The Effects of Bilingualism on Executive Functioning in Infants

Hannah Frank

Mentor: Dr. Rachel Barr

Georgetown University
Abstract

Children, young adults, and the elderly have demonstrated a “bilingual advantage” in cognitive flexibility (Adi-Japha, Berberich-Artzi, & Libnawi, 2011). Specific bilingual advantages have been found for all three components of executive functioning (EF) – inhibition, working memory (WM), and cognitive flexibility (Feng, Bialystok, & Diamond, 2009; Kovacs & Mehler, 2009). EF has not previously been investigated in toddlers due to a difficulty with finding age-appropriate tasks. However, WM tasks for 18- to 30-month-olds have recently been developed, allowing researchers to investigate EF at an age previously not investigated (Hughes & Ensor, 2005; Bernier, Carlson & Whipple, 2010). In the present study, 18- and 24-month-old bilingual and monolingual infants were given age-appropriate EF tasks. Given previous research on bilingual cognitive advantages of bilingualism (Bialystok, Craik Green, & Gollan, 2009), we predicted that bilinguals would outperform monolinguals on both WM tasks. Our results did not show a bilingual advantage in WM at this age, suggesting a need for additional research on EF tasks for young infants, as well as additional longitudinal analysis of these tasks.
The Effects of Bilingualism on Executive Functioning in Infants

Executive functioning (EF) refers to the set of higher order cognitive processes that underlie flexible goal-directed behavior (Hughes & Ensor, 2005). Best and Miller (2010) discuss different theoretical frameworks for EF, one of which divides EF into three interrelated but distinct components – attentional flexibility, inhibitory control and working memory. EF skills across these three domains are necessary for expanding cognitive, social and psychological capacities in children and adults. Higher-order executive functioning abilities play an integral role in daily activities, including success in school (i.e. math and language skills), organization and time management. More broadly, higher order executive functioning abilities include skills such as problem solving, reasoning and planning. These skills rely on EF to control thoughts and behavior in areas such as selective attention and response inhibition. In order to maintain cognitive control and to succeed at the aforementioned higher-order cognitive processes, EF abilities must be strong (Blakemore & Choudhury, 2006). Understanding the trajectory of EF development is critical for ensuring positive developmental outcomes.

Executive functioning is dependent on the prefrontal cortex (PFC) (Shimamura, 2000), an area of the brain that is not fully developed until late adolescence (Toga, Thompson & Sowell, 2006). Investigating the course of development of EF is important, though it is often difficult to assess in young children for a number of reasons. First, the PFC is in its early stages of development prior to adolescence, indicating a long developmental trajectory that coincides with that of EF. When investigating EF in young children, it is necessary to find tasks that account for the developmentally early nature of the PFC. In addition, it is often difficult to verify that tasks used for young infants are measuring the areas of cognitive processing that they are intended to measure. This difficulty is compounded by the fact that some areas of EF that are
differentiated in later adulthood may not yet be differentiated in early childhood. It is necessary
to be creative in designing tasks that accurately reflect cognitive abilities, particularly when
studying EF. The three components of EF (attentional flexibility, inhibitory control and working
memory) have been studied in children, but few tasks have been developed to measure these
abilities in children younger than three years old.

One of the areas given most recent attention in terms of task development for young
children is working memory (WM). Working memory requires a cognitive representation and
manipulation of information. Baddeley’s model of working memory identifies WM as a control
mechanism (the executive control) with two storage buffers, the phonological loop and the
visuospatial sketchpad that are used to manipulate information (Baddeley, 1986). The executive
control commands the two buffers, selecting and maintaining information in temporary storage
(Shimamura, 2000). Recent research had led to an increase in the number of tasks available to
measure EF and WM in young children, though the development of such tasks has a long history.

Many early EF tasks were designed for testing with non-human primates with brain
lesions. In 1965, Pinto-Hamuy and Linck investigated the effect of frontal lesions on
performance of sequential tasks by monkeys. When tested on an externally ordered sequence
test, in which they had to push on a series of panels as determined by external cues, monkeys
with and without frontal lobe lesions did not perform differently on the task. Milner (1971)
devised a similar task, the Corsi blocks task, which also used an externally ordered sequence test
as a measure of visuospatial WM. The forward-span version of the Corsi blocks task requires
participants to tap a series of irregularly positioned blocks in the same order as the experimenter.
The reverse Corsi blocks task requires the participant to tap the blocks in the opposite order
demonstrated by the experimenter. Both versions of the Corsi blocks task, as well as Pinto-
Hamuy and Linck’s externally ordered task measured WM with external cues. The availability of external cues in these tasks might be a critical factor that allows this task to be completed regardless of frontal lobe functioning.

On the second part of Pinto-Hamuy and Linck’s (1965) study, they used an internally ordered task in which monkeys had to push on all of several cued panels without repetition. As opposed to the externally ordered task, monkeys with frontal lesions had difficulty with the internally ordered task. While both tasks required WM, the externally ordered task offered external spatial cues that could also be used to solve the task, whereas the internally ordered task did not. The internally ordered task measured a number of components of EF, particularly planning abilities, since the task required an internal representation of a behavioral sequence in order to succeed. The fact that this task is contingent upon frontal lobe functioning is unsurprising, given significant subsequent research demonstrating the role of the frontal lobe, particularly the PFC in executive functioning (e.g. Diamond, 1990). In addition to prompting additional work on the role of the PFC in EF, Pinto-Hamuy & Linck’s (1965) study formed the basis for Petrides and Milner’s (1982) self-ordered pointing task (SOPT). This task was used to measure deficits on self-ordered tasks after frontal- and temporal lesions in man, as opposed to lesions in non-human primates. The SOPT requires subjects to point to unique abstract items on each page of a booklet without repeating any of the objects. This task again demonstrated the role of the frontal lobe in EF, particularly when completing internally ordered WM tasks.

Following the development of the SOPT, which gave more specific insight into cognitive processes involved in EF, Petrides (1988) devised a working memory task for monkeys called the Multiple Boxes Task. This task required subjects to retrieve a reward from each of three different shaped and colored boxes without choosing any of the same boxes twice. It was
similar to the SOPT in that it was a self-ordered task that required subjects to open all of the boxes without repeating a choice. Unsurprisingly, success on this task relied on the dorsolateral PFC in the frontal lobe, the same region required for success on the SOPT. Like SOPT and Pinto-Hamuy and Linck’s (1965) task, the Multiple Boxes Task relied on internally ordered processing and WM.

Diamond, Prevor, Callender and Druin (1997) used a stationary and scrambled version of the Multiple Box Task, as well as a six-box variation of the task, to study cognitive functioning in children with moderate PKU. This task was part of a battery of EF tests that Diamond used to investigate cognitive abilities in children ages 6 months – 7 years. Diamond and colleagues’ (1997) findings indicated that performance on the stationary box task was not affected by frontal lobe impairments. This finding can likely be attributed to the fact that the task could be completed using spatial cues in addition to WM. However, individuals with frontal lobe impairments were unsuccessful at the scrambled version of the task, likely because the removal of visual cues meant that task performance relied solely on WM abilities. Despite the differences in the scrambled and stationary versions of the tasks, they both rely on the same internally self-ordered component as Pinto-Hamuy and Linck’s (1965) original task.

In the same way that self-ordered tasks require participants to inhibit previous responses, Hughes and Ensor’s (2005) Spin the Pots task was developed to assess WM and inhibition in young children. This task was one of five executive functioning tasks used to study associations between executive functioning and theory of mind in two-year-olds. This multi-location search task was originally designed for children between 24 and 36 months of age. Like Pinto-Hamuy & Linck’s (1965) internally ordered sequence task, Spin the Pots participants were allowed to choose any cup for the first response, but whether the next responses were correct were
contingent on the previous action. In addition, like the scrambled version of the Multiple Box Task (Petrides, 1988), fixed spatial cues could not be used due to the rotation of the cups.

