying infants and young children with developmental disorders in the medical home: An algorith for developmental surveillance and screening. Pediatrics. 2006; 118(1): 405–420. [PubMed: 16818591]

Dutta T. Evaluation of the Kinect (TM) sensor for 3-D kinematic measurement in the workplace. Applied Ergonomics. 2012; 43(4):645–649. [PubMed: 22018839]

Flynn JR. Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin. 1987; 101:171–191.

Gartland D, Strosnider R. Learning disabilities and young children: Identification and intervention. Learning Disability Quarterly. 2007; 30(1):63–72.

Gekoski MJ, Fagen JW, Pearlman MA. Early learning and memory in the preterm infant. Infant Behavior & Development. 1984; 7(3):267–276.

Haley DW, Grunau RE, Oberlander TF, Weinberg J. Contingency Learning and Reactivity in Preterm and Full-Term Infants at 3 Months. Infancy. 2008; 13(6):570–595. [PubMed: 20717491]

Harris SR, Backman CL, Mayson TA. Comparative predictive validity of the Harris Infant Neuromotor Test and the Alberta Infant Motor Scale. Developmental Medicine and Child Neurology. 2010; 52(5):462–467. [PubMed: 19863637]

Heathcock JC, Bhat AN, Lobo MA, Galloway JC. The performance of infants born preterm and full-term in the mobile paradigm: Learning and memory. Physical Therapy. 2004; 84(9):808–821. [PubMed: 15330694]

Heathcock JC, Lobo M, Galloway JC. Movement training advances the emergence of reaching in infants born at less than 33 weeks of gestational age: A randomized clinical trial. Physical Therapy. 2008; 88(3):310–322. [PubMed: 18096650]

Hess CR, Papas MA, Black MM. Use of the bayley infant neurodevelopmental screener with an environmental risk group. Journal of Pediatric Psychology. 2004; 29(5):321–330. [PubMed: 15187171]

Horner TM. Single versus repeated assessments of infant abilities using the Bayley Scales of Infant Development. Infant Mental Health Journal. 1988; 9(3):209–217.

Izard, CE.; Dougherty, LM.; Hembree, EA. A System for Identifying Affect Expressions by Holistic Judgements (Affex). The University of Delaware; 1989.

Janssen AJWM, Akkermans RP, Steiner K, de Haes OAM, oostendorp RAB, Kollee LAA, Nijhuis-van der Sanden MWG. Unstable longitudinal motor performance in preterm infants from 6 to 24 months on the Bayley Scales of Infant Development - second edition. Research in Developmental Disabilities. 2011; 32:1902–1909. [PubMed: 21530160]

Jongbloed-Pereboom M, Janssen A, Steenbergen B, Nijhuis-van der Sanden MWG. Motor learning and working memory in children born preterm: A systematic review. Neuroscience and Biobehavioral Reviews. 2012; 36(4):1314–1330. [PubMed: 22353425]

Kelly-Vance L, Ryalls BO, Glover KG. The use of play assessment to evaluate the cognitive skills of two- and three-year-old children. School Psychology International. 2002; 23(2):169–185.

Kodric J, Sustersic B, Paro-Panjan D. Assessment of general movements and 2.5 year developmental outcomes: Pilot results in a diverse preterm group. European Journal of Paediatric Neurology. 2010; 14(2):131–137. [PubMed: 19540782]

Koseck K. Review and evaluation of psychometric properties of the revised Bayley Scales of Infant Development. Pediatric Physical Therapy. 1999; 11:198–204.

Lloyd M, Burghardt A, Ulrich DA, Angulo-Barroso R. Physical Activity and Walking Onset in Infants With Down Syndrome. Adapted Physical Activity Quarterly. 2010; 27(1):1–16. [PubMed: 20147766]

Lobo MA, Galloway JC. Postural and object-oriented experiences advance early reaching, object exploration, and means-end behavior. Child Development. 2008; 79(6):1869–1890. [PubMed: 19037955]

Lung FW, Chen PF, Shu BC. Predictive and construct validity of the Bayley Scales of Infant Development and the Wechsler Preschool and Primary Scale of Intelligence with the Taiwan Birth Cohort Study Instrument. Psychological Reports. 2012; 111(1):83–96. [PubMed: 23045850]

Matthews A, Ellis AE, Nelson CA. Development of preterm and full-term infant ability on AB, recall memory, transparent Barrier Detour, and Means-End tasks. Child Development. 1996; 67(6): 2658–2676. [PubMed: 9071757]

McGrath E, Wypij D, Rappaport LA, Newburger JW, Bellinger DC. 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. 2004; 114(5):E572–E576. [PubMed: 15492354]

McManus BM, Carle AC, Poehlmann J. Effectiveness of Part C Early Intervention Physical, Occupational, and Speech Therapy Services for Preterm or Low Birth Weight Infants in Wisconsin, United States. Academic Pediatrics. 2012; 12(2):96–103. [PubMed: 22230187]

Munakata Y, Bauer D, Stackhouse T, Landgraf L, Huddleston J. Rich interpretation vs. deflationary accounts in cognitive development: The case of means-end skills in 7-month-old infants. Cognition. 2002; 83:B43–B53. [PubMed: 11934406]

Myers CL, McBride SL, Peterson CA. Transdisciplinary, play-based assessment in early childhood special education: An examination of social validity. Topics in Early Childhood Special Education. 1996; 16(1):102–126.

