ELSEVIER

Early Childhood Research Quarterly 23 (2008) 141–158

Early Childhood Research Quarterly

Touch your toes! Developing a direct measure of behavioral regulation in early childhood

Claire E. Cameron Ponitz [a,*], Megan M. McClelland [b], Abigail M. Jewkes [c], Carol McDonald Connor [d], Carrie L. Farris [b], Frederick J. Morrison [a]

[a] *University of Michigan, Department of Psychology, United States*
[b] *Oregon State University, Department of Human Development and Family Sciences, United States*
[c] *Saginaw Valley State University, College of Education, United States*
[d] *Florida Center for Reading Research, Florida State University, United States*

Received 29 August 2006; received in revised form 23 January 2007; accepted 28 January 2007

Abstract

Behavioral aspects of self-regulation, including controlling and directing actions, paying attention, and remembering instructions, are critical for successful functioning in preschool and elementary school. In recent years, several direct assessments of these skills have appeared, but few studies provide complete psychometric data and many are not easy to administer. We developed a direct measure of children's behavioral regulation, the Head-to-Toes Task, and report performance of participants aged 36–78 months, including a group of Spanish-speaking children, from two different sites ($N = 353$; $N = 92$). We examined construct validity, examiner reliability, sources of variation, and associations between task scores and background characteristics. Results showed that the task was valid, reliable, and demonstrated variability in children's scores. A cross-classified hierarchical growth curve analysis indicated that girls, participants assessed in English, and higher-socioeconomic status (SES) children achieved slightly higher average scores than did boys, Spanish-speaking and lower-SES children, but effect sizes were small. Older participants achieved higher scores than did younger children, and there were no effects for site. Results suggest that the Head-to-Toes Task is an informative and easy-to-administer direct assessment of children's behavioral regulation. We discuss implications for its use in early childhood settings.

Keywords: Behavioral regulation; Cross-classified hierarchical model; Early childhood; Growth curves; Observational measure

Young children's academic and social success depends on a variety of skills, including emergent literacy and basic knowledge of math and vocabulary (Miller, Kelly, Zhou, & Campbell, 2005; Whitehurst & Lonigan, 2001). Beyond early academic competence, the ability to regulate behavior also signifies whether or not children will prosper in school. *Self-regulation* describes a broad construct representing the skills involved in controlling, directing, and planning emotions, cognitions, and behavior, and is important for functioning in varied contexts, including classrooms (Baumeister & Vohs, 2004; Shonkoff & Phillips, 2000).

In early educational settings, strong self-regulation has been linked with effective classroom behavior and high achievement, whereas poor self-regulation forecasts future problems in school (Blair, 2002; Bronson, 2000; Fabes,

* Corresponding author at: Center for Advanced Study of Teaching & Learning (CASTL), University of Virginia, 350 Old Ivy Way, Suite #300, Charlottesville, VA 22908, United States. Tel.: +1 734 647 9439; fax: +1 734 615 0573.
E-mail address: cameron.ce@gmail.com (C.E. Cameron Ponitz).

0885-2006/$ – see front matter
doi:10.1016/j.ecresq.2007.01.004

Martin, Hanish, Anders, & Madden-Derdich, 2003b; Howse, Calkins, Anastopoulos, Keane, & Shelton, 2003a; Shonkoff & Phillips, 2000). Furthermore, teachers report that many children enter formal schooling without adequate levels of these vital skills. Reports reveal substantial numbers of young students, ranging from 15 to 50% in two different studies, have trouble paying attention to and remembering instructions, converting automatic behaviors into controlled responses (such as raising their hand before participating and taking turns), completing tasks independently, and transitioning between tasks (McClelland, Morrison, & Holmes, 2000; Rimm-Kaufman, Pianta, & Cox, 2000). These basic competencies lay a foundation for individual learning and also contribute to overall classroom functioning, because one or two children with poor regulation can derail a teacher's plans for the entire class.

In recent years, a variety of conceptualizations of self-regulation have been established in a steadily emerging area of inquiry, including executive function, executive control, learning-related skills, emotion regulation, and behavioral regulation (Eisenberg & Spinrad, 2004; Howse et al., 2003a; McClelland & Morrison, 2003; Pintrich, 2000; Rueda et al., 2004). Research on self-regulation during early childhood shows that neurological, linguistic, and motor functions develop to enable independent and deliberate actions (Epsy, Kaufmann, Glisky, & McDiarmid, 2001; Shonkoff & Phillips, 2000). In this study, we focus on "cool" or cognitively mediated *behavioral regulation* skills, which are refined during early childhood and contrast with "hot" or affectively mediated regulation skills (i.e., delay of gratification, controlling aggression). Emotion regulation skills also play a role in children's early school success but are not presently highlighted (Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Mischel, Shoda, & Rodriguez, 1989; Zelazo, Müller, & Goswami, 2002).

Behavioral regulation includes inhibitory control, attention, and working memory. Together and separately, these three skills predict achievement and broader measures of functioning in early learning environments (McClelland, Cameron, Wanless, & Murray, 2007). They are also primarily implicated in executive function and relate to success in school settings (Blair, 2002; Bronson, 2000). *Attention* includes focusing, sustaining, and shifting attention (Barkley, 1994; Rueda, Posner, & Rothbart, 2005). In one study, children who could exercise sustained attention had higher academic achievement than their peers (Howse, Lange, Farran, & Boyles, 2003b). *Working memory*, or holding information in mind while processing new information, is also important for academic success (Adams, Bourke, & Willis, 1999). In addition, there is evidence that *inhibitory control*, or deliberately stopping an automatic response to exhibit another behavior, is the main regulatory deficit in Attention Deficit Hyperactivity Disorder (ADHD; Thorell & Wahlstedt, 2006). Lack of inhibition often impairs social and academic efficacy (Barkley, 2004; Blair, 2003).

Reliable and valid assessments of behavioral regulation are vital for early childhood researchers and the practitioners who work to ensure the well-being of every child as they transition to formal school settings, which pose ever increasing attentional and behavioral demands (Entwisle & Alexander, 1998; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005). The aim of this study was to create an easy-to-administer behavioral regulation task, which assessed behaviors similar to those required in classroom settings. We developed the Head-to-Toes Task, a direct measure adapted from the "Head and Feet" task described by McCabe, Cunnington, and Brooks-Gunn (2004a). The Head-to-Toes Task taps behavioral regulation, which includes inhibitory control, attention, and working memory, and contributes to children's ability to control and change their own behavior. We examined task reliability and validity, as well as longitudinal data, background correlates, and teacher reports of behavior for 445 children (30% ethnic minority) from two sites, including a group of Spanish-speaking children assessed in Spanish.

1. Direct measures of behavioral regulation

Traditional assessments of behavioral regulation have often utilized caregiver reports (Howse et al., 2003a; McClelland et al., 2000; Schultz, Izard, Ackerman, & Youngstrom, 2001). Increasingly common are direct observational measures, which demonstrate construct validity and predict social and academic outcomes without the potential for observer bias (Diamond, Kirkham, & Amso, 2002; Gathercole & Pickering, 2000; Hongwanishkul et al., 2005; Kochanska, Coy, & Murray, 2001; Manly et al., 2001; McCabe, Hernandez, Lara, & Brooks-Gunn, 2000; McCabe, Rebello-Britto, Hernandez, & Brooks-Gunn, 2004b; Pickering & Gathercole, 2004; Welsh, Pennington, & Groisser, 1991). Nevertheless, a review of existing observational instruments revealed a number of shortcomings because many tasks were designed for use in the laboratory or with clinical populations (Fahie & Symons, 2003; Pickering & Gathercole, 2004).