Bernier, Carlson and Whipple (2010) recently adapted the Spin the Pots task for 18-month-olds. The 18-month-old Hide the Pots task uses three distinctly colored pots from which the participant has to retrieve an item hidden by an experimenter. In this task, WM is activated on each trial when children update their mental representation of where the ball is located. This task is similar to the stationary Multiple Box Task in terms of the appearance of the apparatus, but it differs in that only one item is hidden at a time and the search for the item is not self-ordered. Because the cups are not moved, spatial cues can be used to solve the Hide the Pots task, while they are not available when solving the Spin the Pots task. Moreover, Spin the Pots is internally ordered, while Hide the Pots is not. These minute differences may have implications in terms of overall performance on each of the tasks.

In addition to the Multiple Boxes Task, Diamond et al. (1997) measured cognitive functioning in children with PKU with age-appropriate variations of the Stroop Task and the A-not-B task, among others outlined in Appendix C. Most relevantly, the A-not-B task has frequently been adapted to allow for testing of WM across a broad spectrum of age groups. Diamond’s research (1990) with adult and infant non-human primates demonstrated that the dorsolateral PFC is a major contributor to performance on the original A-not-B task. A looking version of the A-not-B task has also been developed for young, pre-verbal children. With this task, children are required to “search” for a hidden toy by making an eye movement to one of two possible hiding locations (e.g., Bell and Adams, 1999). Much like the reaching version of the A-not-B working memory task (Diamond, 1990; Diamond et al., 1997), the looking version requires infants to constantly update their memory of where the toy was hidden through a series
of displacements. Simultaneously, they must inhibit looking back toward a previously rewarded hiding place (Bell and Adams, 1999). Bell (2012) recently conducted a study using the looking version of the A-not-B task with 8-month-old infants, providing more evidence for the role of the frontal lobe in working memory. Cuevas & Bell (2011) demonstrated even earlier changes in brain activation as a function of WM processes using the looking version of the A-not-B task in 5-month-old infants. When being tested on this task, infants engage WM by forming and updating temporary representations of objects and their locations. However, this varies from the adult version of the A-not-B task in that adults completing the task rely on verbal instructions and an unambiguous set of WM processes. The HTP and STP tasks are differentiated from the infant version of the A-not-B task in that task completion does involve verbal directions. This difference might have important implications regarding language comprehension abilities, which are liable to vary in 18- and 24-month children.

Despite the difficulty of differentiating between cognitive processes in young children, both the Hide the Pots and Spin the Pots task have been used to evaluate the development of EF in infants. Both tasks activate WM when children update representations of where objects are located. The development of these tasks has enabled researchers to examine EF development in younger infants than previously studied (e.g. Bernier, Carlson, Deschenes, Matte-Gagné, 2011). As a result, there is a growing body of research on mediating components in the development of EF. Carlson (2009) discussed the role of social and cultural interactions, such as parental discipline style and scaffolding, in toddler’s development of EF. Similarly, Bernier and colleagues (2010) found that higher quality parenting would predict better child EF performance. Despite the increasing amount of research on EF development, particularly on parent-child interactions and EF, only one study has examined the effects of bilingualism as a mediating
component of EF in infants (Poulin-Dubois et al., 2010). Early effects of bilingualism on EF is a particularly relevant topic given that two-thirds of children around the world grow up learning two languages simultaneously (Werker & Byers-Heinlein, 2008). However, until recently, tasks were not even available to measure the effects of EF on bilingualism in children younger than 3 years old.

Despite previous misconceptions, recent research has rapidly emerged supporting the idea that bilingualism is an advantage rather than a disadvantage (e.g. Diamond, 2010; Westly, 2011). In addition to reaching the same developmental milestones as monolingual children, bilinguals differentiate languages early and can use them appropriately by 19 months (Nicoladis & Genesee, 1996). Furthermore, bilinguals have been shown to have particular life-long advantages in cognitive flexibility and other components of EF as a result of their sensitivity and awareness to different languages (e.g. Bialystok, Craik & Luk, 2008). Enhanced cognitive skills lead to a developmental advantage in nonlinguistic tasks that necessitate cognitive flexibility (i.e. Carlson & Meltzoff, 2008), as well as an increased capacity for learning and problem solving. These advantages are truly lifelong, as evidenced by Craik, Bialystok & Freedman’s (2010) study demonstrating the role of bilingual cognitive advantages in delaying the onset of Alzheimer’s symptoms by 5 or 6 years (Craik et al., 2010). Given these impressive bilingual advantages, a number of theories have been proposed to explain them.

Perhaps the most relevant explanation for cognitive bilingual advantages is Green’s (1998) inhibitory control model. Green (1998) proposed that the representation of both languages is distinct within the bilingual child, but that both languages remain active during the use of either language. Green (1998) further suggested that it is the inhibitory processes that are responsible for suppressing the irrelevant language. At a young age, these inhibitory processes
are being developed, leading to the common erroneous perception that language mixing is confusion on the part of the child. Instead, this serves as an indication of the constant practice bilingual children obtain in gaining inhibitory control. If this model is accurate, then bilingual children experience extensive practice of inhibitory functions from early in development and should perform better on executive functioning tasks. Inhibition relates to WM, in particular, because of the need to remember the linguistic goal in order to know what to inhibit. WM is also required when completing a linguistic goal (e.g. retrieving a word from one language or the other) because bilingual individuals must simultaneously inhibit distractions to concentrate and process the relevant information. This updating and manipulation of information demonstrates the link between inhibition and WM in bilinguals.

As mentioned previously, the only study to date examining the effects of bilingualism on EF in young children was Poulin-Dubois and colleagues (2011) study with 24-month-olds. EF performance was assessed using a battery of tests including a multilocation search task, shape Stroop, reverse categorization and a set of delay tasks. Twenty-four-month-old bilingual children outperformed their monolingual counterparts on the Stroop task, an inhibition task. This finding necessitates further research on specific bilingual EF advantages at this age, particularly in the context of Green’s (1998) inhibitory control model. In order to further investigate EF differences, the *Hide the Pots* and *Spin the Pots* task were selected for the present study as age-appropriate tasks to measure WM in bilingual infants as young as 18-months.

This study will attempt to replicate and extend the findings of Bernier and colleagues (2010) and Hughes and Ensor (2005) in a longitudinal study examining individual differences on performance in these tasks. Previous studies have examined EF in young children (e.g. Poulin-Dubois, 2011), but this study will investigate WM in both monolingual and bilingual children at
an even earlier age (before three years old). This study will explore whether there are age-related differences in a WM task for monolingual and bilingual infants. In addition, error patterns during the tasks will be assessed in detail. Analysis of error patterns have yielded important findings in other executive functioning tasks with older children and adults (e.g. Bull & Scerif 2001) but to date have not been examined within these tasks or at such a young age.

**Methods**

**Participants** - Seventy-eight participants were recruited through community fairs, flyers and postings on online parent groups. There were 22 longitudinal participants who completed the *Hide the Pots* task at 18 months and the *Spin the Pots* task at 24 months. An additional 43 18-month participants completed just the *Hide the Pots* task and another 13 24-month participants completed just the *Spin the Pots* task. To be considered bilingual, infants had to have at least one parent who is a non-native English speaker and have heard two languages since birth (“crib bilinguals”). Children had to have at least 25% exposure to the second language to be considered bilingual. Thirty-four participants in this sample were bilingual. A parent-administered questionnaire determined children’s exposure to each language.