Nordhov SM, Ronning JA, Ulvund SE, Dahl LB, Kaaresen PI. Early Intervention Improves Behavioral Outcomes for Preterm Infants: Randomized Controlled Trial. Pediatrics. 2012; 129(1):E9–E16. [PubMed: 22184645]

Oberg GK, Campbell SK, Girolami GL, Ustad T, Jorgensen L, Kaaresen PI. Study protocol: an early intervention program to improve motor outcome in preterm infants: a randomized controlled trial and a qualitative study of physiotherapy performance and parental experiences. Bmc Pediatrics. 2012; 12

Orton J, Spittle A, Doyle L, Anderson P, Boyd R. Do early intervention programmes improve cognitive and motor outcomes for preterm infants after discharge? A systematic review. Developmental Medicine and Child Neurology. 2009; 51(11):851–859. [PubMed: 19732117]

Piek JP, Gasson N. Spontaneous kicking in fullterm and preterm infants: Are there leg asymmetries? Human Movement Science. 1999; 18(2–3):377–395.

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 and Child Neurology. 2012; 54(3):240–246. [PubMed: 22188215]

Pritchard VE, Clark CAC, Liberty K, Champion PR, Wilson K, Woodward LJ. Early school-based learning difficulties in children born very preterm. Early Human Development. 2009; 85(4):215–224. [PubMed: 19022593]

Ramey CT, Campbell FA, Nicholso.Je. Predictive power of the Bayley Scales of Infant Development and Stanford-Binet Intelligence Test in a relatively constant environment. Child Development. 1973; 44(4):790–795.

Redding RE, Morgan GA, Harmon RJ. Mastery motivation in infants and toddlers - Is it greatest when tasks are moderately challenging. Infant Behavior & Development. 1988; 11(4):419–430.

Rovee CK, Rovee DT. Conjugate reinforcement of infant exploratory behavior. Journal Of Experimental Child Psychology. 1969; 8:33–39. [PubMed: 5804591]

Ruskin EM, Mundy P, Kasari C, Sigman M. Object mastery motivation of children with Down syndrome. American Journal on Mental Retardation. 1994; 98(4):499–509. [PubMed: 8148126]

Sajaniemi N, Hakamies-Blomqvist L, Katainen S, von Wendt L. Early cognitive and behavioral predictors of later performance: a follow-up study of ELBW children from ages 2 to 4. Early Childhood Research Quarterly. 2001; 16(3):343–361.

Sullivan MW, Lewis M, Alessandri SM. Cross-Age Stability in Emotional Expressions During Learning and Extinction. Developmental Psychology. 1992; 28(1):58–63.

Taylor HG, Klein N, Anselmo MG, Minich N, Espy KA, Hack M. Learning problems in kindergarten students with extremely preterm birth. Archives of pediatrics & adolescent medicine. 2011; 165(9):819–825. [PubMed: 21893648]

Vincer MJ, Cake H, Graven M, Dodds L, McHugh S, Fraboni T. A population-based study to determine the performance of the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale to predict the mental developmental index at 18 months on the Bayley Scales of Infant Development-II in very preterm infants. Pediatrics. 2005; 116(6):E864–E867. [PubMed: 16263973]

Vohr BR, Stephens BE, Higgins RD, Bann CM, Hintz SR, Das A. Eunice Kennedy Shriver Natl, I. Are Outcomes of Extremely Preterm Infants Improving? Impact of Bayley Assessment on Outcomes. Journal of Pediatrics. 2012; 161(2) 222-+.

Yang P, Jong YJ, Hsu HY, Lung FW. Role of assessment tests in the stability of intelligence scoring of pre-school children with uneven/delayed cognitive profile. Journal of Intellectual Disability Research. 2011; 55:453–461. [PubMed: 21375639]

Highlights

- We assessed learning abilities from birth through 2 years in full-term and preterm infants.
- Learning differences were evident in preterm infants by 3 months of age.
- Learning ability remained similar for individuals from infancy through toddlerhood.
- Novel problem-solving assessments were best at identifying infants with delays.

