BRIDGING NATURALISTIC AND LABORATORY MEASURES OF SELF-REGULATION: THE DEVELOPMENT AND VALIDATION OF CHALLENGE TASKS

by

Xuezhao Lan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education and Psychology)
in The University of Michigan
2009

Doctoral Committee:

Professor Frederick J. Morrison, Chair
Professor Kevin F. Miller
Associate Professor William J. Gehring
Associate Professor Priti R. Shah

This dissertation is dedicated to my father for his profound influence on me in the 27 years of my life, and to my grandmother, who raised me and whom I deeply loved.

ACKNOWLEDGEMENTS

Thanks goes to my dissertation chair Frederick J. Morrison in guiding me through the entire process; and to Priti Shah, Kevin Miller and Bill Gehring for their generous support and comments. This dissertation work would not have been possible without the support of my "Pathways to Literacy" lab mates, my research assistants Jean Kwek and others, and the teachers and principals who generously allowed me into their classrooms.

In addition, I want to thank Kevin, Priti and Fred again for working with me on various projects in the past four and half years and advised me not only academically but also in many aspects above and beyond. Thanks must go to Yibo Ling, Duohong Xu and Tianjun Ling for their love and support from the first day I arrived at Michigan.

Lastly, I would like to acknowledge my mother for always supporting me.

Table of Contents

List of Tables

List of Figures

List of Appendices

ABSTRACT

BRIDGING NATURALISTIC AND LABORATORY MEASURES OF SELF-REGULATION: THE DEVELOPMENT AND VALIDATION OF CHALLENGE TASKS

by

Xuezhao Lan

Chair: Frederick J. Morrison

Self-regulation (or "executive function", including inhibition, working memory and attention) has drawn a great deal of attention recently as a result of its role in predicting academic achievement and social competence. However, age appropriate and ecologically valid measurements of the concept are seriously lacking. Most existing assessments measure the concept either exclusively in the laboratory setting or directly in the classroom through observation of natural behaviors. In contrast, the current study introduces a series of field-experimental measures (called, the "challenge tasks") to evoke self-regulation behaviors in the classroom. To validate these measures, 139 demographically diverse children were recruited from the suburban Midwestern area. Results showed that the challenge tasks, 1) have good inter-coder and test-retest reliability, 2) relate to existing experimental tasks as demonstrated by correlation analysis and confirmatory factor analysis, and 3) predict math and reading achievement. Further, the inhibition and working memory challenge tasks added additional variance over and above in predicting both reading and math achievement. The implications, limitations and future direction of this work are discussed.

Chapter 1

Introduction

Self-regulation (including related concepts such as behavioral self-regulation and executive control) in early childhood has drawn a great deal of attention in recent years. It is important for school readiness, academic achievement (e.g. McClelland, Cameron, Connor, Farris, Jewkes & Morrison, 2007) and social competence (e.g. Hughes, 1998). Over the last 10-15 years, while efforts have been made to create developmentally appropriate laboratory assessments of self-regulation (e.g. Carlson, 2005), measurements that are ecologically valid and robust for diagnosis and assessment by non-researchers are seriously lacking (e.g. Blair, 2002) proposed to develop tools that predict school readiness). The goal of this study was to develop and validate easy-to-use field-experimental measures (named the "Challenge tasks") of self-regulation.

A large variety of direct self-regulation measures have been used in two mutually exclusive contexts: in the lab (i.e. performance based tasks including computerized tasks and behavioral experimental tasks) and in the naturalistic settings such as the classrooms. Indirect self-regulation measures such as teacher report have also been used frequently. There are both advantages and drawbacks to the currently used self-regulation measurements: Laboratory measures tend to be precise but are narrow and they may not accurately capture how children behave in the real world (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007); naturalistic observations (classroom coding) usually incorporate various social factors (e.g. teachers, peers and classroom structure) but tend to be less precise; furthermore, researchers have no control over targeted behaviors. Separating relatively stable factors such as students' self-regulation skills from variant features such as situational peer and teacher influence complicates this strategy (Seidel & Prenzel, 2006). Furthermore, comparing observations or scores across studies is difficult since these observation tend to be embedded in varying features such as a particular day's observation; Teacher reports add to the ecological validity of measuring self-regulation because they are based on many days of observation of child behavior (as oppose to single data points measured by other tasks); however, teacher ratings are often more subjective and subject to biases. For example, Derks, Hudziak, & Boomsma (2007) found that teachers scored boys with more ADHD related symptoms than girls while parent reports did not show a difference.

In addition, studies have found weak correlations as well as inconsistent results using these measures. In one study, teacher ratings of inhibition did not correlate with performance-based inhibition measurements from computerized tasks (Bodnar et al., 2007); in another study investigating gender differences in self-regulation, teacher reports revealed strong gender differences in self-regulation in China, Korea and Taiwan but a direct measure of self-regulation revealed no gender difference in any of these countries (Lan & Morrison, 2008).