**Language Exposure.** Productive vocabulary was assessed via parent-report using the short form of the MacArthur-Bates Communicative Developmental Inventory (MCDI) (Fenson et al., 2000). The MCDI is a 100-item vocabulary list administered in order to indicate the child’s productive vocabulary in each language spoken. The parents indicate which words on the list their child produces. Bilingual parents also reported the words in the second language that their child spoke. Given the number of language that were spoken by participants, it was not possible to have a separate language inventory for each language.
Design – A 2 (Age: 18- and 24-months) x 2 (Experimental: monolingual, bilingual) mixed design was employed. Vocabulary (indexed via MCDI scores), parent’s education and SES were added as covariates to the overall model. The overarching purpose of the study was to assess working memory abilities in bilingual and monolingual infants. In order to best assess these abilities, the study was broken down into three smaller experiments. Experiment 1a included all 18-month participants, both longitudinal and not, and measured overall patterns of executive functioning performance in bilinguals compared to monolinguals. In order to assess executive functioning, 18-month old participants completed the *Hide the Pots* task. Experiment 1b included all 24-month participants, longitudinal and not, but used the *Spin the Pots* task as measure of executive functioning. Experiment 1c included participants who completed EF tasks at both 18 and 24 months. These longitudinal participants completed the *Hide the Pots* task at 18 months and the *Spin the Pots* task at 24 months. We compared their data across time in a mixed design with repeated measures of working memory across time. This allowed us to gain a better understanding of when monolingual and bilingual differences in executive functioning emerge. Based on current bilingual literature providing evidence of bilingual advantages in executive functioning (Poulin-Dubois, Blaye, Coutya, & Bialystok, 2010), we expected that bilinguals would outperform monolinguals on both the *Hide the Pots* task and the *Spin the Pots* tasks.

**Experiment 1a**

Experiment 1a included only the 18-month participants who completed the *Hide the Pots* task described above. This included both longitudinal and non-longitudinal participants.

**Methods**

**Participants** - The sample for the 18-month *Hide the Pots* task consisted of 65 monolingual (n = 37) and bilingual (n = 24) participants (34 males and 31 females). The mean
rank of socioeconomic index (SEI, Nakao & Treas, 1992) for the families who completed this task was 66.34 ($SD = 20.69$). Education was ranked on a 6-point scale as follows: 1 – 8th grade or below, 2 – some high school, 3 – high school, 4 – some college, 5 – college, 6 – graduate school. The mean amount of education completed by parents was through college level ($M = 5.16$, $SD = 1.34$). Participants were African-American ($n = 5$), Asian ($n = 2$), Caucasian ($n = 32$), Latino ($n = 8$), of mixed ethnic origin ($n = 17$), and one family did not report. Four participants were dropped either due to equipment failure ($n = 3$) or due to sibling interference ($n = 1$).

**Apparatus** – The *Hide the Pots* (Bernier et al., 2010) task was used for the 18-month olds. Three distinctly colored pots were used for this portion of the study. All three cups fit under a box, which was closed at one point during each of three trials. During each trial, a small ball was hidden under each of the cups, in turn. See Figure 1 in Tables and Figures for images of the apparatus. Productive vocabulary was assessed via parent-report using the short form of the MacArthur-Bates Communicative Developmental Inventory (MCDI) (Fenson et al., 2000).

**Procedure** – During the 18-month visit, the child completed the *Hide the Pots* task. A warm-up phase acquainted the child with the apparatus and the expectations for the trials. During the test phase, a small ball was placed under one of three opaque pots of different colors. A lid then covered the pots for two seconds before the child was instructed to find the ball on each trial. There were a total of three trials, with the ball hidden under a different cup during each trial. The order in which the ball was hidden was counterbalanced across individuals. Each trial required the child to hold the location of the ball in memory and then reach for the pot where the ball was hidden. Parents were asked to complete a general questionnaire (including contact information, education, career, parental education, child language exposure and media exposure). In addition, they were asked to complete an MCDI in order to indicate the child’s
productive vocabulary in each language spoken.

**Results**

**Coding and Reliability** - All data collection visits were videotaped for later coding. The coding schemes were extended based on examination of the adult and comparative data (e.g. Petrides 1982, Ross et al., 2007). Performance on the *Hide the Pots* task was given a score (range = 0-3). This score was calculated based on the number of trials in which the child selected the correct pot on the first search attempt. In addition, performance on this task was coded for a number of errors. The following errors were coded from the videotaped version of the visit: perseveration, defined as incorrectly choosing the cup that was selected on the previous trial (max = 2); corrections, defined as touching an incorrect cup and then picking up the correct cup (max = 2); and picking up two cups simultaneously (max = 3). Multiple codes could apply to one trial. Latency was recorded as the lapsed time between when the lid was raised until the child picked the first cup. Total time on the task was also recorded. See Appendix A for the entire *Hide the Pots* coding protocol.

Based on 26% of the sessions, the inter-observer reliability for the total score on the *Hide the Pots* task was 97.96% (Kappa = 0.96).

**Descriptive statistics.** The mean scores and standard deviations for the *Hide the Pots* task at 18 months are reported in Table 1. Overall scores, errors, total time and MCDI scores were included in the analysis.

Differences between monolinguals and bilinguals were examined using an independent samples t-test. In contrast to our hypothesis, monolinguals performed significantly better than bilinguals on the task overall, where scores ranged from zero to three, \( t(59) = 2.74, p = .01 \). There were no differences between monolinguals and bilinguals in the number of perseverations,
In terms of other factors that were assessed, there were no differences between monolinguals and bilinguals in the length of time it took to complete the task, \( t(59) = -1.148, p = .88 \). There were also no differences in the combined language MCDI scores (which included the total number of words known across all languages) between the two groups, \( t(59) = .418, p = .68 \). There were also no differences between monolinguals and bilinguals on the English-only MCDI, \( t(59) = 1.62, p = .11 \).

First order correlations were run between demographic and outcome variables. Not surprisingly maternal education and SES were highly correlated, \( r(58) = .73, p < .01 \). Hide the Pots scores negatively correlated with the number of perseverations made, \( r(61) = -.42, p < .01 \), an unsurprising finding as well, given that perseverations are coded as an error. There was also a moderate negative correlation between the amount of second language exposure and Hide the Pots scores, \( r(61) = -.28, p = .03 \). There were moderate positive correlations between education and English-only MCDI scores, \( r(57) = .28, p = .03 \) and between education and the amount of second language exposure, \( r(64) = .28, p = .02 \).

A linear regression model was used to predict Hide the Pot outcomes. Second language exposure and socioeconomic status (SES) were the two variables selected for this model based on the first order correlations with the HTP scores. The model was as follows:

\[
(HTP)^\hat{} = 1.66 - .02 \text{(second language exposure)} + .01 \text{(SES)}
\]

Second language exposure was statistically significant in predicting Hide the Pots outcome \( (p = .01) \) and SES was marginally significant in predicting Hide the Pots outcome \( (p = .07) \). This linear regression demonstrated a positive association between Hide the Pots scores
and higher SES. In addition, there was a trend toward a negative association between *Hide the Pots* scores and greater exposure to a second language. The amount of second language exposure and SES explained 16% of the variance in *Hide the Pots* scores, $R^2 = .16, F(2, 53) = 4.74, p = .01$.

**Experiment 1b**

Experiment 1b included all 24-month participants who completed the *Spin the Pots* task described above.

**Methods**

**Participants** - The sample for the 24-month *Spin the Pots* task consisted of 34 monolingual ($n = 24$) and bilingual ($n = 10$) participants (18 males and 16 females). The mean rank of socioeconomic index (SEI, Nakao & Treas, 1992) for the families who completed this task was 74.65 ($SD = 15.51$). Education was ranked on a 6-point scale as described previously. The mean amount of education completed by parents was through college level ($M = 5.80, SD = .47$). Participants were Asian ($n = 1$), Caucasian ($n = 21$), Latino ($n = 2$), and of mixed ethnic origin ($n = 11$). One participant was dropped due to equipment failure.