Table 1 displays thirteen commonly used direct assessments of aspects of behavioral regulation (see also Carlson, 2005, for a discussion of additional measures). All measures except Luria's Handgame require specialized materials

Table 1
Tasks measuring behavioral regulation

Task	Recent reference; sample age; sample *N*	Description	Time and materials required	Scoring
Attention Network Task	Rueda et al. (2005); 6–9 years; 48	A modified flanker task using fish: children are asked to help feed the central fish by pressing a button corresponding to the direction which the middle fish is swimming.	30 min; computer and software	Reaction time
Bear and Dragon	Carlson et al. (2004); 3–4 years; 49	Children are given commands by a bear and dragon puppet, and told to respond only to the bear's commands.	Several minutes; puppets	0–10
Day–Night Stroop	Berwid et al. (2005); 3.5–7 years; 71	Children are shown 16 cards on which "day" or "night" is pictured; they are instructed to say "day" to the night card and "night" to the day card.	4–5 min; 16 cards, videotape, stopwatch	0–16
Dimensional Change Card Sorting Task	Müller, Dick, Gela, Overton, and Zelazo (2006); 3–5 years; 52	Cards with colored shapes that can be sorted differently, either by color or shape. Children are first told to sort by one dimension and then the other.	About 10 min; cards with pictures	0–12
Go/No-Go Task	Simpson and Riggs (2006); 3.5 years; 120	Children must press or stop from pressing a button when a particular stimulus occurs.	15–20 min; apparatus and/or laptop	0–10
Grass and Snow	Carlson (2005); 3–5 years; 262	Children must point to a white card when the experimenter says "Grass," and a green card when she says "Snow."	5–10 min; colored cards	0–16
Luria's Hand-Game	Fahie and Symons (2003); 5–9 years; 26	Children first copy the researcher when he or she knocks on the table or points a finger; children are then instructed to point when the researcher makes a fist and vice versa.	Several minutes	0–2 (pass/fail)
The Matching Familiar Figures Test	Welsh et al. (1991); 3–12 years; 100	Out of six figures, children choose the one that most precisely matches the standard. A speed-accuracy tradeoff typically produces more errors.	Several minutes; pictures for matching	1–4
Self-Ordered Pointing	Hongwanishkul et al. (2005); 3–5 years; 98	Children are presented a binder composed of pages divided into four sections, children are instructed to go through each page and point to a different picture on each page.	15–20 min; binder with pictures	Number of items remembered (until failure)
The Test of Everyday Attention	Manly et al. (2001); 6–16 years; 293	It is a non-computerized measure that uses game-like tests to assess three different forms of attention. Battery costs $530.	60 min; administration materials, stimulus cards	Standard scores are calculated
Tower	Kochanska et al. (1996); 2–3.5 years; 99	Children are asked to build a tower with an experimenter, and shown how to take turns adding blocks to the tower. Score is based on the number of turns children offer to the experimenter.	20 blocks of varying sizes	0–0.5 (proportion of blocks placed by tester)
Tower of London	Ward, Shum, McKinlay, Baker-Tweney, and Wallace (2005); 7–21 years; 90	Children are asked to make their tower match a picture, after the trial run they are asked to make their tower match the picture as quickly as possible, with a restriction of number of times they can move the balls.	15–20 min; book with tower pictures, ball and peg apparatus, stopwatch	0–27
Working Memory Test Battery for Children	Pickering and Gathercole (2004); 4–15 years; 83	Children are tested individually in three sessions over a 10-day period. Of the ten tests, nine are working memory subtests. Testing kit costs approximately $260.00	60 min; manual, record form, mazes	Standard are scores calculated with norms

and/or substantial time to administer (Hughes, 1998). In only a few tasks – Bear and Dragon (Carlson, Moses, & Claxton, 2004; Reed, Pien, & Rothbart, 1984), Grass and Snow (Carlson, 2005), and Tower (Kochanska, Murray, Jacques, & Koenig, 1996) – are children asked to respond with gross motor movements. Assessments requiring fine motor coordination may mask young children's capabilities (i.e., they fail because they do not have the fine motor coordination to complete the task, not because they lack behavioral regulation skills). In the Head-to-Toes Task, children respond in ways that are comparable to demands in classrooms, namely controlling and directing overt behavior.

Different research teams have developed and reported on early childhood measures of behavioral regulation and inhibitory control (Carlson, 2005; Carlson et al., 2004; Kochanska et al., 2001, 1996). The Head-to-Toes Task incorporates features from many of these assessments, including a game-like format and responding in an unnatural way. The task resembles traditional measures of inhibitory control such as the Bear and Dragon task (Reed et al., 1984), and Luria's Handgame (Hughes, 1998), and also requires executive control aspects of attention and working memory (Rueda et al., 2004). The tasks in Table 1 have proven valuable, but we found that most reports lacked information on variability, validity, and reliability (Arizmendi, Paulsen, & Domino, 1981). For example, samples did not include large numbers of children whose ages spanned early childhood (i.e., 3–6 years), nor were data about the range of scores at different ages provided, since many studies were cross-sectional.

Except for gender, relations between performance and contextual variables, including socioeconomic status (SES), ethnicity, native language, and childcare experience was not typically assessed, nor was task correspondence with other indicators such as teacher reports (Simpson & Riggs, 2006). Hence, it is not clear how these widely utilized tasks relate to observer reports of children's behavioral regulation or with common background characteristics (Isquith, Crawford, Espy, & Gioia, 2005). We did find consistent information on task reliability, but most assessments have not been used with a large sample of typically developing children. In this study, we examined the variability, construct and convergent validity with observer reports, and the reliability of the Head-to-Toes Task, administered to 445 children from two sites. We also sought to determine if the task was sensitive to developmental and individual influences in 3–6-year-old children.

2. Individual and ecological factors in behavioral regulation

Sociocultural influences should be considered with new observational assessments, especially given the diversity in SES and culture that typifies American children and families (McCabe et al., 2000). Some research has found that ethnic minority children exhibit lower levels of behavioral regulation, but these effects often disappear after controlling for SES (Connell & Prinz, 2002). Although fairly uncommon, it is desirable to administer measures in Spanish to Spanish-speaking children (Buela-Casal, Carretero-Dios, De los Santos-Roig, & Bermúdez, 2003; Hart & Risley, 1995; Isquith et al., 2005; McCabe et al., 2004b).

In addition to sociocultural differences, girls have been found to have stronger inhibitory control, persistence, and more adaptive classroom behavior compared to boys (Deater-Deckard, Petrill, Thompson, & DeThorne, 2006; Kochanska et al., 2001; McCabe et al., 2004a; McClelland et al., 2000; Ready, LoGerfo, Burkam, & Lee, 2005). However, this has not been supported in some research on aspects of attention or working memory (Brocki & Bohlin, 2004; Carlson, Moses, & Breton, 2002; Renninger & Wozniak, 1985). We also do not know how or when gender differences in behavioral regulation emerge as children develop. In the present study, we examined gender differences in overall performance and rates of improvement on the Head-to-Toes Task, expecting that girls would perform better overall on the measure, which has a strong inhibitory control component.

Although space limits a full discussion of interactions among childcare, family factors, and child characteristics affecting behavioral regulation, including findings from the National Institute of Child Health and Development (NICHD) Early Child Care Research Network (ECCRN), research suggests that childcare experience may help children practice behavioral regulation (Connell & Prinz, 2002; Fabes, Hanish, & Martin, 2003a; NICHD ECCRN, 2003a; NICHD ECCRN, 2004; Wachs, Gurkas, & Kontos, 2004). In the present study, we examined the effects of sociocultural variables, gender, and months of prior childcare experience on children's performance and rates of improvement on the Head-to-Toes Task.

Finally, because longitudinal assessment and educational data are usually nested (e.g., assessments are nested within individuals over time, and children are nested within classrooms), different units of analysis are important to consider (Haney, 1980; Raudenbush & Bryk, 2002). Multi-level modeling statistically assesses different sources of variance in children's skills. In the present study, up to four task administrations of the behavioral regulation assessment were nested within each child, and on average, four children were nested in each classroom. Thus, we considered development, child, and classroom sources, and predicted variation with background variables such as gender and ethnicity to explain differences in Head-to-Toes Task scores.

3. Rationale and hypotheses

The purpose of this study was to develop a direct measure of behavioral regulation, sensitive to developmental as well as individual differences, which might be used to supplement popular survey measures (Meisels, 2006). We asked the following research questions. First, is the Head-to-Toes Task a valid and reliable measure of behavioral regulation during the early childhood period? Second, how much variation is due to development, differences among children, and among classrooms? And third, how do Head-to-Toes Task scores vary with key background and sociocultural characteristics?