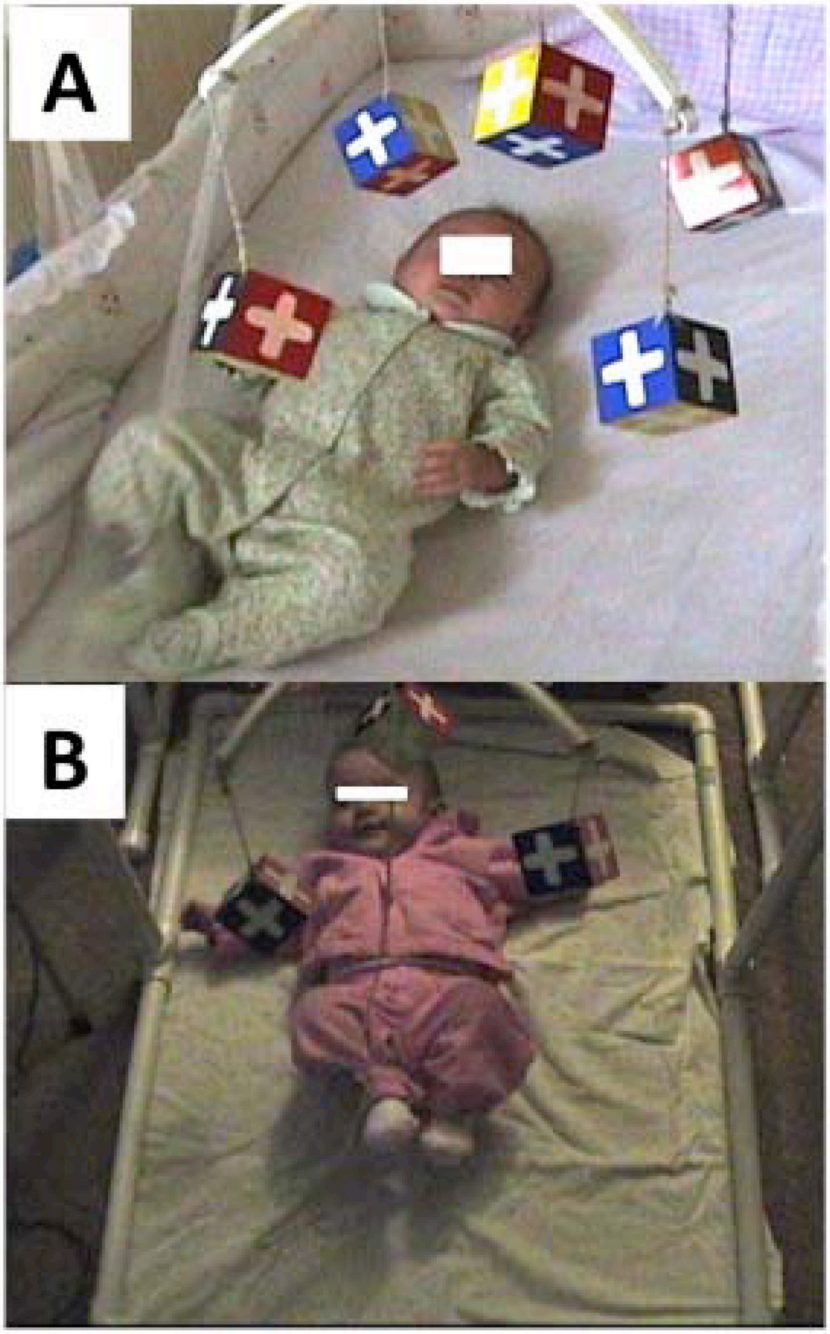

Figure 1.

In the mobile paradigm assessment, we determined if infants 3 to 4 months of age could learn that kicks of their tethered leg would result in movement of an overhead mobile. We used infants' cribs when they were available and fit the mobile apparatus (A). We used supports made of plastic piping in other instances (B).[2]

---

[2] Parental consent has been obtained for sharing of all photographs printed in this article.