Therefore, it is quite possible that each of these assessments measure a unique aspect of self-regulation embedded in one particular context. Thus it is essential 1) to examine to what extent these measurements share common variance, and 2) to develop more ecologically valid measurements that bridge the gap between the laboratory and naturalistic assessments which can be used by researchers and childcare practitioners alike.

The aims of this dissertation were, 1) to develop a series of classroom-based challenge tasks that evoke self-regulation behaviors in the naturalistic settings (i.e. preschool classrooms). In contrast to traditional classroom measures over which researchers have no control, these ‘structured’ measures stemmed from experimental measures that evoke targeted behaviors (e.g. inhibition, working memory, attention control, 2) to investigate relationships between the challenge tasks, existing experimental tasks, and teacher report, and 3)to investigate the efficacy of field-experimental tasks in predicting academic achievement.

<u>Definition of self-regulation</u>

Self-regulation refers to a set of cognitive skills used for purposeful, future-oriented behavior, including inhibition of inappropriate responses, processing multiple pieces of information simultaneously in working memory, controlling attention (a.k.a. interference resolution or the ability to avoid being distracted by distractors) and organizing and planning behaviors (Blair, 2002; Eslinger, 1996; Klein, 2003; Shonkoff & Phillips, 2000; Zelazo, Muller, Frye, & Marcovitch, 2003). There are two approaches to studying self-regulation (Blair & Razza, 2007). One is a temperament-based approach where motivation and emotion are emphasized, while the other is an executive function-based approach which studies volitional control. This dissertation adopts the executive function approach and will use the two terms (executive function and self-regulation) interchangeably.

Research suggested that self-regulation contains multiple components, including response inhibition (inhibiting inappropriate responses), working memory and attention control (resistance to distractors) (Barkley, 1997; Muller, Zelazo, Hood, Leone, & Rohrer, 2004). It has been suggested that cognitive flexibility (flexibly switching between rules) is also a critical component of self-regulation (Zelazo, Frye, & Rapus, 1996). Though it is generally accepted that the components of self-regulation are loosely correlated constructs in adults (Miyake et al., 2000), fewer attempts have been made to examine the association in children (especially under six years of age). Some studies have suggested that self-regulation contains only one general component (Wiebe, Espy & Charak, 2008), while other studies rebut this notion (Klenberg, Korkman, & Lahti-Nuuttila, 2001; Welsh, Pennington, & Groisser, 1991).

The current study adopts a multi-dimensional approach using tasks that measure multiple components of self-regulation; in contrast to existing studies which only use experimental tasks, the present study also uses measures from multiple contexts (i.e. computer based, behavioral and field-experimental tasks). In addition to developing and validating field-experimental tasks, the results of the current study also contribute to the current debate over the structure of self-regulation.

<u>Self-regulation and academic achievement</u>

Self-regulation has been found to be important for both reading (McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, in press) and math achievement (Bull & Scerif, 2001). For instance, Bull & Scerif (2001) found that students with better self-regulation skills including inhibition, working memory and task switching scored higher in math. Inhibition was also found to be a unique predictor of math achievement (Espy, 2004, Blair & Razza, 2005) and children's literacy in kindergarten beyond general intelligence and attention control. Strong working memory skills are consistently associated with higher academic achievement (Adams, Bourke, & Willis, 1999; Gathercole & Pickering, 2000). Attentional control was not found to uniquely predict math or reading ability in Blair & Razza (2005).

The aforementioned studies used experimental self-regulation measures that may be less than ideal predictors of children's actual achievement because young children are often unable to comprehend or concentrate on unfamiliar tasks especially those with heavy verbal demands. The current study expands the horizon of experimental measurements and develops field-experimental measures that measure the same cognitive constructs while accounting for situational variation in natural settings. In the following sections, I will review traditional measures of self-regulation in laboratory and naturalistic settings and then describe the field-experimental measures of self-regulation developed for this work.

Measuring self-regulation in the laboratory setting

In the laboratory setting, scientists have employed two types of performance based self-regulation measurements, in part stemming from the nature and demands of different fields; Neuroscientists and cognitive psychologists have computerized tasks that permit the record accuracy rate and reaction time in milliseconds (e.g. the stop signal task which asks children to respond to certain stimuli but stop responding when hear other stimuli, Aron et al., 2003). Such measurements often involve minimal social context with only simple behavioral reactions (e.g. button press). Another measure of self-regulation created by developmental and educational scientists is less rigorously time constrained (e.g. a gift wrapping game measuring delay of gratification, in which children had to resist peeking at the gifts, Kochanska et al, 1996). Such measures typically require more complicated behavioral reactions (e.g. peeking the gift) and a longer time to react (several seconds) but still involve minimal social context. Children are usually assessed individually in the laboratory or school setting by a trained experimenter. In the current study, we refer to the first type as computerized experimental measurement and the second type as behavioral experimental measurement.