We first hypothesized that the Head-to-Toes Task would show evidence of substantial individual and developmental variability and demonstrate acceptable inter-examiner reliability and convergent validity by correlating with teacher reports. Such results would substantiate the construct validity and reliability of the Head-to-Toes Task (Anastasi, 1992). Second, we anticipated that significant variation in scores would be attributed to development, child characteristics, and classroom-level influences. Third, we predicted that membership in a minority group and Spanish administration of the task would not influence performance on the Head-to-Toes Task, after the effect of parent education as a proxy for SES was considered. We expected girls to exhibit stronger performance compared to boys on the task, although in the absence of prior work demonstrating otherwise, we predicted girls and boys to improve at similar rates. Finally, participants with prior exposure to childcare and those with parents with higher education levels were expected to earn higher scores on the Head-to-Toes Task, compared with children with less childcare experience and parents with less education.

4. Method

4.1. Participants

Children were recruited from two sites, drawn from a heterogeneous community near a major metropolitan area in Michigan ($N = 353$), and a university area in a small city in Oregon ($N = 92$). The percentages of non-Caucasian ethnic minority participants in the two sites (Michigan and Oregon) were 24% and 52%, respectively, making the overall percentage of ethnic minority participants 30%.

In Michigan, children were recruited through their preschool programs, located within a single district, to participate in a 5-year longitudinal study of early school transition. Preschools were part of a district-sponsored fee-based program housed within six elementary school buildings; the district was initially approached and all schools agreed to participate in the longitudinal study. One school was a Title One site for two Head Start classrooms, and the other preschool classrooms were licensed by the state of Michigan. All 3- and 4-year-old children entering preschool were recruited through fall orientations and backpack mail during the first 2 years; children from the Title One school were also recruited during the third year. Once the target sample size had been achieved, recruiting efforts were stopped. This procedure resulted in approximately 38% of the district's children enrolled in preschool during the first year participating in the longitudinal study. The present study reports data from the second (Times 1 and 2 for this study) and third (Times 3 and 4) years of the overall study. Children were nested in 62 classrooms, including 36 preschool and 26 kindergarten classrooms, at Times 1 and 2, and were followed into 97 classrooms, including 24 preschool, 35 kindergarten, and 38 first grade classrooms, at Times 3 and 4.

In Oregon, children and parents were part of a study investigating factors related to preschoolers' behavioral functioning. Three NAEYC-accredited and three Head Start preschools were invited to participate in the study and all agreed. The three NAEYC-accredited preschools were center-based community preschools and all six preschools included 3- and 4-year-olds. Letters inviting parents of children entering the prekindergarten year were sent home from school with approximately 165 four-year-olds, and 95 families (or 58%) agreed to participate. These 95 children and their families were studied during the fall and spring of their prekindergarten and kindergarten years. Two participants were excluded because neither English nor Spanish was their first language, and one was missing all testing data, making the final Oregon sample 92. Children were nested in 12 classrooms during the prekindergarten year (Times 1 and 2 for this study) and followed into 25 kindergarten classrooms (Times 3 and 4).

At both sites, teachers were recruited to participate if they had a study child in their class. In Michigan, 99 teachers were recruited and all participated in the study. In Oregon, 42 teachers were recruited and all participated. Teachers received a $20 gift card for each classroom they taught with enrolled child participants.

At the Michigan site, the sample consisted of 353 children (178 girls). The average education level attained for both mothers and fathers was a bachelor's degree ($M = 16.0$; S.D. = 1.6). On average, children had spent about 13.9 months in childcare prior to data collection at Time 1 (S.D. = 16.15). Of the 319 children for whom we could obtain ethnicity, 76% were Caucasian, 8% were African American, 7.5% were East Indian or Asian, 2.5% were Latino/a, and 6% were Middle Eastern. The mean ages of children over Times 1–4 were 53, 60, 66, and 72 months, respectively.

At the Oregon site, 92 children (48 girls) participated. The mean parent education level was an associate's degree ($M = 14.6$ years; S.D. = 3.9) and 24% of parents had a high-school education or less ($N = 22$). On average, children had spent about 22 months in childcare prior to data collection at Time 1 (S.D. = 18.07 months). Twenty-five percent ($N = 23$) of children and families were Latino/a, of which 19 (83%) were primary Spanish-speakers; teachers identified the Spanish-speaking children in their class. The remaining sample was 48% Caucasian, 19% Asian, and 8% other ethnic groups. The mean ages of children at Times 1, 2, 3, and 4, were 55, 62, 68, and 74 months, respectively. Table 2 presents descriptives for each sample.

We conducted simple comparisons of the two sites, using an alpha level of $p < 0.05$ for all tests, and found five differences, which are also shown in Table 2. Compared with Michigan children, Oregon participants were significantly older by 2 months (Time 1, $t = 3.59$, $p < 0.001$). Second, the proportion of total ethnic minority participants was higher in Oregon than in Michigan, $t = 5.40$, $p < 0.01$. Analysis of each site's ethnic make-up revealed that there was a significantly greater proportion of African American/biracial and Middle Eastern children in Michigan and a significantly greater proportion of Latino/a and Asian children in Oregon. In focused comparisons, Caucasian and Asian children's Head-to-Toes Task scores were not significantly different from one another, although they were significantly higher than scores of children in other (i.e., African American, Middle Eastern, and Latino/a) minority ethnic groups. Thus we created a non-Asian minority variable to use in subsequent analyses. Third, on average, parents in Michigan had significantly more years of education than did parents in Oregon, with 15.9 years versus 14.6 years, $t = -2.85$, $p < 0.01$. Fourth, Oregon children spent more time outside the home in childcare, $t = 4.34$, $p < 0.001$. Finally, Head-to-Toes Task scores did not differ significantly by site at Times 1–3, but Michigan participants scored higher at a marginal level of significance at Time 4, $t = -1.99$, $p = 0.051$.

Table 2
Descriptive statistics by site

	Michigan ($N = 353$)			Oregon ($N = 92$)		
	M	S.D.	Range	*M*	S.D.	Range
1. Age Time 1 (months)[*]	53	9.8	34–71	55	3.3	49–62
2. Non-Asian minority[*,a]	0.17	0.38	0–1	0.33	0.47	0–1
3. Parent education (years)[*]	15.95	1.62	11–18	14.64	3.88	4–21
4. Male[b]	0.49	0.50	0–1	0.49	0.50	0–1
5. Childcare (months)[*]	12.00	15.09	0–50	22.14	18.0	1–57
6. Missing data[c]	0.20	0.40	0–1	0.21	0.41	0–1
7. CBRS Time 2	3.80	0.65	1.63–5	4.13	0.64	2.88–5
8. CBRS Time 4	3.93	0.69	2.00–5	4.10	0.59	2.50–5
9. HTT Time 1	10.02	8.01	0–20	8.86	7.52	0–20
10. HTT Time 2	12.77	7.03	0–20	13.23	6.83	0–20
11. HTT Time 3	15.15	5.77	0–20	15.06	5.34	0–20
12. HTT Time 4[*]	17.24	4.57	0–20	15.55	5.95	0–20
13. Spanish-speaking[*,d]	–	–	–	0.21	0.41	0–1

Note: HTT—Head-to-Toes Task; CBRS: Child Behavior Rating Scale.
[a] Non-Asian minority = 0; minority = 1.
[b] Female = 0; male = 1.
[c] Complete data = 1; missing data = 0.
[d] Spanish-speaking = 1; English-speaking = 0.
[*] $p < 0.05$.

4.2. Procedure

At both sites in the fall and spring of each year, research assistants administered the Head-to-Toes Task as part of a battery of assessments. Each battery was administered in individual sessions lasting 30–40 min in Michigan and 10–15 min in Oregon. The Head-to-Toes Task took about 5 min to administer and in Oregon, Spanish-speaking children were given the task in Spanish. At both sites, data were collected in the fall and spring over 2 years, resulting in four administrations on average 6.1 months apart. Assessments were completed at children's schools in a quiet hallway outside their classroom. In Michigan, participants were given stickers as they went through the battery. At both sites, teachers completed child behavior ratings in the spring.

4.3. Measures

4.3.1. Questionnaires

In the first year of the study, parents at both sites completed a background questionnaire in English or Spanish (Oregon site only) which contained questions about children's age, gender, childcare experience, ethnicity, and parent education level. In Oregon, the background questionnaire was translated into Spanish and back translated into English for Spanish-speaking parents by two native Spanish-speakers. Families in both sites received a total of $20 in gift certificates per year for participating in any part of the study.