**Apparatus** – For the 24-month olds, the *Spin the Pots* paradigm (Hughes & Ensor, 2005) was used to measure the working memory component of executive functioning. For this task, eight distinctly decorated cups were hidden in a circle on a lazy Susan apparatus with a cover. Six stickers were hidden under six different cups. Productive vocabulary was assessed via the MCDI. See Figures 1 and 2 in Tables and Figures for images of the apparatus.

**Procedure** – Twenty-four month olds completed the *Spin the Pots* paradigm. The infants placed attractive stickers in six of the eight pots, leaving two of the pots empty. After hiding the stickers, the child was reminded which two pots did not contain stickers, and then were
instructed to find the stickers. Before each trial, the child was told that he or she had one turn to pick one sticker. Then, the tray holding the pots was covered with a lid and spun around. The lid was spun 180 degrees on each trial in order to counterbalance the position of the pots. After the experimenter removed the lid, the child was invited to choose a pot where a sticker could be found. Each time the child chose a pot, he or she was given appropriate feedback before the pots were covered and rotated again. If a correct sticker had been chosen, it was removed from the beneath the cup and given to the parent or child. The child had up to sixteen trials to find all six stickers. This task required children to hold the location of the cups that did not have stickers in mind and to update this memory after each trial. Trials ended either when the child found all six stickers or became too bored or distracted to continue with the task. Parents completed a general questionnaire and the MCDI (described above).

Results

Coding and Reliability – Performance on the Spin the Pots task was first assessed by counting how many of six stickers (range = 0-6) were chosen after completing all trials. A span score, the number of correct responses before choosing an incorrect response (Ross et al., 2007) was also recorded. In addition, the trial number when three cups in a row were incorrectly chosen, the end trial number, and the elapsed time after five, ten and sixteen trials was recorded. The position of the cup (top or bottom, relative to the child’s positioning) and the color of the cup chosen in each trial were also recorded. Because two of the eight cups never contained stickers, the number of times a child chose these empty cups was also recorded as a specific type of error. Performance on this task was also coded for perseveration, correcting a response and choosing two cups at once. A total error score was also calculated to include all trials when a child chose the incorrect cup.
A number of different overall scores were calculated in order to gain a comprehensive understanding of the child’s performance on the task. The “original task score”, as calculated by Hughes & Ensor (2005), was calculated as sixteen minus the number of errors made. An additional score, the “trial score”, was calculated by subtracting the number of errors from the number of trials completed. Finally, another total score, the “all trials score”, was calculated in order to better compare children who had not completed all 16 trials with those who had. This score was adjusted so that all uncompleted trials (through 16) were counted as errors, except for children who had found all 6 stickers. For the children who found all of the stickers, their total score remained the same as the “original task score”, 16 minus the number of errors actually made (not accounting for all 16 trials). For children who did not find all 5 stickers, their error score included trials that had not been completed. This error score was then subtracted from 16. For example, if a child found 4 stickers and became too distracted to complete the task after 10 trials, their error score would be 16 trials – 4 correct. Their total score would be 16 minus this new error score (12), resulting in a score of 4. To explain in another way, children who found 6 stickers had more variance in their scores based on the number of errors they made before finding all 6 stickers. For all other children, their total score equated the total number of stickers they found. For this calculation, a higher number indicated a higher score. See Appendix A for the entire Spin the Pots coding protocol.

Based on 18% of the sessions, the inter-observer reliability for the Spin the Pots task was 97.6% (Kappa = .94).

Descriptive Statistics: The mean scores and standard deviations for the Spin the Pots task at 24 months are reported in Table 2. Overall scores, errors, total time and MCDI scores were included in this analysis. Differences between monolinguals and bilinguals were examined using
an independent samples t-test. Contrary to our expectations, there were no differences between monolinguals and bilinguals in overall performance on the task, based on the “all trials” score, \( t(32) = 1.32, p = .20 \). Scores on this measure ranged from 2 (lowest) to 15 (highest). No differences emerged between monolinguals and bilinguals on task performance according to any of the Spin the Pots composite scores. There were also no differences between monolinguals and bilinguals in the number of perseverations, \( t(32) = -0.98, p = .35 \), the number of corrections, \( t(32) = -0.84, p = .41 \), nor in the number of times two cups were chosen simultaneously, \( t(32) = .04, p = .97 \). The number of times the empty cups were chosen also did not differ for monolinguals and bilinguals, \( t(32) = -0.34, p = .74 \).

In terms of other factors that were assessed, there were no differences between monolinguals and bilinguals in the length of time it took to complete the task, the combined MCDI scores (including the total number of words known across all languages), the English-only MCDI scores, the proportion of total stickers found (relative to total number of trials), nor in the total number of errors made (relative to total number of trials), as shown in Table 2.

Table 2 includes the means for all three total score calculations for Spin the Pots. These scores were all computed differently in order to look at the data from a number of perspectives. First order correlations were run between total score measures, demonstrating correlations between all three types of calculations. The only other significant correlation was between education and SES, unsurprisingly, \( r(32) = .43, p = .01 \).

**Experiment 1c**

Experiment 1c included the 18- and 24-month old longitudinal participants who completed both the Hide the Pots task at 18-months and the Spin the Pots task at 24-months.
This experiment aimed to investigate specific changes in working memory task performance at 18- and 24-months.

**Methods**

**Participants** – Participants for experiment 1c were a subset of the participants included in experiments 1a and 1b. Twenty-two infants, 12 monolingual and 10 bilingual, were included in this sample (13 males and 9 females). The mean rank of socioeconomic index (SEI, Nakao & Treas, 1992) for the families who completed this task was 73.94 ($SD = 13.72$). Education was ranked on a 6-point scale as described previously. The mean amount of education completed by parents was through college level ($M = 5.86, SD = .35$). Participants were Caucasian ($n = 12$), Latino ($n = 1$) and of mixed ethnic origin ($n = 9$). Five participants were dropped either due to equipment failure ($n = 4$) or to sibling interference ($n = 1$).

**Coding and Reliability** – Coding and reliability for both *Hide the Pots* and *Spin the Pots* was the same as described above in experiment 1a and 1b, respectively.

**Results**

In order to assess changes in performance across longitudinal participants, a number of first order correlations were run between language and outcome variables, shown in Table 3. *Hide the Pots* performance at 18-months was not correlated with any measure of *Spin the Pots* at 24-months, including the “all trials score” that accounted for all 16 trials. While performance on *Hide the Pots* and on *Spin the Pots* were not correlated, other factors were closely correlated with *Spin the Pots* performance. Most notably, both the 18-month combined language MCDI, $r(20) = .74$, $p < .01$ and the 18-month English-only MCDI, $r(20) = .74$, $p < .01$, were positively correlated with the “original score” (16-errors) for *Spin the Pots* at 24 months. Scores on the 24-month English-only MCDI were also correlated with *Spin the Pots* “original score” performance, $r(17)$
= .53, \( p = .03 \). However, the 24-month combined language MCDI was not correlated with any measure of *Spin the Pots* performance. Another *Spin the Pots* outcome, the proportion of correct cups chosen to the total number of trials completed, was correlated with both the 18-month combined language MCDI, \( r(21) = -0.46, p = .03 \) and the 18-month English-only MCDI, \( r(20) = -0.47, p = .03 \).

While *Spin the Pots* performance was closely correlated with MCDI scores, language ability was not correlated with *Hide the Pots* scores. Neither the 18 month English-only MCDI, \( r(17) = -0.24, p = .92 \), nor the combined language MCDI, \( r(17) = -0.98, p = .71 \) at 18-months were correlated to *Hide the Pots* scores. In terms of language scores overall, English-only MCDI scores at 18 and 24 months were correlated with each other, \( r(17) = .65, p < .01 \), though combined language MCDI scores at 18 and 24 months were *not* correlated, \( r(21) = .02, p = .94 \).

All other correlations for the longitudinal participants are shown in Table 3.