Though most experimental assessments include all three components of self-regulation, individual tasks are designed to primarily assess a single component. For instance, while completing the shape Stroop task (Kochanska et al., 2000) requires working memory (remembering instructions) and attention control (paying attention to the task instead of other distractions), response inhibition is the primary focus of this test since the primary task is to point to the small shape when seeing the big shape and vice versa. Various experimental tasks have been developed to measure each component of self-regulation for preschool and kindergarten children (e.g. Carlson, 2005). Tasks used to measure response inhibition include variation of the Simon says task (e.g. Head to Toes, Knees to Shoulders, Cameron, McClelland, Mathews, & Morrison, 2007) where children have to do the opposite of what experimenters says) and the Stroop like task such as the shape stroop (Kochanska et al., 2000) and grass/snow task (Carlson and Moses, 2001); Go/no-go task (Mesulan, 1985) and Stop signal task and variations, where children have to respond when signaled and stop responding when signaled otherwise;

Working memory tasks usually require children to process two sets of information at the same time remembering one of them, such as backward digit span where children have to remember the digits and repeat them backwards, and Sentence Completion (Towse, Hitch & Hutton, 2002), where children have to complete sentences while remembering the last words of each sentence;

Typical attention control task include Flanker (Mezzzacappa, 2004) where children are asked to pay attention to the center fish and ignore the side fishes and pair cancelation (Woodcock & Mather, 2000) where children have to circle pairs of stimuli and ignore the distractors. Distinctions between these components is clear for experimental tasks but are rarely made noticeable for naturalistic measurements such as natural observations and teacher report.

Despite the advantages of having precise indicators (e.g. reaction time or accuracy rate) and well-defined tasks for measuring each component of self-regulation, laboratory based tasks have been criticized for being de-contextualized as they incorporate neither larger nor immediate environments, which were both shown to influence children's performance (McCabe & Brooks-Gunn, 2007).

Experimental tasks typically do not capture the variance explained by peers, teachers or interactions between various non-cognitive factors and correlate weakly with assessments that incorporate such variances. For example, the correlation between laboratory measurements of self-regulation and teacher reports was reported to be very low (Lan & Morrison, 2008). One possible explanation of such inconsistency is that laboratory tasks and teacher reports measure distinct variances; another possible explanation is that the laboratory task is a one-time measure which does not reliably capture children's true performance as observed by teachers on a daily basis. A third possibility is that teacher reports are subjective and may or may not reflect children's actual behavior.

Measuring self-regulation in the naturalistic setting

Fewer studies have measured self-regulation in more naturalistic settings. These investigations include classroom video observation, adapted experimental tasks and indirect measures such as teacher and parent report.

Classroom observations are based on children's real time performance in the classroom; however, researchers can only passively observe what is happening in the classroom and have no control over triggering behaviors of interest. Components of self-regulation were also often not differentiated using such measures. One study investigating children's behavioral self-regulation skills in the classroom setting (Lan et al., 2009) found Chinese students to have greater behavioral regulation (% time on-task throughout a math class) than their American counterparts. However, one can hardly tell whether students are "regulating themselves" to be on task or simply complies with teacher expectations. In other words, in a well-managed classroom, self-regulation behaviors may not be apparent.

A third measure that is commonly used and regarded as ecologically more valid than performance-based tasks is teacher rating of self-regulation (such as The Child Behavior Rating Scale, Bronson Tivnan, & Seppannen, 1995). However, teacher measures can be subjective and biased (e.g. Derks, Hudziak, & Boomsma, 2007). Some teacher reports show inconsistent results with performance-based measures (e.g. Bodnar et al., 2007). Furthermore, existing teacher reports do not typically differentiate components of self-regulation.

To increase the ecological validity of experimental measures, some have adapted experimental tasks into forms that can be used in small groups in the classroom. Such measures are less decontextualized than experimental tasks but still might not create similar-to-reality context that maximize children's real world performance. Mccabe, and Brooks-Gunn (2007) adapted the gift-wrapping task for groups of children in the classroom. They found that children had more difficulty inhibiting themselves in the group setting than when they were alone. It is unclear if the decline in performance resulted from a lack of familiarity with the tasks, which increased the difficulty in understanding the rules in group settings.

In summary, on one hand, experimental measures are decontextualized and so may not capture non-cognitive variances; on the other hand, naturalistic measures either only capture natural behaviors so that researchers do not have control over the experiment, or are subjective and potentially biased. The following methods proposed is by no means a perfect solution to the problem, but is a first attempt to tackle this difficult dilemma.