To assess children's behavioral regulation, teachers at both sites completed the Child Behavior Rating Scale (Bronson, Tivnan, & Seppanen, 1995) in the spring (Times 2 and 4). The CBRS contains eight items chosen to assess behavioral regulation from an overall mastery behaviors scale, which has shown strong reliability in previous research (McClelland & Morrison, 2003). Example items include "Complies with adult directives, giving little or no verbal or physical resistance, even with tasks," "Observes rules and follows directions without requiring repeat reminders," and "Completes learning tasks involving two or more steps (e.g., cutting and pasting) in an organized way." In the present study, the internal consistency reliability for the eight behavioral regulation items was 0.91 (Time 2) and 0.92 (Time 4) in Michigan, and 0.87 (Time 2) and 0.89 (Time 4) in Oregon.

4.3.2. Direct measure of behavioral regulation

At both sites, we used the Head-to-Toes Task as a direct assessment requiring three aspects of behavioral regulation (inhibitory control, attention, and working memory). Children were asked to play a game where they were instructed to touch their head, and then to do the opposite, and touch their toes. After two questions to check understanding, children were given four practice tests and the instructions were repeated up to three times during the practice tests. Then ten test commands were given verbally in random order, without feedback. Children received two points for a correct response, one point for a self-correct, defined as making any discernable motion (ranging from slight to complete) toward the incorrect response but ending with the correct response, and zero points for an incorrect response. Thus, the total score possible on ten trials was 20 points. During the testing portion, the experimenter stated the behavioral commands without modeling any actions.

At the Oregon site, the measure was translated into Spanish and back translated into English by two Spanish-speakers, including a professor of Spanish. Spanish-speaking children in Oregon were given the task in Spanish. We used a dummy variable denoting the language in which children received the Head-to-Toes Task as a control in analyses.

5. Results

5.1. Variability and psychometric properties of the Head-to-Toes Task

Our first goal was to examine variability in children's scores and establish construct validity and reliability of the Head-to-Toes Task. Variability was present for all ages of children, with the widest range emerging in scores of 48–60-month-olds. In addition, data were fairly bimodal, with most children scoring relatively low or high on the task. To assess the normality of the distributions, we calculated skewness and kurtosis within each site. General guidelines suggest kurtosis absolute values greater than 8.0 and skewness absolute values greater than 3.0 indicate severe deviations from normality (Kline, 2005). As shown in Table 3, skewness and kurtosis values were not extreme for children across

Table 3
Distribution of scores on the head-to-toes task (Time 1)

Age group (site)	*N*	Skewness	Kurtosis	% scoring at floor	% scoring at ceiling
36 months (MI)	37	0.65	−1.67	65	0
42 months (MI)	54	0.53	−0.96	56	2
48 months (MI)	54	.02	−0.44	33	6
48 months (OR)	30	−0.30	−0.30	30	3
54 months (MI)	69	−0.16	1.78	13	7
54 months (OR)	48	−0.35	1.48	17	6
60 months (MI)	44	0.85	0.86	2	20
60 months (OR)	14	−1.57	0.50	21	0
66 months (MI)	36	1.87	1.57	0	17

Note: Percent is based on the number of children in that age group. MI: Michigan; OR: Oregon.

ages. In addition to these values, percentages of participants scoring at floor or ceiling levels were comparable across sites.

Although more than half of the children who were age 36 months at the first task administration in Michigan scored zero on the task, only about a third of 48-month-olds and about a sixth of 54-month-olds in both sites scored zero on the task (see Table 3). At Time 1, the majority of children scored at least one point on either the practice or test portion of the task: 72% of all 3-year-olds, 91% of all 4-year-olds, and all except one 5-year-old.

We also graphed developmental trends across 6-month age groups (e.g., 36-month-olds, 42-month-olds, 48-month-olds, etc.) for each time point in Fig. 1. Average scores are shown on the *y*-axis for different age groups. Note that this differs from the standard practice of graphing time on the *y*-axis, but more easily shows potential differences in scores across multiple task administrations. Fig. 1 demonstrates relative consistency in scores for children in each age group, across up to four task administrations. An analysis of Head-to-Toes Task scores by task administration number (Times 1–4) and age group demonstrated scores differed by 6-month age group, $F(8, 1320) = 29.55$, $p < 0.01$, but not task administration, $F(3, 1320) = 1.45$, $p = 0.23$; the interaction of age group by task administration was not significant, $F(18, 1320) = 1.30$, $p = 0.18$. This suggests age group, but not number of task administrations, was associated with improvement on the task, and that this pattern was the same across age groups.

In addition to demonstrating construct validity by showing a mean increase in scores as children developed, we sought to establish convergent validity for the Head-to-Toes Task by comparing task scores to CBRS ratings. *At the Michigan site*, children who received high scores on the Head-to-Toes Task were rated higher by their teachers on the eight behavioral regulation items of the CBRS at Time 1, $r = 0.20$, $p < 0.01$, Time 2, $r = 0.15$, $p < 0.05$, Time 3, $r = 0.19$, $p < 0.05$, and Time 4, $r = 0.15$, $p < 0.05$. *At the Oregon site*, higher scores on the Head-to-Toes Task were related to higher teacher ratings on the CBRS at Time 1, $r = 0.42$, $p < 0.01$, Time 2, $r = 0.47$, $p < 0.01$, and Time 3, $r = 0.26$, $p = 0.06$, but not Time 4, $r = 0.10$, $p = 0.48$. Correlations for the Michigan sample are shown in Table 4, and correlations for the Oregon sample are shown in Table 5. Correlations among background variables, Head-to-Toes Task scores,

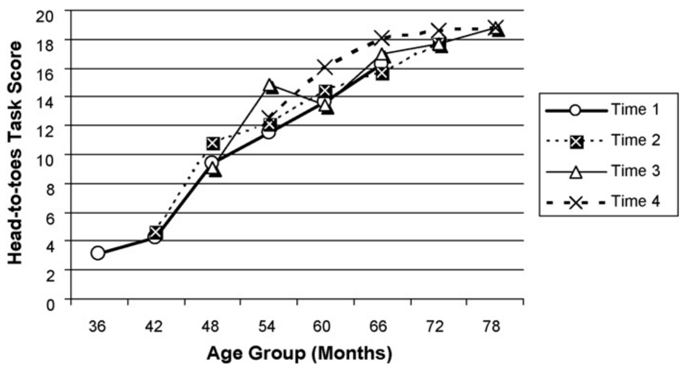

Fig. 1. Developmental trends across age groups for four task administrations (Times 1–4) of Head-to-Toes. Scores are based on the chronological age group into which children fell when they were assessed.

Table 4
Correlations for the Michigan site (maximum $N = 353$)

	1	2	3	4	5	6
1. Age Time 1 (months)	–	0.01	−0.05	0.08	0.10[t]	−0.06
2. Non-Asian minority[a]		–	−0.17**	0.06	−0.02	0.18**
3. Parent education (years)			–	0.03	0.17**	[c]
4. Male[b]				–	0.001	0.004
5. Childcare (months)					–	−0.06
6. Missing data[d]						–
7. CBRS Year 1						
8. CBRS Year 2						
9. HTT Time 1						
10. HTT Time 2						
11. HTT Time 3						
12. HTT Time 4						

	7	8	9	10	11	12
1. Age Time 1 (years)	0.15*	0.10	0.62**	0.57**	0.50**	0.42**
2. Non-Asian minority[a]	−0.07	−0.20**	−0.13*	−0.13*	−0.13*	−0.12[t]
3. Parent education (years)	0.12[t]	0.06	−0.01	−0.002	0.07	0.04
4. Male[b]	−0.23**	−0.29**	0.03	−0.05	−0.08	−0.14*
5. Childcare (months)	−0.03	−0.07	0.08	0.13*	0.11	0.05
6. Missing data	−0.07	0.02	−0.13*	−0.13*	−0.22**	−0.13*
7. CBRS Year 1	–	0.49**	0.20**	0.15*	0.20**	0.14[t]
8. CBRS Year 2		–	0.20**	0.19*	0.19*	0.15*
9. HTT Time 1			–	0.57**	0.45**	0.36**
10. HTT Time 2				–	0.59**	0.48**
11. HTT Time 3					–	0.61**
12. HTT Time 4						–

Note: HTT—Head-to-Toes Task; CBRS—Child Behavior Rating Scale.
[a] Non-Asian minority = 0; minority = 1.
[b] Female = 0; male = 1.
[c] Invalid because missing data were calculated based on parent education.
[d] Complete data = 0; missing data = 1.
[t] $p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.

and CBRS ratings demonstrated a similar pattern across sites. However, the correlation between Oregon Head-to-Toes Task Time 2 and Time 3 scores was lower than other within-task correlations, at $r = 0.29$, $p < 0.05$.