**Discussion**

Though previous research has demonstrated cognitive advantages for bilingual children (e.g. Bialystok & Shapero, 2005), little research has been conducted to assess earlier development of these advantages. Poulin-Dubois and colleagues (2010) completed the only study to date investigating bilingual EF advantages in children as young as two-year olds. However, after using a battery of EF tests, this study only demonstrated significant bilingual advantages on the Stroop task, a conflict and inhibition task. Given the mixed evidence regarding the effect of bilingualism on WM (see Bialystok et al., 2009), the current study attempted to assess the role of WM at an even earlier age than previously studied. Furthermore, this study provided a more in-depth understanding of the differences in the WM component of EF in 18- and 24-month old infants.
Experiment 1a – Contrary to our hypothesis, monolinguals outperformed bilinguals on the 18-months *Hide the Pots* (HTP) task. Only the overall task score was significantly different in monolinguals and bilinguals. However, bilinguals had a higher mean of perseverations than monolinguals. More importantly, bilinguals made more errors than monolinguals, even though these errors were not confined to a particular error type. The differences in overall performance and in number of errors may serve as an indication that monolinguals and bilinguals are solving the task differently. This is possibly due to the transition at this age from comprehension to production of language. Based on Green’s (1998) model, the shift to language production would require the development of inhibitory processes in bilinguals. If language production is a key phase in this shift, it is possible that WM and inhibition processes are just beginning to develop in bilinguals, meaning that an advantage would not yet have emerged at this point.

The regression model for the HTP task showed that higher SES had a positive effect on HTP scores, while greater second language exposure had a negative effect. The SES effect is unsurprisingly, particularly given recent research demonstrating the effects of SES on infant cognitive flexibility (Clearfield & Nima, 2012). However, the second language exposure effect is a more surprising finding. The more second language exposure a child has, the worse they perform on the HTP task. Though this is unexpected, it does correspond to the finding that bilinguals performed worse on the task. It might also be explained by Green’s model (1998), given that these infants are constantly being given (linguistic) input that requires inhibition and WM, but they have not yet mastered the implicit ability to switch between languages. Given that the factors used for the regression model predicted HTP scores in different directions, we were not able to run interactions on this model. However, the role of SES and second language
exposure in EF task performance play an important role in understanding the development of EF in infants.

**Experiment 1b** – While there were significant bilingual and monolingual differences on the HTP score at 18-months, there were no significant differences on the 24-month *Spin the Pots* (STP) task. Though the differences were insignificant, monolinguals did continue to outperform bilinguals on this task, but to a lesser extent than at 18-months. One critical component in analyzing performance on the *Spin the Pots* task was the way in which to calculate the overall task performance score. When Hughes and Ensor (2005) developed the task, they used it as part of a battery of tests with 24-36 month olds, not differentiating more narrowly across this age group. As a result, only one participant out of 140 failed to complete all 16 trials and was not included in the analysis. Similarly, Bernier and colleagues (2010) used this task on 26-month olds and all 80 participants were able to complete the 16 trials.

The older age groups assessed in previous analyses of this task provide several reasons for why STP performance may not have differed between monolinguals and bilinguals in the present study. First, these older participants were able to complete all 16 trials of the task, whereas the majority of our 24-month participants did not complete all 16 trials. Most 24-month olds were not able to attend to the task for the full trial length, largely frustration caused by their inability to find the stickers. For this reason, we had to develop different ways to assess overall task performance to account for the trials that were not completed. As described in more detail in the coding section above, we counted uncompleted trials (up to 16) as errors unless children selected all 6 stickers. Based on this error score, we were better able to the see variations in performance as if children had completed all 16 trials. However, there were still no significant differences in monolingual and bilingual performance using any of the different score calculations. Given that
the task was originally designed to include all 16 trials and that few participants in our 24-month sample found all 6 stickers, this indicates the possibility of a floor effect on the task at this age. If this were to be the case, it might explain the lack of significant differences between monolinguals and bilinguals on this task. Additional research needs to be completed to assess whether this task is an effective measure of WM at 24 months, particularly because our sample is limited in power by its small size ($n=34$).

If the STP task is accurately measuring EF in 24-month olds, the lack of significant differences between monolinguals and bilinguals may be explained in two additional ways. First, because WM has not been studied in monolinguals at bilinguals at such a young age, it is possible that WM differences have not yet emerged. If there is a only a gradual progression in the development of WM abilities (and advantages) in bilinguals, this explanation aligns with the findings at 18-months. It is possible that at 18-months, bilinguals are performing worse because they are beginning to process inhibition and selective attention differently as productive vocabulary emerges. By 24-months, there are no significant differences between monolinguals and bilinguals, though monolinguals are still performing slightly better than bilinguals. At this point, the increased practice that bilinguals have with language production in multiple languages may begin to foster WM and inhibitory control abilities, leading to improvement on tasks such as STP. Future research with a larger sample size would benefit from gaining an even more detailed analysis of error patterns made by bilinguals at 24-months, including perseverations. If bilinguals make errors indicative of inhibitory processing, even if this processing is faulty, it would provide evidence for a gradual emergence of cognitive advantages over monolinguals at this age.
Alternately, it is possible that WM differences in monolinguals and bilinguals do not exist at all. While Poulin-Dubois and colleagues (2010) found bilingual advantages at this age, the only significant advantages found were in a different type of task, the Stroop Task. The Stroop task is primarily an inhibition task and not a WM task. It is possible that bilinguals have advantages in some areas of EF and not in others. This possibility is further reinforced by the fact that previous studies have shown mixed evidence of WM advantages in bilinguals, as discussed by Bialystok and colleagues (2009). This review article (Bialystok et al., 2009) cited contradictory findings from previous studies of bilingual performance on different types of WM task.

**Experiment 1c:** First order correlations between language and EF task outcome yielded some significant results, though STP and HTP performance was not correlated. Language performance on the 18-month MCDI positively correlated with better STP scores 6 months later. The fact that language at 18-months may predict STP performance at 24-months provides further evidence for the possibility that Green’s model (1998) does indeed come into play as productive language skills are emerging. Bilingual children who have more productive vocabulary at 18-months gain substantial practice with inhibition and WM by 24-months, at which point they have improved performance on a novel WM task. The theoretical implications of this explanation merit further research on the role of language as a mediating component for EF development in monolinguals and bilinguals.

Though significant language correlations were evident among the longitudinal participants, there were no correlations between HTP and STP. Bernier and colleagues (2010) found positive correlations between HTP at 18-months and the EF dimension (Conflict) that included STP at 26-months. Three possible explanations might explain the lack of a significant correlation between HTP and STP.
First, as was discussed in relation to experiment 1b, there is a possible floor effect for the STP task as a measure of WM for 24-month-olds. If this were the case, STP outcomes would not be a meaningful measure of WM at 24-months. If STP is not successfully measuring WM at this age, it is unsurprising that it would not correlate with the 18-month WM measure (HTP).

Another possibility is that our small sample size ($n = 22$) resulted in insufficient statistical power, which would explain the insignificant correlation between HTP and STP. Bernier and colleagues (2010) had a much larger sample ($n = 80$) and thus did not have concerns relating to power when correlating their sample.

A third potential reason for the lack of a significant correlation between STP and HTP is that both tasks might be measuring a slightly different component of WM. This possibility is further substantiated by the fact that Bernier and colleagues (2010) grouped STP into a composite score when correlating task performance with HTP.

**General Discussion:** The results of this study showed no apparent bilingual advantages in WM at 18- and 24- months. The two WM tasks, HTP and STP were also not correlated. As outlined above, there are a number of possible theoretical explanations for these findings.