Developing field-experimental tasks

One possible way of bridging the laboratory and naturalistic measurements is to develop experimental tasks in natural settings. The purpose of these field-experimental tasks is to evoke self-regulation behaviors in the classroom and introduce some degree of experimental control. Such tasks should be familiar to children so that the instructions are easily understood even in group settings. In addition, such tasks should be short and easy to use so that multiple time points can be easily collected. In the present study, we seek to develop and investigate how these challenge tasks relate to traditional self-regulation measures and academic achievement.

Specifically, I highlight the following advantages: challenge tasks 1) serve as a bridge between experimental measurements and naturalistic measurements: on one hand, they capture variances explained by contextual factors and interactions not captured by traditional lab measurements; on the other, they make it easy to measure behaviors in naturalistic settings, 2) can be easily implemented by teachers and educators 3) can assess all the components of self-regulation, providing an accurate and comprehensive picture.

To accurately assess self-regulation, we used a multidimensional approach with separate challenge tasks to measure inhibition, working memory and attention control. Correlating children's performance from these components with well-defined inhibition, working memory and attention control measures from the laboratory (e.g. measures of inhibition with the challenge inhibition task) would allow us to evaluate how the challenge tasks relate to basic self-regulation constructs as assessed in the laboratory setting. Further, we designed teacher questionnaires that target the three components of self-regulation. Correlating challenge task performance on each self-regulation component with teacher rated inhibition, working memory and attention control would reveal how the two assessment strategies relate to each other.

In collaboration with teachers and principals and based on common games children often play, we developed a series of three interrelated tasks that evoke children's self-regulation skills in classrooms. During these tasks, a teacher and students stand in a circle in the classroom and march to recorded music. This is a common activity in a preschool and permit introduction of experimental manipulation. For example, the inhibition challenge task requires children to march to the music and freeze into a statue when the music stops. By varying task demands, children could be required to exhibit their degree of inhibition in an everyday preschool environment. The working memory task requires children to remember instructions (e.g. jump three times) while marching and act when the music stops. The instruction become gradually more complicated as the game goes on (e.g. jump three times and clap twice; jump three times, clap twice and go three step backwards). During the attention control task, children sit in small groups and each of them work on a tangram puzzle. They are instructed to stay on task instead of being distracted by distractiors (such as cell phone rings). All tasks are described in detail in the method section. It takes on average eight to ten minutes to complete.

Tasks were piloted in two classes (a head start class and a tuition-based class) and revised before used in all classrooms.

Research questions and hypothesis

The broad aim is to design and validate a series of field-experimental tasks. Specifically, the study explores the following three research questions:

1. What is the nature, variability and reliability of the challenge tasks, as well as other self-regulation tasks used in this study?

All the tasks are expected to be normally distributed and show variability as well as reliability. Each class was tested multiple times on both inhibition and working memory challenge tasks and consistency between these time points was assessed.

2. What are relations between the challenge tasks and laboratory tasks and teacher report?

The challenge tasks are expected to correlate with both experimental tasks and teacher report.

Additionally, results from Confirmatory Factor Analysis (CFA) will be reported. There are several possible hypotheses:

a. *Single versus multiple components of self-regulation.* According to previous literature (Miyake et al., 2000), if the three factors model fits the data the best with each factor tapping one component of self-regulation, namely inhibition, working memory and attention control (or cognitive flexibility), it would mean that the components of self-regulation are loosely correlated constructs that are independent of measurements/contexts. Given the finding of Wiebe and colleague

(2008) in exploring components of self-regulation in children, if one-factor model fits the data, it would reveal that components of self-regulation are essentially one correlated construct.

c. *Single versus multiple contexts*. However, if two or three factors loaded on different measurements (behavioral self-regulation measures, field-experimental measures and/or computer based tasks), it would mean that the contexts in which we measure self-regulation contribute the most to the differences in variances, independent of components.

d. *Other alternative models*. A few other alternative models are also possible and will be tested, factors grouped by verbal (tasks that involve processing or remembering a verbal rule) versus non-verbal tasks (tasks that involve button press); visual heavy tasks versus tasks that do not require visual processing; response time based tasks (tasks that use reaction time as the main dependent variable) versus accuracy based tasks (tasks that use accuracy as the main dependent variable).

3. What are the unique and joint (with experimental tasks) contributions of challenge tasks to academic achievement?

If experimental tasks measure children's basic cognitive skills and challenge tasks measure how these skills are used in the naturalistic environment, challenge tasks alone should be predictors of math and reading achievement. In addition, if the non-shared variance (variances other than those underling children's basic cognitive skills) between the two kinds of tasks alone matters to achievement, challenge tasks should still predict achievement controlling for experimental tasks (i.e. unique contributions).