Next we assessed reliability across examiners for the Head-to-Toes Task scores for both sites. *At the Michigan site*, looking within schools because examiners were assigned to assess children in particular schools, there were no significant differences in average Head-to-Toes Task scores at any task administration. *At the Oregon site*, no significant differences in Head-to-Toes Task scores across examiners (also looking within schools) were found at Times 1–3, but a significant difference was found at Time 4, $F(7, 51) = 2.95$; $p < 0.05$. This was due to one examiner, who only tested Spanish-speaking children. These children scored lower on the Head-to-Toes Task compared to other children in the study (see Table 5). We also tested for examiner differences in scoring self-corrected responses on the Head-to-Toes Task. *At the Michigan site*, there were no significant differences in the average number of self-corrects scored by examiner within schools at Times 1–3. At Time 4, a significant difference emerged, $F(5, 242) = 5.26$, $p < 0.01$, and closer inspection revealed that two out of six raters gave significantly higher average numbers of self-corrected responses, although this did not produce a difference in overall score. *At the Oregon site*, there were no significant differences by examiner in the average number of self-corrected responses on the Head-to-Toes Task at Times 1–4. Finally, in another study, inter-rater reliability was obtained by having six examiners score videotapes of twelve children, selected at random, completing the Head-to-Toes Task (Connor et al., 2007). Examiners achieved excellent inter-rater reliability (alpha = 0.95 for self-corrections, 0.98 overall). Overall, these results meet generally established reliability levels (Landis & Koch, 1977).

Table 5
Correlations for the Oregon site (maximum $N = 92$)

	1	2	3	4	5	6
1. Age Time 1 (months)	–	0.30**	−0.11	0.003	0.19	−0.08
2. Non-Asian minority[a]		–	−0.70**	0.06	−0.14	0.13
3. Parent Ed. (years)			–	−0.18	0.33**	[c]
4. Male[b]				–	−0.18	0.33**
5. Childcare (months)					–	−0.04
6. Missing data[d]						–
7. CBRS Year 1						
8. CBRS Year 2						
9. HTT Time 1						
10. HTT Time 2						
11. HTT Time 3						
12. HTT Time 4						
13. Spanish-speaking[e]						

	7	8	9	10	11	12	13
1. Age Time 1 (years)	0.18[t]	0.09	0.18[t]	0.03	0.24[t]	0.21	0.20[t]
2. Non-Asian minority[a]	0.004	−0.04	−0.002	−0.15	−0.23[t]	−0.27*	0.73**
3. Parent Ed. (years)	0.04	0.24[t]	0.18	0.34**	0.32*	0.32*	−0.67**
4. Male[b]	−0.13	−0.30*	−0.11	−0.14	−0.09	−0.01	0.15
5. Childcare (months)	−0.11	−0.14	0.20[t]	0.18	0.15	0.14	−0.18
6. Missing data[d]	−0.07	−0.09	−0.16	−0.17	−0.18	−0.20	0.22*
7. CBRS Year 1	–	0.51**	0.42**	0.47**	0.38**	0.40**	−0.15
8. CBRS Year 2		–	0.33*	0.46**	0.26[t]	0.10	−0.04
9. HTT Time 1			–	0.54**	0.35**	0.51**	−0.20[t]
10. HTT Time 2				–	0.29*	0.39**	−0.31**
11. HTT Time 3					–	0.71**	−0.35**
12. HTT Time 4						–	−0.45**
13. Spanish-speaking[e]							–

Note: HTT—Head-to-Toes Task; CBRS—Child Behavior Rating Scale.
[a] Non-Asian minority = 0; minority = 1.
[b] Female = 0; male = 1.
[c] Invalid because missing data were calculated based on parent education.
[d] Complete data = 0; missing data = 1.
[e] English-speaking = 0; Spanish-speaking = 1.
[t] $p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.

5.2. Sources and predictors of variation in behavioral regulation

Our second question asked about the amount of variability in behavioral regulation within individuals over time (i.e., variability due to development), among children, and among classrooms, after taking into account background predictors. Our third research question examined the relation of sociocultural variables, gender, and childcare experience, with Head-to-Toes Task scores. We investigated both questions by modeling growth curves using a single cross-classified random effects feature, known as HCM2, in the HLM 6.02 program (Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, & Congdon, 2004). Behavioral regulation data were cross classified, because most children were members of two classrooms during the study. We included an examination of site effects, which demonstrated consistency in our findings across site and provided justification for including both samples in the same analysis.

In the growth curve analysis in HCM2, multiple task administrations were modeled at level-1, and child characteristics were modeled at level-2. The program applies a full-maximum likelihood procedure that uses each score to calculate fitted growth curves based on the child's exact age at score attainment. The average score (intercept) as well as an average effect of age, or growth rate (slope), and quadratic trend (acceleration) are modeled.

The technique is robust to non-normal data distributions, as long as model residuals follow a normal distribution (Raudenbush & Bryk, 2002). A histogram of the residuals from the final model (see final model below) upheld the assumption of normality. One assumption that holds from traditional regression techniques is homogeneity of level-1 residual variance; we note that variance on the task decreased over time because more children scored near ceiling levels. It is not yet possible to assess the homogeneity assumption with the cross-classified program, but a model representing a good fit to the data likely accounts for any heterogeneous variance across level-1 scores using predictors such as growth rate and acceleration. Finally, even if the variance assumption were violated, coefficients for child background predictors will be relatively unchanged, although the estimation process may be less efficient (Raudenbush et al., 2004).

Missing data for parent education and childcare experience were imputed with the expectation maximization (EM) algorithm from SPSS, which is a maximum likelihood procedure that uses all available background (level-2) variables on participants to iteratively calculate missing values (Acock, 2005). A dummy variable was then created and coded "1" for children missing data (20% at each site). Total sample sizes for the analysis were $N = 1472$, child $N = 445$, and classroom $N = 213$. Seventeen children (12 in Michigan; 5 in Oregon) were assessed in the laboratory and assumed to be in their own classroom (17 + 159 Michigan classrooms + 37 Oregon classrooms = 213).

Prior to building the final model, we assessed the extent of variation in intercept, growth rate, and acceleration at different levels. We therefore created a level-1 growth model without level-2 predictors (Model 1). The final model (Model 2; see below) then incorporated predictors of the intercept and the acceleration term. These are variables producing deflections from the mean score or mean acceleration at age t (centered at 48 months), holding all other variables constant at their mean or reference value (for dummy variables, the reference value was zero). We added predictors, one at a time, first to the intercept, then to the growth rate, and finally to acceleration, trimming predictors that did not reach significance to maintain parsimony (Raudenbush & Bryk, 2002). We also ran the model with age centered at the grand mean (63 months) and found highly similar results. We report the final model, shown below, with age centered at 48 months because it is a meaningful reference point for interpreting findings, since many children enter preschool around age 4 years, and undergo significant development in the skills underlying behavioral regulation during this period (Bronson, 2000; Diamond et al., 2002; Shonkoff & Phillips, 2000).

$$\text{Level 1}: \quad Y_{tjk} = \pi_{0jk} + \pi_{1jk}(\text{child age}) + \pi_{2jk}(\text{quadratic of child age}) + e_{jk} \tag{1}$$

$$\text{Level 2}: \quad \begin{aligned} \pi_{0jk} &= \theta_0 + b_{00} + c_{00} + (\gamma_{01})\,\text{non-Asian minority} + (\gamma_{02})\,\text{Spanish version} \\ &\quad + (\gamma_{03})\,\text{SES (parent education)} + (\gamma_{04})\,\text{gender} + (\gamma_{05})\,\text{missing data} \\ \pi_{1jk} &= \theta_1 \\ \pi_{2jk} &= \theta_2 + (\gamma_{22})\,\text{SES (parent education)} + (\gamma_{23})\,\text{missing data} \end{aligned} \tag{2}$$

In Eq. (1), a child's expected score at age t is made up of four parts—the intercept (π_{0jk}), slope or growth rate calculated from child age (π_{1jk}), acceleration or quadratic of child age (π_{2jk}), and error for the individual score (e_{jk}). In Eq. (2), the intercept, or average score at age t, is in turn comprised of the grand mean of all scores, plus random child and classroom effects, plus the effect of child-level predictors such as SES and gender. Random effects associated with child j, averaged across all classrooms, are represented by b_{00}, and c_{00} is the random classroom effect associated with membership in classroom k, averaged across children. The slope, or growth rate (π_{1jk}), is defined as the average growth rate at age t, and the acceleration (π_{2jk}) is defined as the average acceleration at age t, plus the effects of SES and having missing background data. Because of the relatively small number of children in classrooms (on average four), the error terms for growth rate and acceleration were fixed across children and classrooms (Raudenbush et al., 2004).