Though Bernier and colleagues (2010) and Hughes & Ensor (2005) developed the tasks used in the present study, the current study examined performance on these tasks in more detail. As a result, a number of theoretical differences could be identified. Because STP scores had to be calculated differently for the younger participants in the current study, minute discrepancies in task performance were identified. Both HTP and STP were examined for error patterns, specifically perseverations. Though STP may perhaps be too difficult a task for 24-month olds, HTP and STP performance also may not have correlated because the tasks are measuring different components of WM.
The differences between these two tasks can closely be related to the stationary and scrambled versions of the Multiple Boxes Task (Petrides, 1988). The stationary variation of the Multiple Boxes task allows participants to use a simple spatial search strategy or to use WM to solve the task. Because spatial cues are available, individuals with frontal lobe damage can complete this task. Similarly, the cups on the HTP task do not change spatial locations. Therefore, participants can use both WM and spatial cues in order to solve the task. On the other hand, the scrambled version of the Multiple Boxes task removes spatial cues so that only WM can be used to complete the task. This is similar to STP in that the cups in STP change location, eliminating spatial cues. Individuals with frontal lobe damage cannot complete such tasks. Despite these subtle variations, both STP and HTP involve components of WM. However, they can be differentiated by the availability of spatial cues. Another relevant component that distinguishes these tasks from one another is the self-ordered nature of STP, as opposed to the non-self-ordered nature of HTP, discussed in more detail in the introduction outlining the history of EF task development. Both of these tasks require WM, though one definitely requires frontal lobe involvement and the other may not.

In converting STP to a simpler task for younger children, Bernier and colleagues (2010) inadvertently shifted some of the mechanisms needed to solve the task. As noted in the comparison between the stationary and scrambled version of the Multiple Box Task, HTP and STP may measure different components of WM and may require different areas of the brain. Though STP is primarily a WM task, when correlated with HTP, Bernier and colleagues (2010) grouped STP with a set of tasks that strictly measured inhibition, including the Stroop task. HTP may measure components of inhibition in addition to WM. Because STP was part of a composite score primarily including inhibition tasks, this could have been the reason for the
correlation between these tasks in Bernier and colleagues’ study (2010). This method of comparing the tasks would have exaggerated their direct relevance, explaining the lack of a correlation in the present study. Further research is warranted in order to better identify the underlying components of these tasks and to ultimately determine whether they are related.

In summary, while both STP and HTP involve WM, they may involve a different component of WM. Bernier’s correlation of HTP with the composite score including STP does not necessarily provide evidence that the two tasks are related. However, these differences will be easier to distinguish with additional longitudinal data that allows for better insight into EF development.

This study adds to the current literature of the role of WM and bilingualism. Given the mixed evidence in previous studies of whether bilinguals even have an advantage in bilingualism (see Bialystok et al, 2009), this study demonstrates additional ways in which WM tasks need to be refined before conclusions can be made at any age. It is even more difficult to determine WM mechanisms in young children, given that it is unknown whether such processes are even differentiated at this age, or whether they are the same as in adults. Tasks such as HTP and STP assume that infants’ EF processes operate analogously to those of adults, when this may not be the case at all. Furthermore, even if different components of EF can be separated out in infants, there is still substantial uncertainty regarding the best way to measure EF and whether performance on these tasks will relate to later EF abilities. Investigating bilingual infants in comparison to monolingual infants is one possible way to discern early differences in EF. Within the context of this study, the HTP task demonstrated differences in inhibition between bilinguals and monolinguals at 18-months, a finding that warrants further investigation into the development of inhibition specifically. Given that Poulin-Dubois’ (2010) finding showed
bilingual advantages in inhibition at 24-months, this is perhaps an area of EF development that varies according to mediating factors such as bilingualism. On the other hand, tasks that solely rely on WM have more ambiguous results, both in this study and in previous studies of older and younger adult bilinguals (Bialystok et al., 2009). The ambiguous and contradictory findings related to WM and bilingualism could be an issue of the WM tasks currently being used, both for infants and for adults. Alternatively, it could be that WM on its own does not truly vary between monolinguals and bilinguals.

The findings in this study have prompted numerous questions for additional research, many of which will hopefully be addressed when 30-month longitudinal data is collected for the participants in this study. Data collection is ongoing in our lab, both in EF tasks and on other potential mediating components of EF, such as parent-child interactions. Given previous research on the role of children’s social environment on EF development (e.g. Carlson, 2009), there is reason to explore further factors that may influence EF. Finally, this ongoing data collection will include imitation tasks as a representational task measuring long-term memory as opposed to WM. This data will provide insight into the integration of memory systems, which should also help to explain the early development of EF. This set of ongoing data collection, in addition to the proposed areas for future research, provide a myriad of methods by which to gain additional insight into EF development in children, particularly in bilingual children.
References


ambiguous figures. *Developmental Science, 8*(6), 595-604.


Feng, X., Bialystok, E., & Diamond, A. (*submitted* 2009). Do bilingual children show an advantage in working memory?


## Tables and Figures

Table 1.

Scores on *Hide the Pots* task (± 1 SD) and MCDI (± 1 SD) measures for 18-month participants in Experiment 1a as a function of monolingual and bilingual language exposure

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bilingual</th>
<th>t-test (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HTP (18 mo.)</strong></td>
<td>N= 37</td>
<td>N=24</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2.27 (.80)</td>
<td>1.71 (.75)</td>
<td>.01*</td>
</tr>
<tr>
<td>Perseveration</td>
<td>.24 (.50)</td>
<td>.50 (.66)</td>
<td>.11†</td>
</tr>
<tr>
<td>Corrections</td>
<td>.03 (.16)</td>
<td>.13 (.34)</td>
<td>.20</td>
</tr>
<tr>
<td>Two Cups</td>
<td>.14 (.42)</td>
<td>.08 (.41)</td>
<td>.64</td>
</tr>
<tr>
<td>MCDI Combined</td>
<td>38.76 (29.96)</td>
<td>35.38 (28.64)</td>
<td>.68</td>
</tr>
<tr>
<td>MCDI English-only</td>
<td>36.19 (31.30)</td>
<td>23.17 (25.20)</td>
<td>.11†</td>
</tr>
</tbody>
</table>

*P < .05, † p < .10
Table 2.

Scores on Spin the Pots task (± 1 SD) and MCDI (± 1 SD) measures for 24-month participants in Experiment 1a as a function of monolingual and bilingual language exposure

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bilingual</th>
<th>t-test (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STP (24 mo.)</td>
<td>N= 24</td>
<td>N=10</td>
<td></td>
</tr>
<tr>
<td>All Trials Score</td>
<td>6.50 (3.41)</td>
<td>5.10 (2.51)</td>
<td>.20</td>
</tr>
<tr>
<td>Original Task Score</td>
<td>9.75 (2.63)</td>
<td>9.10 (3.11)</td>
<td>.54</td>
</tr>
<tr>
<td>Trial Score</td>
<td>4.92 (1.10)</td>
<td>4.50 (.85)</td>
<td>.29</td>
</tr>
<tr>
<td>Perseveration</td>
<td>.25 (.53)</td>
<td>.60 (1.08)</td>
<td>.35</td>
</tr>
<tr>
<td>Corrections</td>
<td>.54 (.78)</td>
<td>.80 (.92)</td>
<td>.41</td>
</tr>
<tr>
<td>Two Cups</td>
<td>.21 (.51)</td>
<td>.20 (.63)</td>
<td>.97</td>
</tr>
<tr>
<td>MCDI Combined</td>
<td>53.26 (22.58)</td>
<td>39.72 (27.76)</td>
<td>.17</td>
</tr>
<tr>
<td>MCDI English-only</td>
<td>53.26 (22.58)</td>
<td>37.17 (29.42)</td>
<td>.11†</td>
</tr>
</tbody>
</table>

*P < .05, † p < .10
Table 3.
First-order correlations between *Spin the Pots* task, *Hide the Pots* task and MCDI measures for longitudinal participants in Experiment 1c