Factor analysis and Analytical Plan

Previous research investigating the relationships between self-regulation measures often uses exploratory factor analyses (EFA) or principal component analysis (PCA), which cannot examine specific model patterns thus leaving limited space for explanation (Wiebe et al., 2008). Further, many studies used Varimax rotation assuming orthogonality of the factors, which could provide biased results (Gorsuch, 1997), given that the components of self-regulation are often correlated (Friedman & Miyake, 2006).

In contrast, confirmatory Factor Analysis (CFA) not only allows one to specify model patterns but also model constraints (such as coefficient and correlations) in order to test model fit. Some researchers (e.g. Miyake et al., 2000) argue that CFA extracts more purified latent factors because one can build different model patterns based on what is known about the task demands. It also provides multiple indicators of model fit allowing explicit comparison across different models.

Unfortunately, very few studies have used CFA to investigate the relationships between self-regulation measures in children. Using experimental tasks, Wiebe and colleagues (2008) demonstrated that all three measures of self-regulation were essentially one component. The current study was the first to use CFA in child population in investigating measures across various contexts.

In addition, missing data were handled carefully in the current study. M*plus* 5.2 (Muthén and Muthén, 1998-2007) was used to calculate the correlations and estimate CFA parameters adjusting for missing data. Contrary to common solutions such as listwise deletion which often produces biased estimators and weaker power for samples that are not sufficiently large (Acock, 2005), M*plus* uses all the available information (i.e. the optimal Full Information Maximum Likelihood (FIML) approach) and provides a maximum likelihood estimation to estimate correlations and parameters in CFA. This Maximum likelihood approach is one of the two recommended modern approaches to deal with missing data in psychological research (e.g. Acock, 2005)

Chapter 2

Method

Participants:

Teachers and children were recruited from two preschools in a rural and suburban county in the Midwest. Parent consent forms were distributed and collected by teachers. All parents received a description of the study and a consent form, 90% the parents consented to participate in the study. The final sample included 139 children from two Head Start classrooms, six state-funded school readiness programs, and three tuition-based preschool classrooms in two schools. 91% of the teacher questionnaires and 60% of the parent background questionnaire were returned for analysis. Parent reports indicated that these children came from families with a mean family income of $41,400 for a family of 4.25 (range: $4160-$150,000). The majority of the children in the study were Caucasian (69.2%); the rest were Chaldean (10.8%), Asian Americans (9.2%), African-Americans (6.2%) and Latino/a (4.6%). Preschoolers had a mean age of 57.4 months (SD =6.5), and the sample was 43% female.

Procedure:

Undergraduate research assistants were trained to administer the tasks. They went through two group training sessions and practiced giving the battery of tasks five times to each other as well as to persons unfamiliar with the study. They were tested and judged on performance by the project director before being certified to administer the tests. During testing, workstations were set up with two chairs facing each other around a small table. During the computer sessions, experimenters and children sat on one side of a table. All the tests were administered either in a quiet room or a corner in the hallway away from the classrooms. The order and version of the tasks were counter-balanced. The battery of tasks was split into four sessions each including 2-3 tasks that take 10-15 minutes to finish. Any given child was tested on one session on a given day (or maximally two if the first session was too short). The sessions were further broken down into those with only 1-2 task(s) each as needed. Testing lasted from April 15th to May 22nd of 2008. All experimenters were working full time either for credits or payment. Video challenge tasks were also videotaped during this period time (see the challenge task section for detailed description). Teachers and parents completed background questionnaires at the end of May.

Tasks:

Assessments included tasks from both laboratory and naturalistic settings tapping each of the three components of self-regulation (inhibition, working memory and attention control). A cognitive flexibility task was also included in the battery. The laboratory tasks were chosen because they 1) they fit the definition of the component, for example, for working memory tasks, we chose tasks that allow children to process two kinds of information at the same time over those tap primarily short-term memory (e.g. digit span); 2) have used by previous studies and shown to have good reliability and validity; 3) appropriate for children who are four to five years old with a wide range of abilities.

Table 1 provides an overview of all the self-regulation tasks.

Table 1. Overview of all self-regulation tasks.

Component and context	Inhibitory control	Working memory	Attention control	Other (cognitive flexibility)
Laboratory tasks	Stop signal	Operation span	Flanker	Card sorting (3DCCS)
	HTKS	Sentence completion	Pair cancellation	
Challenge tasks	Challenge task Freeze game	Challenge task Jumping game	Challenge task-resistance to distractor	
Teacher report	TQ-inhibition	TQ-working memory	TQ-attention control	

Note. HTKS=Head to Toes, Knees to Shoulders. TQ=Teacher Questionnaire. 3DCCS = three Dimension-Change Card Sorting.