5.3. *Sources of variation in behavioral regulation: random effects*

To address our second research question asking about variance at three different levels, we used results from the random effects portion of Table 6. The initial model, Model 1, which included growth and acceleration, explained 11.0% of the variance among individual children's different test scores (level-1 error variance), 21.0% of the variance among children, and 89.5% of the variance among classrooms. The final model, Model 2, explained 18.5% of the variance remaining among children after accounting for growth, and 66.2% of the variance remaining among classrooms. Overall, the final model, including all significant predictors, explained 41.4% of the total variance, and 8.1% of the

Table 6
Cross-classified growth curves modeling head-to-toes task scores, age centered at 48 months

Fixed effects	Model 1[a]			Model 2			
Predictor	Coeff.	d.f.	t	Coeff	d.f.	t	d[b]
Intercept	13.50	1469	61.96***	14.65	1462	50.63***	
Non-Asian minority				−1.04	1462	−1.94[t]	0.06
Male				−1.07	1462	−2.89**	0.08
Spanish version				−2.92	1462	−2.43*	0.08
Missing data				−1.56	1462	−3.23**	0.09
Parent education				0.23	1462	2.13*	0.07
Linear growth rate (slope)							
Intercept	0.59	1469	18.55***	0.60	1462	20.59***	0.95[c]
Quadratic trend (acceleration)							
Intercept	−0.01	1469	−7.80***	−0.01	1462	−9.13***	0.10
Missing data				<0.01	1462	2.09*	<0.01
Parent education				<0.01	1462	−0.70 (ns)	

Random effects	Model 1			Model 2		
	Variance	d.f.	χ^2	Variance	d.f.	χ^2
Times (e, intercept)	19.66			19.99		
Children (b_{00}, intercept)	10.21	444	1126.94***	8.32	439	1033.10***
Classrooms (c_{00}, intercept)	1.48	211	521.54***	0.50	207	327.46***

[a] We also ran the fully unconditional model (without the growth rate or acceleration), and obtained $e = 22.08$, $b_{00} = 12.92$, and $c_{00} = 14.16$.
[b] Effect sizes were calculated for a 1 S.D. change in the coefficient.
[c] The effect size for age represents the effect for a 1 S.D. increase in age (11.25 months), so corresponds roughly to the effect size associated with being 1 year older rather than 1 month older (the coefficients shown are for 1 month, as in the actual model results).
[t] $p = 0.052$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

variance in average Head-to-Toes Task scores after accounting for the effects of maturation. About four percent of the initial classroom-level variation remained, $\chi^2 = 327.46$, $p < 0.001$. Thus, significant differences among classrooms were present even after considering developmental and child characteristics.

5.4. *Predictors of behavioral regulation: fixed effects*

We also examined the extent to which child characteristics, such as gender and parent education, changed the predicted growth trajectory, controlling for all other variables. These fixed effects are shown in Table 6. The top right portion of Table 6 lists the coefficients for each predictor. The final model describes a fitted or mean growth curve where the predicted outcome for a child at a certain age depends on the intercept, slope, and acceleration. For “Fixed Effects,” coefficients under “Intercept” represent deflections from the average score at age 4 years associated with predictor values.

The sample average Head-to-Toes Task score for children at age four (grand mean) was 14.65, which differed significantly from zero, $t = 50.63$, $p < 0.001$. Receiving the Spanish version of the Head-to-Toes Task resulted in a significant negative deflection from this average score of almost three points, $t = -2.43$, $p < 0.05$, and being a boy was associated with scoring about one point lower than average, $t = -2.89$, $p < 0.01$. Being a member of a family missing background data was associated with scoring 1.6 points lower than average, $t = -3.23$, $p < 0.01$. Using the standard deviation of the outcome (7.06) to calculate effect size, these effects were small (see Table 6). Having a parent with two more years of education than average was associated with about a 0.5-point score increase, $t = 2.13$, $p < 0.05$. Finally, being minority (non-Asian) had a negative effect at a level of marginal significance, $t = -1.94$, $p = 0.052$.

To explore the nature and extent of the negative effect of the non-Asian minority variable on the intercept, we conducted a post-hoc analysis with an interaction between minority and site. After including this interaction in the model, the initial marginally significant effect of minority status disappeared, $t = 0.35$, $p = 0.59$. We graphed scores by site and minority status, and found that children from Oregon scored significantly lower compared to minority children from Michigan and non-minority participants from both groups, but only at Time 4.

No predictors affected the growth rate, which was positive for all children, $t = 20.59$, $p < 0.001$, such that children's scores increased, on average, about seven points per year with steadily decreasing rates of growth over time (i.e., negative quadratic trend suggests deceleration over time in rates of HTT growth). The missing parent education data dummy variable significantly predicted the quadratic trend. In general, the growth rate for children's scores flattened out (i.e., decelerated; see also Fig. 1) over time, but scores for children with complete data decelerated more rapidly, $t = -9.13$, $p < 0.001$, compared to scores for children with missing data, $t = 2.09$, $p < 0.05$. Parent education did not predict acceleration, but it was associated with the missing data variable and thus included in Model 2.

In sum, significant variability was found in children's behavioral regulation assessed directly with a task demonstrating reliability and validity. Except for age, effect sizes were small for child characteristics including gender, parent education (SES), being administered the task in Spanish, and non-Asian minority status. There were also no significant site differences.

6. Discussion

The purpose of this study was to develop a simple and easy-to-administer assessment of behavioral regulation, the Head-to-Toes Task. The task requires children to apply inhibitory control as well as attention and working memory as they respond to complex commands. Results indicated variability in scores. The measure may be most useful with children 4–5 years of age and valuable only as a screening tool for older children, who were more likely to score at ceiling levels. Future work with a nationally representative sample should assess longitudinally whether older children who do poorly on the Head-to-Toes Task develop other problems related to behavioral regulation, including ADHD, or exhibit poor academic performance (Bredekamp & Copple, 1997; Kalpidou, Power, Cherry, & Gottfried, 2004). These results highlight the challenge of creating increasingly complex versions of behavioral regulation assessments, which require multiple cognitive components but are also simple enough for younger children. This is evidenced by a dearth of measures of this domain spanning from early to middle childhood (S. E. Rimm-Kaufman, personal communication, June 23, 2006). Additional research could improve existing measures by incorporating an increasing number of rules to remember, which would broaden the scoring range for younger and older children. Nonetheless, the majority of scores for all participants showed variability and were between floor and ceiling levels.

The Head-to-Toes Task also demonstrated convergent validity with observer reports of behavioral regulation. Children who did well on the task received higher scores on a teacher-report measure, while those who did worse received lower teacher ratings. Correlations between task scores and teacher ratings were stronger in the Oregon site than in Michigan when Oregon children were younger (Times 1 and 2), but all correlations were statistically significant, except at Time 4 in the Oregon site. The size of each correlation at Time 4 was similar in both sites, but may have been non-significant in Oregon due to that site's smaller sample size. Stronger correlations in Oregon early in the study may be due to increased variability in that sample. In addition, preschool teachers in Oregon were at NAEYC-accredited and Head Start schools, whereas all but two Head Start teachers in Michigan were housed within a public school district without NAEYC accreditation. It is possible that compared with teachers in Michigan, teachers in Oregon had more training in developmentally appropriate practices or were more informed about age-appropriate behaviors (Bredekamp & Copple, 1997). This may have produced a stronger match between their ratings of child behavior and Head-to-Toes Task performance.[1]

It is also important to note that in general, the strength of these correlations was somewhat weak in both sites, which may be associated with the lack of high predictive validity for skills assessed in early childhood, a time marked by individual differences, rapid gains in development, and inconsistency in performance (Pianta & McCoy, 1997). Limited convergent validity with teacher ratings may also signal different demands of the classroom environment, where children must respond to competing social and emotional goals while simultaneously regulating behavior. Thus, children may have behavioral regulation as evidenced by the Head-to-Toes Task, but may not be able to demonstrate these skills as well to teachers (Miller et al., 2003). Future psychometric efforts should include measures of discriminant

[1] We are grateful to an anonymous reviewer for making this point.

and convergent validity to adequately address this issue. For example, it would be useful to determine whether language development demonstrates discriminant validity with children's behavioral regulation.