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HTP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STP: All Trials</td>
<td>.09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STP: Original</td>
<td>.01</td>
<td>.57*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STP: Proportion Correct</td>
<td>.13</td>
<td>-.87**</td>
<td>-.86**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>STP: Proportion Errors</td>
<td>.24</td>
<td>.87**</td>
<td>.86**</td>
<td>-.25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MCDI 18 English</td>
<td>-0.24</td>
<td>.34</td>
<td>.74*</td>
<td>-0.47*</td>
<td>.36</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MCDI 18 Combined</td>
<td>-0.98</td>
<td>.25</td>
<td>.74*</td>
<td>-0.46*</td>
<td>.30</td>
<td>.91**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MCDI 24 English</td>
<td>.34</td>
<td>.24</td>
<td>.53*</td>
<td>-0.30</td>
<td>.23</td>
<td>.65*</td>
<td>.38</td>
<td>-</td>
</tr>
<tr>
<td>MCDI 24 Combined</td>
<td>.02</td>
<td>.10</td>
<td>.02</td>
<td>-.15</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
<td>.99**</td>
</tr>
</tbody>
</table>

**p < .01  *p < .05, † p < .10**
Figure 1: Hide the Pots task
Figure 2: Spin the Pots apparatus
Figure 3: Spin the Pots task
Appendix A

Protocols

_hide the Pots_ Protocol

1. Print coding sheet.
2. Enter participant’s code number, age, gender and the study participated in in the corresponding columns.
3. Code a first pass of the video, beginning when the lid is first lifted (skip practice trials):
   a. Use a timer to record latency for each trial. Begin timing when the lid is raised and end when the child picks the first cup (whether right or wrong) for each trial. _Only_ code the first cup that is lifted by the child, regardless of whether the second cup is chosen immediately after (_unless_ two cups are chosen simultaneously). Record the time in seconds in the appropriate columns (“Latency 1” for trial 1, “Latency 2” for trial 2 & “Latency 3” for trial 3).
   b. Note for each trial whether the child picks the right or wrong cup. Enter this data in the column labeled “T1” for trial 1, “T2” for trial 2, and “T3” for trial 3.
      i. If the child is correct, write “c.”
      ii. If the child is incorrect write “x”
      iii. If the child picks two cups, one of which is right, write “2cr”
      iv. If the child picks two cups, both of which are wrong, write “2cw”
      v. If the child perseverates (picks the same cup that was chosen on the previous trial, whether the cup was right or wrong on the previous trial, max = 2), write “p”
         1. Note: if the child picks up the same cup on multiple trials but it is correct the second or third time, code this as correct, not as perseveration
         2. Note: If a child picks 2 cups and one (or both) had been chosen previously, this should be coded as _both_ a perseveration and 2 cups. If the code is “2cw” and “p,” this should be counted as incorrect. If the code is “2cr” and “p,” this should be counted as correct.
      vi. If the child touches one cup and then picks up a different cup that is incorrect, write “cw”
      vii. If the child touches one cup and then picks up a different cup that is correct, write “cr”
   c. Tally EF score (range 0-3) and enter it in the “Total EF Score” column.
      i. “2cr,” “c,” “2cr, p” and “cr” indicate a correct answer
      ii. “x,” “cw,” “2cw,” “p” and “2cw, p” indicate incorrect answers
   d. Note how many times the child perseverated (“p”) in the column labeled “#p,” Also, note how many times the child corrected his/her response _only if it was changed to a correct answer_ (“cr”) in the column labeled “#cr.” Finally, record how many times the child picked up 2 cups (“2cr” or “2cw”) in the column labeled “#2c.”
   e. Note the order in which the balls were hidden (eg. 3,1,2) in the column labeled “Combination”
i. Cup #1: cup closest to box hinge
   ii. Cup #2: middle cup
   iii. Cup #3: cup furthest from box hinge

4. Code a second pass of the video:
   a. Time entire length of trial (in seconds) from when the box is lifted for the first trial to when the first cup is selected by the child in the third trial
   b. Confirm that coding was accurate on the first pass
   c. Record any language(s) heard during the demonstration and trials in the column labeled “language(s)”

5. Enter data into electronic spreadsheet.
Appendix B

*Spin the Pots* Protocol

1. Print coding sheet.
2. Use the coding sheet for individual trials (coding sheet #1) while watching the video the first time through. Enter participant’s code number. Before the child begins selecting cups, note which two cups *never* had a sticker on the summary sheet. For each trial, note whether the child picks a cup with or without a sticker in the column corresponding to the trial number. *Only* code the first cup that is lifted by the child, regardless of whether the second cup is chosen immediately after (unless two cups are chosen simultaneously). Enter information until the child picks all 6 stickers or stops participating in trials. Stop coding at 16 trials even if additional trials are completed.
   a. If the child picks a cup with a sticker underneath, write “c.”
   b. If the child picks a cup without a sticker underneath, write “x.”
   c. If the child picks two cups, one of which has a sticker, write “2cr.”
   d. If the child picks two cups, neither of which has a sticker, write “2cw.”
   e. If the child picks two cups, both of which have a sticker, write “2c2r,” but only count this as 1 correct (i.e., max score is now 5, unless experimenter only allows the child to pick one sticker).
      i. Note: If the 2nd cup picked up after the child looks under the first cup (i.e., not simultaneously) and the 2nd cup is either not examined by the child or is blocked by experimenter, do not code this as 2 cups.
   f. If the child perseverates (picks the same cup as chosen on the previous trial, whether this cup was right or wrong on the previous trial, max = 15), write “p.”
      i. Note: if the child picks up the same cup on multiple trials but it is correct the second or third time, code this as correct, not as a perseveration.
      ii. Note: If a child picks 2 cups and one (or both) had been chosen previously, this should be coded as both a perseveration and 2 cups. If the code is “2cw” and “p,” this should be counted as incorrect. If the code is “2cr” and “p,” this should be counted as correct.
   g. If the child touches one cup and then picks up a different cup that does not contain a sticker, write “cw”
   h. If the child touches one cup and then picks up a different cup that contains a sticker, write “cr.”
3. On the same coding sheet (coding sheet #1), note the total time elapsed (since the lid was lifted for the first trial) after 5 trials, after 10 trials and after the last trial. The last trial is when the child finds the 6th sticker, when the experimenter ends the trials, or when the child has completed 16 trials. List all time in seconds.
4. After completing coding sheet #1, print the coding sheet for recording color and position of the cup chosen (coding sheet #2). Enter the participant’s code number.
   a. Watch the video a second time, recording the color of the cup chosen (according to the key at the bottom of the coding sheet) and the position of the cup (on the top or bottom half of the apparatus, relative to the child’s positioning). The color and the position of the cup should be recorded in the column corresponding to the trial number.
5. After completing coding sheet #2, print the summary coding sheet (coding sheet #3). Enter participant’s code number, age, gender, tape number, study name, monolingual or bilingual and any languages heard during the video.

6. Use the data collected on coding sheets #1 and #2 to enter data in coding sheet #3.
   a. Record the total number of cups correctly chosen (i.e. number of stickers found) in the first 5 trials. Enter this information in the column labeled “stickers 5 trials” on coding sheet #3. Repeat this step for the first 10 trials, if applicable, and for the total number of trials completed.
      i. “c,” “cr,” “2cr,” “2cr, p,” “2c2r, p” and “2c2r” indicate a correct answer
      ii. “x,” “cw,” “2cw,” “2cw, p” and “p” indicate incorrect answers
   b. In the column labeled “STP Total,” enter the number resulting from 16 minus the number of errors (errors include “x,” “cw,” “2cw,” and “p”). In the column labeled “STP Trial Score,” enter the number resulting from the # of trials minus the number of errors (again, errors include “x,” “cw,” “2cw,” and “p”).
   c. Determine the trial number when the child consecutively chose 3 cups without stickers underneath them. Any of the following responses would indicate incorrect responses: “x,” “cw,” “2cw,” “2cw, p” or “p.” Record the relevant trial number in the column labeled “trial # when 3 wrong in a row” on coding sheet #3.
   d. Determine the number of consecutive correct responses before the child chose any incorrect response. Enter this number into the column titled “Span Score” on coding sheet #3. If the first cup the child chose had no sticker, write “0” in this column. If the child found all 6 stickers without selecting a cup without a sticker, write “6” in this column.
   e. In the column labeled “End Trial #” on coding sheet #3, enter the last trial number in which the child participated.
   f. Record how many times the child perseverated (“p”) in the column labeled “#p.” Also, note how many times the child corrected his/her response only if it was changed to the correct answer (“cr”) in the column labeled “#cr.” Finally, enter how may times the child picked up 2 cups (“2cr,” “2cw,” or “2c2r”) in the column labeled “#2c” based on the information in coding sheet #1.