The self-regulation battery: The self-regulation battery comprised seven tasks including two computer tasks. The tasks were:

*Staircase Stop Signal (*Adapted from Aron et al., 2003): Children were asked to respond to stimuli on a computer screen and stop responding upon hearing a beeping sound. It measures children's ability to withhold their responses. The task included two blocks and two types of trials (the go trial and stop trial). In the "nonstop" block, children were asked to press the red bottom (button "1") when they saw a red heart on the screen and the green bottom (button "4") when saw a green tree. Stimuli appear for 1500 ms after a 700-fixation point. In the "stop" block, they were instructed to respond similarly to stimuli unless they hear a beeping sound (the stop trial), at which time they had to stop responding. A stop trial appeared after a go trial. The time interval between the go and stop signals (SSD) indicates children's ability to withhold a response. The longer the SSD, the more difficult it was for children to stop. In contrast to the traditional stop-signal paradigm where SSD was fixed, the stop-signal paradigm used here was a staircase stop-signal where SSD decreases (to make it easier for the subject to stop at the stop signal) if the subject fails at a previous stop trial and increases if the subject succeeds in order for a child to achieve a success rate of 50% for stop trials (Levitt 1970). A ''critical'' SSD (called the "step" in the current study) is computed as a major dependent measure that represents the time delay required for the subject to succeed in withholding a response in the stop trials 50% of the time. Another dependent measure was the accuracy rate for the go trials in the non-stop block.

There were 6 trials for each practice blocks (2 in total) followed by 40 trials for each testing block. In the stop block, 50% of the trials were stop trials. There were one non-stop and one stop block.

Fish Flanker (Mezzacappa, 2004; Rueda et al., 2004): a center fish flanked by two fish on either side appeared on the computer screen for 4000 ms or until a button press. Children were asked to press "right" (if the center fish faced right) or "left" (if the center fish faced left) button of the mouse to "catch" the fish. The fish (targets) appeared for 400ms after a cue. In a control trial, the center fish faced the same direction as the flanked fish. In an interference trial, the center fish faced the opposite direction as the other fish. A feedback appeared 500ms after button press. Children were given "hooray!" (for positive responses) or "beep" (for negative responses) feedback during both practice and formal testing. One testing block appeared (40 trials) with half control and half interference trials after a practice block (6 trials). Both reaction time and accuracy were recorded. Figure 1 showed an example of an incongruent trial.

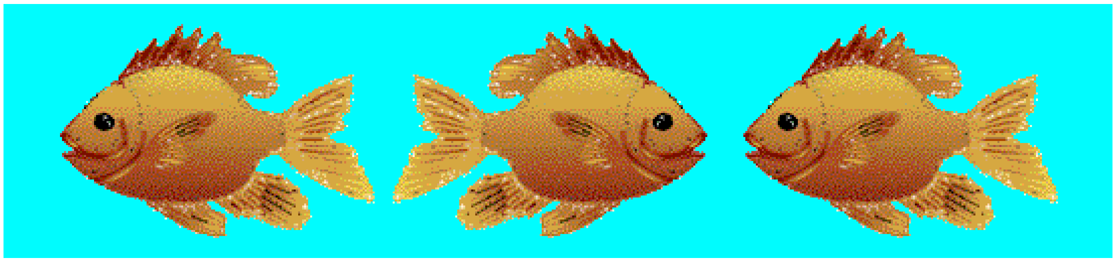

Figure 1. Sample incongruent trial in the flanker task

Operation Span (Blair & Razza., 2005). In this task, children had to remember the animals living in each house while processing the color information of each house. Children were shown a booklet with one-, two-, three- or four-item pictures of colored houses and animals in the houses. The test started from a one-item picture where children were asked to identify the color of a house on the front page (e.g. "what color is this house?" answer: "red") and remember the animal in the house (e.g. "what animal lives in this house? Answer: "a dog"). The experimenter then flipped the page with only a colored house at the back. Children had to recall the animal that lived in the house (e.g. "what animal lived in this house?"). Children must remember the animals while processing the color information, and then recall only the animals living in the house. The task became harder as the number of houses increases. Both accuracy and order of recall were recorded. Figure 2 is an example of a 2-item set.