Overall, results suggest that the Head-to-Toes Task is a reliable and valid measure of developing behavioral regulation in the early childhood period. There was also an absence of practice effects, since mean scores improved as children gained years, but not experience with the task. For the most part, task scores correlated positively, with children performing well at one time point maintaining high scores relative to other children at later time points. One exception was the correlation between Time 2 and Time 3 scores in the Oregon sample. This correlation may have been low because as children in this sample went from prekindergarten to kindergarten (i.e., from Time 2 to Time 3), more of them scored at ceiling levels on the Head-to-Toes Task—even those who had scored low in prekindergarten. Despite this finding, all other within-task correlations were fairly strong and positive across samples.

Moreover, a recent study revealed that prekindergarteners who achieved higher scores on the Head-to-Toes Task demonstrated significantly greater growth on early reading, math, and vocabulary in prekindergarten (McClelland, Cameron, Connor, et al., 2007). Thus, the Head-to-Toes Task appears to be an important indicator that children will succeed in school, and we offer three recommendations for its use in early educational settings. First, we encourage administering the entire task, including the practice portion, to individual children in a quiet space with limited distractions. Second, employing multiple types of assessments at different times is important during early childhood to gather a representative skill portrayal (Bredekamp & Copple, 1997; Meisels, 2006). Third, it is also necessary to interpret results while considering that variability in skills is typical during this developmental phase (Díaz, Neal, & Amaya-Williams, 1990; Shonkoff & Phillips, 2000). Care in administration and interpretation of the Head-to-Toes Task will also help to ensure skills are fairly and accurately assessed, and that appropriate aid can be provided if children consistently have difficulty regulating their behavior.

Significant variation arose from development, differences among individuals, and differences among classrooms; we identified some child and sociocultural characteristics that are associated with variation in average scores and deceleration. Classroom-level contributors to scores were untested in this study, but differences among classrooms remained even after considering development and child variables. This could be due to actual classroom (e.g., teacher practice) differences, aggregate effects of children in those classrooms, or community effects. Although it was beyond the scope of the paper to model classroom differences in growth rate or acceleration, the possibility that teachers may differentially promote acquisition of these skills is compelling. This limitation highlights an avenue for future research and intervention. Assessing whether certain classroom practices or activities promote behavioral regulation would be informative in bolstering skill development. However, not much is known about teaching factors that improve behavioral regulation. One recent intervention study demonstrated that teaching preschoolers with special needs problem-solving strategies increased their task persistence, which requires sustained attention and inhibitory control, on challenging tasks (Karnes, Johnson, & Beauchamp, 2005). Other work has revealed that practicing a game requiring complex rule use is associated with improvement in preschoolers' inhibitory control (Dowsett & Livesey, 2000). These studies suggest that directly teaching desired behaviors, as well as allowing children opportunities to practice behavioral regulation skills, are important for promoting their development, but we must emphasize that future studies employing randomized control trials in educational settings are needed to properly evaluate these implied mechanisms.

Chronological age was by far the strongest contributor to Head-to-Toes Task score gains. Before introducing ethnic minority status, Oregon participants scored lower, compared with Michigan participants, which may be due to the higher proportion of minority participants in Oregon. Site effects disappeared after considering ethnic minority status, which in turn fell to a level of marginal significance after introducing parent education and the missing data variable. Post-hoc analyses suggested that the effect associated with ethnic minority status was present for Oregon participants and only at the final time point. Furthermore, Spanish-speaking children had parents with the lowest education levels of all groups, which may help explain why receiving the Head-to-Toes Task in Spanish had an additional negative impact on scores. Other sociocultural factors, including risk factors not fully captured by the SES variable (parent education only), may account for the negative effects associated with the Spanish-speaking variable.

Future work should assess other markers of SES, including family income, occupational prestige, and employment status. Notably, the effect sizes of all sociocultural variables were small. Our findings, showing small effects for parent education, ethnicity, and administration language, suggest that the Head-to-Toes Task may be valid for children of diverse backgrounds. However, additional work with a more diverse sample of Spanish-speaking children is needed to confirm the validity of this measure's Spanish version. Future research should also include a nationally representative sample to better examine the psychometric properties of the Head-to-Toes Task as a measure of behavioral regulation and as a screening tool for older children. Accurate measures of behavioral regulation have become increasingly valuable as practitioners try to identify school-related competencies and needs of individuals (McCabe et al., 2000).

We found small but significant gender differences in children's scores, consistent with other work in this area (Kochanska et al., 2001; McClelland et al., 2000; NICHD ECCRN, 2003b). In our study, girls scored higher compared with boys, although both groups improved at the same rate; to assess when differences emerge, investigations prior to 36 months are still necessary. For example, Fabes et al. (2003b) suggested that girls' versus boys' socialization experiences in preschool may be potential contributors to observed behavioral differences. Moreover, they noted that boys' play is less structured and supervised than that of girls', which may be negatively associated with behavioral regulation. These findings reiterate the value of creating environments conducive to helping both boys and girls develop and practice attention, working memory, and inhibitory control skills.

Present results suggested that the Head-to-Toes Task is a reliable and valid measure of early behavioral regulation, but two final limitations must be mentioned. First, even though we found variability in performance for all age groups, a direct measure appropriate for a broader age range is still needed. Second, the missing data dummy variable deflected scores negatively, although not substantially, so children who were missing background data scored significantly lower, compared with those for whom data were complete. Aspects of families who did not return background questionnaires, which could be associated with lower behavioral task performance, could include a lack of organization or free time to complete questionnaires. Because families volunteered to participate, missing background data may have been more similar to the families in the population who chose *not* to enroll in the study. This may also mean growth curve estimates may be upwardly biased because participants may have been children from more organized homes, with higher levels of behavioral regulation. Unfortunately, the nature of score differences for families missing data remains an unknown source of variation. This reveals another advantage of using a direct measure, however, because parents who did not return the background questionnaire would also be less likely to return any other survey. By using direct observation, we were able to assess behavioral regulation for these participants.

7. Summary

This study indicates the Head-to-Toes Task is reliable and valid, easy to administer, and provides consistent longitudinal information about the development of children's behavioral regulation. The task elicits comparable results when administered in school settings in two geographical and demographic locations in the United States (Michigan and Oregon), since no differences in task performance were found in these two sites after considering key background variables. Further, the effects of minority group membership diminished after controlling for parent education. The Head-to-Toes Task is a valuable direct behavioral assessment of inhibitory control, attention, and working memory, which can be used in concert with other measures to help ensure children have the skills they need to successfully transition to formal school settings.

Acknowledgments

The authors thank the members of the University of Michigan Pathways to Literacy project, the OSU Kindergarten Readiness Study, and the participating families, teachers, and school district administration and personnel at both sites who made this study possible. The Pathways to Literacy project was supported by the National Institute of Child and Human Development and the National Science Foundation under grant numbers R01 HD27176 and 0111754, respectively. Additional funding was provided by the US Department of Education, Institute for Education Sciences under grant number R305H04013. The OSU Kindergarten Readiness Study was supported by Oregon State University.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012–1028.

Adams, A.-M., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. *International Journal of Psychology*, *34*(5), 364–373.

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling & Development*, *70*(5), 610–615.

Arizmendi, T., Paulsen, K., & Domino, G. (1981). The Matching Familiar Figures Test: A primary, secondary, and tertiary evaluation. *Journal of Clinical Psychology*, *37*(4), 812–818.

Barkley, R. A. (1994). The assessment of attention in children. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 69–102). Baltimore, MD: Brookes.

Barkley, R. A. (2004). Attention-deficit/hyperactivity disorder and self-regulation: Taking an evolutionary perspective on executive functioning. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 301–323). New York: Guilford.

Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications*. New York: Guilford.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*(2), 111–127.

Blair, C. (2003). Behavioral inhibition and behavioral activation in young children: Relations with self-regulation and adaptation to preschool in children attending Head Start. *Developmental Psychobiology*, *42*(3), 301–311.