   g. Record how many cups the child chose from the top (further away) half of the apparatus in the column labeled “total top” and the total number of cups chosen from the bottom half in the column labeled “total bottom” based on the information recorded on coding sheet #2.
   h. In the columns labeled “Time 5 trials,” “Time 10 trials” and “Total Time,” on coding sheet #3 enter the corresponding data from coding sheet #1. “Total time” is either the time after 16 trials, or after the last trial is complete.
   i. In the column on coding sheet #3 labeled “cup 1 trial #” write the trial number during which the child first selected the first cup that never had a sticker. In the column labeled “cup 2 trial #” write the trial number during which the child first selected the first cup that never had a sticker. Write the number of times each of the cups that never had stickers was chosen in the corresponding columns, either “# trials cup 1,” or “# trials cup 2.”
   j. In the column on coding sheet #3 labeled “strategy,” note whether or not the child seemed to be using any type of strategy during the trials. If the child did seem to use a strategy, write a brief description of this strategy. For example, the child
may have always chosen the same cup or always chosen the cup directly in front of him/her.

k. In the column on coding sheet #3 labeled “errors” enter how many errors the child made (“cw,” “2cw,” “2cw, p,” “x” and “p”)

l. In the final column on coding sheet #3, write any additional comments.

7. Enter data from both coding sheets into electronic spreadsheets.
## Appendix C

**PKU Tasks (Diamond, 1990)**

<table>
<thead>
<tr>
<th>Task</th>
<th>Change with PKU <em>(High Phe: 6-10 mg/dl vs. Low Phe (6 mg/dl))</em></th>
<th>Brain Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>Inversely related to Phe levels</td>
<td></td>
</tr>
<tr>
<td>Infants: A-not-B Task</td>
<td>High Phe (6-10 mg / dl): shorter delays, worse</td>
<td>Dorsolateral PFC</td>
</tr>
<tr>
<td></td>
<td>performance than normative group, but not significantly</td>
<td></td>
</tr>
<tr>
<td>Infants: Object Retrieval (Large &amp; Small Box)</td>
<td>High Phe: significantly worse performance on front-open and</td>
<td>Dorsolateral PFC</td>
</tr>
<tr>
<td></td>
<td>side-open trials vs. Low Phe and other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparison groups</td>
<td></td>
</tr>
<tr>
<td>Infants: Spatial Discrimination</td>
<td>Groups did not differ significantly in performance</td>
<td>Inferotemporal Cortex</td>
</tr>
<tr>
<td></td>
<td>(High Phe, Low Phe, Normative)</td>
<td>(NOT PFC)</td>
</tr>
<tr>
<td>Infants: Visual Paired Comparison</td>
<td>No significant group differences With 10 minute delay: High</td>
<td>Medial Temporal Lobe</td>
</tr>
<tr>
<td></td>
<td>Phe – compared to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>normative, less likely to remember sample after delay</td>
<td></td>
</tr>
<tr>
<td>Toddlers: A-not-B Invisible Displacement</td>
<td>High-Phe: comparable to other groups (low Phe, normative) until</td>
<td>PFC</td>
</tr>
<tr>
<td></td>
<td>21 months at which point performance leveled off or declined</td>
<td></td>
</tr>
<tr>
<td>Toddlers: Three Boxes (scrambled)</td>
<td>No significant group differences, except higher</td>
<td>Dorsolateral PFC</td>
</tr>
<tr>
<td></td>
<td>perseveration with High Phe levels</td>
<td></td>
</tr>
<tr>
<td>Toddlers: Three Boxes (stationary)</td>
<td>No significant group differences (low Phe, high Phe, normative)</td>
<td>Unknown (NOT PFC)</td>
</tr>
<tr>
<td>Toddlers: Delayed non-matching to sample</td>
<td><em>30 Second Delay</em></td>
<td>Medial Temporal Lobe</td>
</tr>
<tr>
<td></td>
<td>High Phe: performed significantly worse than general population</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Phe: performed similarly to high Phe (?)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>5 Second Delay</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No significant differences between groups</td>
<td></td>
</tr>
<tr>
<td>Toddlers: Global-local (preferential</td>
<td>No significant differences between groups, except mostly</td>
<td>Parietal Cortex</td>
</tr>
<tr>
<td>looking procedure)</td>
<td>unusable sessions between 15-18 months for</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High Phe toddlers because they were too fidgety and distracted</td>
<td></td>
</tr>
<tr>
<td>Children: Day-night Stroop-like test</td>
<td>High Phe: correct on significantly fewer trials than normative</td>
<td>Dorsolateral PFC</td>
</tr>
<tr>
<td></td>
<td>sample and on fewer trials than low Phe (but not significantly)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>due to difficulty keeping 2 rules in mind (not attention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>difficulties); performance significantly related to IQ</td>
<td></td>
</tr>
<tr>
<td>Children: Tapping</td>
<td>High Phe: significantly more likely to fail tapping pretest (but</td>
<td>PFC</td>
</tr>
<tr>
<td></td>
<td>pretesting didn’t allow enough opportunity to demonstrate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>understanding of rules for tapping), including all who failed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pretests high Phe children performed significantly worse than</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other groups</td>
<td></td>
</tr>
</tbody>
</table>
Infants: 6-12 months, Toddlers: 15-30 months, Children: 3.5-7 years old

<table>
<thead>
<tr>
<th>Test Description</th>
<th>Comparison</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children: Three Pegs</td>
<td>High Phe: performed significantly worse than all comparison groups</td>
<td>Not yet studied in relation to brain function</td>
</tr>
<tr>
<td>Children: Corsi-Milner Test of Temporal Order Memory</td>
<td>No significant group differences</td>
<td>PFC</td>
</tr>
<tr>
<td>Children: Six Boxes (boxes scrambled after each reach)</td>
<td>Ages 3.5-4 years, <em>Low Phe</em> children required <em>more</em> reaches than general population and showed most improvement with increasing age, below age 5 performed <em>worse</em> than High Phe children → Inverse relation between concurrent level of Phe and number of reaches to open all boxes.</td>
<td>PFC</td>
</tr>
<tr>
<td><em>High Phe: made first error significantly earlier in trial than children in general population</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children: Stroop Control Condition</td>
<td>No significant group differences</td>
<td>Dorsolateral PFC, anterior cingulated cortex</td>
</tr>
<tr>
<td>Children: Corsi-Milner Test of Recognition Memory</td>
<td>No significant group differences</td>
<td>Medial Temporal Lobe</td>
</tr>
<tr>
<td>Children: Six-Boxes (stationary)</td>
<td>No significant group differences, overall performance significantly better in stationary-box condition than in stationary boxes condition</td>
<td>Unknown (NOT PFC)</td>
</tr>
<tr>
<td>Children: Global-Local (forced-choice)</td>
<td><em>Low Phe: longer</em> to respond than general population, <em>worse</em> at identifying local features small test stimuli and opposite size stimuli. High Phe: didn’t differ from other groups, except little improvement over age</td>
<td>Parietal Cortex</td>
</tr>
<tr>
<td>Children: Line Bisection</td>
<td>No significant group differences</td>
<td>Parietal Cortex</td>
</tr>
</tbody>
</table>