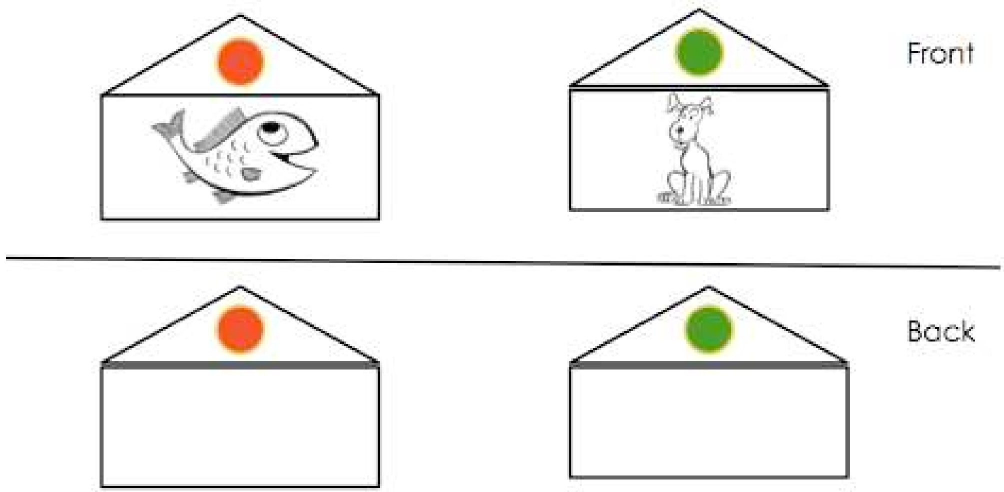

Figure 2. A sample 2-item operation span trial

Head-to-Toes, Knees-to-Shoulders (Cameron et al., 2008; Connor et al., 2007) the task mainly measures inhibitory control (Diamond, 2002). Children were asked to play a game in which they were instructed to do the opposite of what the experimenter told them to do. For example, when asked to touch their head, children had to touch their toes or vice versa. The task included three parts: touching heads and toes, touching shoulders and knees, and the mix of the two. Correct responses on all items received a score of 2, self-corrects (discernible motion to incorrect, with final response given correct) received a score of 1, and incorrect responses received 0 points. The possible highest score was 40. There were a total of 20 items.

Pair Cancellation: This task was from the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock & Mather, 2000). Children were presented with a piece of paper with dogs, balls, and cups on it. They were asked to circle all the ball-dog pairs in which the dog was presented after the ball. They were given three minutes to complete the task. There were 69 correct answers in total. The total number of correct pairs was recorded. Figure 3 showed a sample sheet with correct items marked.