Bredekamp, S., & Copple, C. (1997). *Developmentally appropriate practice in early childhood programs*. Washington, DC: National Association for the Education of Young Children.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*(2), 571–593.

Bronson, M. B. (2000). *Self-regulation in early childhood: Nature and nurture*. New York: Guilford.

Bronson, M. B., Tivnan, T., & Seppanen, P. S. (1995). Relations between teacher and classroom activity variables and the classroom behaviors of prekindergarten children in Chapter 1 funded programs. *Journal of Applied Developmental Psychology*, *16*(2), 253–282.

Buela-Casal, G., Carretero-Dios, H., De los Santos-Roig, M., & Bermúdez, M. P. (2003). Psychometric properties of a Spanish adaptation of the Matching Familiar Figures Test (MFFT-20). *European Journal of Psychological Assessment*, *19*(2), 151–159.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73–92.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319.

Connell, C. M., & Prinz, R. J. (2002). The impact of childcare and parent-child interactions on school readiness and social skills development for low-income African American children. *Journal of School Psychology*, *40*(2), 177–193.

Connor, C. M., Huddleston, C., Travis, Q. M., Phillips, B., Underwood, P., Cameron, C. E., et al. (2007). *Children's self-regulated learning in first grade classrooms*, submitted for publication.

Deater-Deckard, K., Petrill, S. A., Thompson, L. A., & DeThorne, L. S. (2006). A longitudinal behavioral genetic analysis of task persistence. *Developmental Science*, *9*(5), 498–504.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352–362.

Díaz, R. M., Neal, C. J., & Amaya-Williams, M. (1990). The social origins of self-regulation. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 127–154). New York: Cambridge University Press.

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology*, *36*(2), 161–174.

Eisenberg, N., & Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development*, *75*(2), 334–339.

Entwisle, D. R., & Alexander, K. L. (1998). Facilitating the transition to first grade: The nature of transition and research on factors affecting it. *Elementary School Journal*, *98*(4), 351–364.

Epsy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. (2001). New procedures to assess executive functions in preschool children. *Clinical Neuropsychologist*, *15*(1), 46–58.

Fabes, R. A., Hanish, L. D., & Martin, C. L. (2003). Children at play: The role of peers in understanding the effects of child care. *Child Development*, *74*(4), 1039–1043.

Fabes, R. A., Martin, C. L., Hanish, L. D., Anders, M. C., & Madden-Derdich, D. A. (2003). Early school competence: The roles of sex-segregated play and effortful control. *Developmental Psychology*, *39*(5), 848–858.

Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology*, *24*(1), 51–73.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology*, *92*(2), 377–390.

Haney, W. (1980). Units and levels of analysis in large-scale evaluation. *New Directions for Methodology of Social & Behavioral Science*, *6*, 1–15.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*(2), 617–644.

Howse, R. B., Calkins, S. D., Anastopoulos, A. D., Keane, S. P., & Shelton, T. L. (2003). Regulatory contributors to children's kindergarten achievement. *Early Education & Development*, *14*(1), 101–119.

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education*, *71*(2), 151–174.

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology*, *34*(6), 1326–1339.

Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool-aged children. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(3), 209–215.

Kalpidou, M. D., Power, T. G., Cherry, K. E., & Gottfried, N. W. (2004). Regulation of emotion and behavior among 3- and 5-year-olds. *Journal of General Psychology*, *131*(2), 159–178.

Karnes, M. B., Johnson, L. J., & Beauchamp, K. D. F. (2005). Reprise: Developing problem-solving skills to enhance task persistence of handicapped preschool children. *Journal of Early Intervention*, *27*(4), 236–246.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development*, *72*(4), 1091–1111.

Kochanska, G., Murray, K., Jacques, T. Y., & Koenig, A. L. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The differential assessment of children's attention: The Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry*, *42*(8), 1065–1081.

McCabe, L. A., Cunnington, M., & Brooks-Gunn, J. (2004). The development of self-regulation in young children: Individual characteristics and environmental contexts. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 340–356). New York: Guilford.

McCabe, L. A., Hernandez, M., Lara, S. L., & Brooks-Gunn, J. (2000). Assessing preschoolers' self-regulation in homes and classrooms: Lessons from the field. *Behavioral Disorders*, *26*(1), 53–69.

McCabe, L. A., Rebello-Britto, P., Hernandez, M., & Brooks-Gunn, J. (2004). Games children play: Observing young children's self-regulation across laboratory, home, and school settings. In R. DelCarmen-Wiggins & A. Carter (Eds.), *Handbook of infant, toddler, and preschool mental health assessment* (pp. 491–521). New York: Oxford University Press.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*, 947–959.

McClelland, M. M., Cameron, C. E., Wanless, S., & Murray, A. (2007). Executive function, behavioral self-regulation, and social-emotional competence: Links to school readiness. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on social learning in early childhood education.* (pp. 113–137). Charlotte, NC: Information Age.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly*, *18*(2), 206–224.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly*, *15*(3), 307–329.

Meisels, S. J. (2006). *Accountability in early childhood: No easy answers* (Occasional Paper No. 6). Chicago, IL: Herr Research Center for Children & Social Policy, Erikson Institute. Available from http://www.erikson.edu/research.asp?file=publications_topic#assess

Miller, A. L., Gouley, K. K., Shields, A., Dickstein, S., Seifer, R., Magee, K. D., et al. (2003). Brief functional screening for transition difficulties prior to enrollment predicts socio-emotional competence and school adjustment in Head Start preschoolers. *Early Child Development and Care*, *173*(6), 681–698.

Miller, K. F., Kelly, M., Zhou, X., & Campbell, J. I. D. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development. In *In Handbook of mathematical cognition*. New York: Psychology Press., pp. 163–177.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938.

Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the Dimensional Change Card Sort Task. *Child Development*, *77*(2), 395–412.

NICHD ECCRN. (2003a). Families matter—even for kids in child care. *Journal of Developmental & Behavioral Pediatrics*, *24*(1), 58–62.

NICHD ECCRN. (2003b). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development*, *74*(6), 1639–1662.

NICHD ECCRN. (2004). Type of child care and children's development at 54 months. *Early Childhood Research Quarterly*, *19*(2), 203–230.

Pianta, R. C., & McCoy, S. J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology*, *18*(1), 1–22.

Pickering, S. J., & Gathercole, S. E. (2004). Distinctive working memory profiles in children with special educational needs. *Educational Psychology*, *24*(3), 393–408.

Pintrich, P. R. (2000). Issues in self-regulation theory and research. *Journal of Mind & Behavior*, *21*(1), 213–219.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *Elementary School Journal*, *106*(1), 21–38.

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly*, *30*(2), 131–147.

Renninger, K. A., & Wozniak, R. H. (1985). Effect of interest on attentional shift, recognition, and recall in young children. *Developmental Psychology*, *21*(4), 624–632.

Rimm-Kaufman, S. E., La Paro, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in the kindergarten classroom. *Elementary School Journal*, *105*(4), 377–394.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, *15*(2), 147–166.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., et al. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, *28*(2), 573–594.

Schultz, D., Izard, C. E., Ackerman, B. P., & Youngstrom, E. A. (2001). Emotion knowledge in economically disadvantaged children: Self-regulatory antecedents and relations to social difficulties and withdrawal. *Development & Psychopathology*, *13*(1), 53–67.

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC, US: National Academy Press.

Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a 'button-press' go/no-go task. *Journal of Experimental Child Psychology*, *94*(1), 18–26.

Thorell, L. B., & Wahlstedt, C. (2006). Executive functioning deficits in relation to symptoms of ADHD and/or ODD in preschool children. *Infant and Child Development*, *15*, 503–518.

Wachs, T. D., Gurkas, P., & Kontos, S. (2004). Predictors of preschool children's compliance behavior in early childhood classroom settings. *Journal of Applied Developmental Psychology*, *25*(4), 439–457.

Ward, H., Shum, D., McKinlay, L., Baker-Tweney, S., & Wallace, G. (2005). Development of prospective memory: Tasks based on the prefrontal-lobe model. *Child Neuropsychology*, *11*(6), 527–549.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131–149.

Whitehurst, G., & Lonigan, C. (2001). Emergent literacy: Development from prereaders to readers. In S. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11–29). New York: Guilford.

Zelazo, P. D., Müller, U., & Goswami, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 445–469). Malden, MA: Blackwell Publishing.