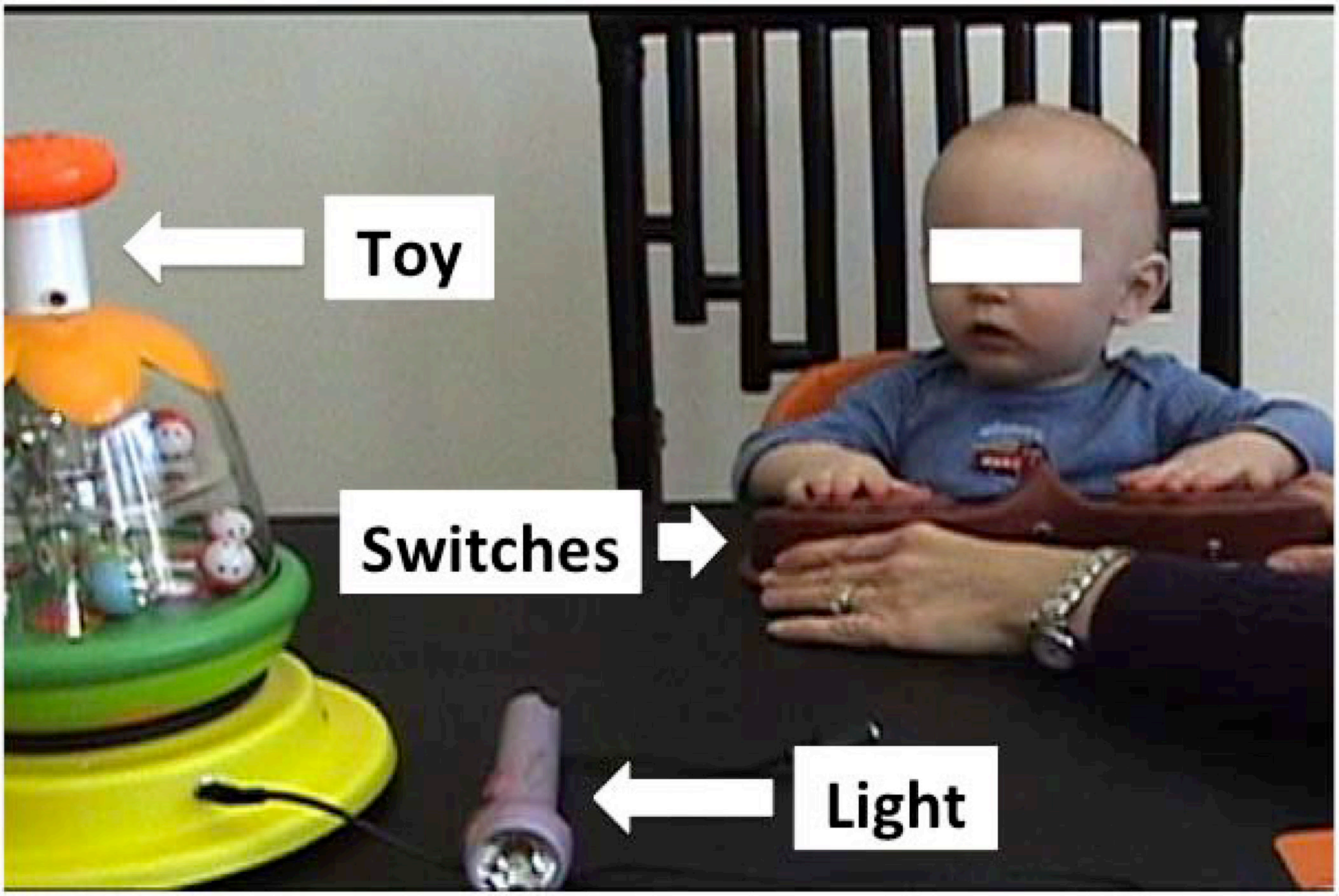

Figure 2.
In the means-end learning assessment, we determined if infants 12 to 24 months of age could learn to activate lights and sound on a distant toy by simultaneously pressing down two proximal switches with their hands.

Table 1

Demographic data

Characteristic	Full-term Group	Preterm Group
Number of males	13 (57%)	10 (34%)
Number of females	10 (43%)	19 (66%)
Number of Caucasians	17 (74%)	12 (41%)
Number of Blacks	4 (17%)	13 (45%)
Number of Asians	2 (9%)	4 (14%)
Number of Hispanics	0 (0%)	4 (14%)
Age at birth (weeks)	39.4 ± .2	26.7 ± .3
Number born very preterm (<32 weeks)	0 (0%)	10 (34%)
Number born extremely preterm (<28 weeks)	0 (0%)	19 (66%)
Weight at birth (grams)	3352.8 ± 141.4	931.6 ± 47.9
Number with very low birth weight (<1500 grams)	0 (0%)	10 (34%)
Number with extremely low birth Weight (<1000 grams)	0 (0%)	19 (66%)

Table 2

Sensitivities, specificities, positive predictive values (PPV), and negative predictive values (NPV) in relation to the outcome of cognition at 24 months measured by the Bayley-III.

Age	Assessment	Sensitivity	Specificity	PPV	NPV
3 to 4 months	Bayley-III cognitive subscale	.11	.91	.2	.83
	Mobile paradigm assessment	.72	.77	.45	.91
1 to 2 years	Bayley-III cognitive subscale	.33	1	1	.87
	Means-end learning assessment	.89	.64	.35	.96