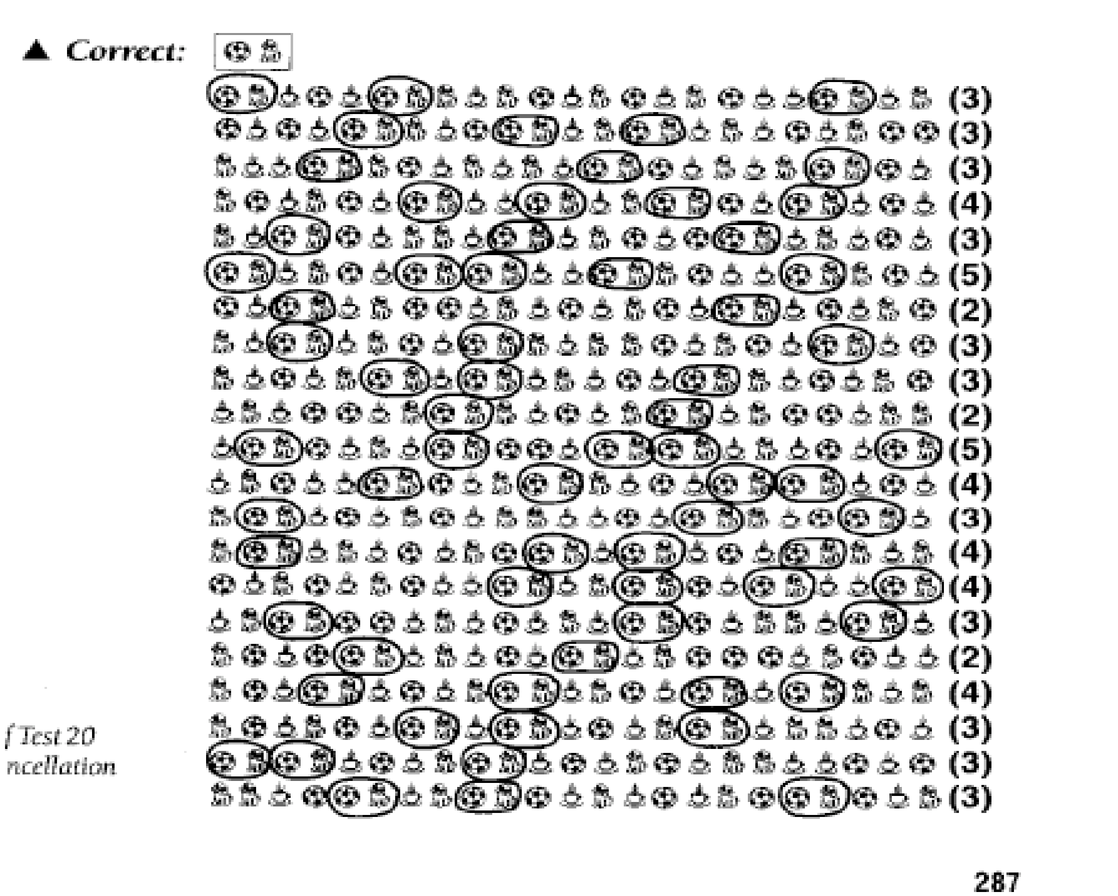

Figure 3. Pair cancelation sample sheet with correct answers

Sentence Completion: Sentence completion task was adapted from the reading span task in Towse et al., (2002) in which children had to complete sentences and remember the words they used to complete the sentences. In this task, children heard a set of short sentences one at a time. Each of these sentence has a missing final word (e..g, "Twinkle, twinkle little ____”). Children then had to complete the sentences and recall all the last words one by one. Children started with one-sentence sets and stopped when they fail to repeat all the words in one set correctly. The maximum sentences in a given set was 4, each of the final word was worth a score of 1. Many sentences used by Towse et al. were included but because the study was based in the US rather than the UK, we excluded some sentences and generated new ones for this study (e.g. twinkle twinkle little star). The total number of correct recall and the order of recall were recorded.

Three-Dimension Card Sorting (3DCCS) (Deák & Narasimham, 2003): children were presented with a set of cards and asked to sort by different dimensions. An experimenter placed three different cards each with different dimensions (e.g. a small yellow bird, a middle blue dog and a large red fish) then asks children to sort by color. Children were then asked to sort by shape and then by size. There was one practice trial before each switch. One distracting card (a big green frog) was used.

Figure 4 showed a sample set for which a child had to sort the card with the medium red dog to one of the four piles according to a rule.

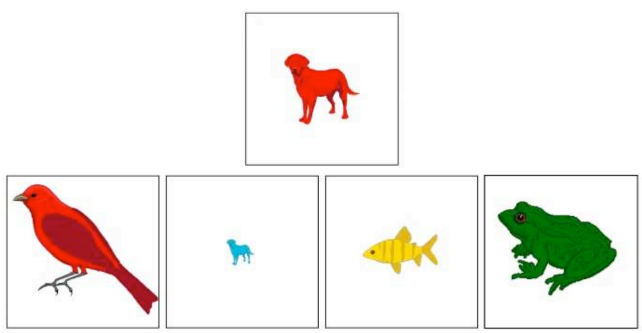

Figure 4. Sample card sorting task.

The questionnaires: The questionnaires included a parent background questionnaire and a teacher rating of self-regulation.

Teacher Questionnaire (appendix): the teacher questionnaire were developed for this dissertation based on existing teacher and parent self-regulation questionnaires, it included inhibition, working memory and attention control items. Sample items in the teacher questionnaire included: “this child leaves seat in classroom in which remaining seated is expected (inhibition)” and “this child follows two-step instructions (working memory)”. Teachers were instructed to rate from 1(strongly disagree) to 7 (strongly agree).

Parent Questionnaire: Parents completed a short questionnaire seeking demographical, economical and health information.

Classroom “Challenge Tasks”:

The challenge tasks were developed collaboratively with teachers and principals. The tasks have three parts. All challenge tasks were videotaped twice (four classes were videotaped three times due to child absence or unexpected interruption during previous recordings) on different days, which were several days apart, and later coded by undergraduate research students.

Challenge - Inhibition (“the freeze game”). Teachers instructed children to march to the music in a circle. When the music stops, children froze into a certain pose. Children could only unfreeze themselves when the teachers said, “unfreeze” or when the music started again. Teachers were the group leaders in this game. An experimenter controlled the marching music and stopped the music at random intervals with around 3 seconds, 5 seconds, 7 seconds, 9 seconds, 11 seconds and 13 seconds in length. There were three fixed trials followed by three varied trials in which teacher decided (or asked a child to decide) which pose children froze into.

The three varied trials were intended to keep the children motivated and the game interesting and were not included in the final analysis.

Challenge – working memory ("the jumping game"). Similar to the inhibition game, children marched to the music in a circle, but prior to marching, teachers instructed students to "jump three times" (one-step instruction) when they heard the music stop. So when children marched in the circle, they had to memorize the instructions thus requiring processing two pieces of information at the same time. After the "one-step instruction", teacher gave "two-step instruction" (jump three times and clap twice) and "three-step instruction" (jump three times, clap twice and go one step backwards) respectively. There were three "varied trials" afterwards (which were also not included in the final analysis) in which a teacher (or a child) decided which two-step instructions they wanted to follow. Figure 4 showed a flow chart of the working memory task in which a teacher gave instructions at the beginning of each trial, followed by student marching, and reaction when music stopped. A new trial started when teacher finished giving instructions.

Figure 5. The working memory challenge task flow chart for fixed trials

Challenge – Attention control ("The puzzle game"). Children were gathered in small groups (groups of 4-6) and were instructed to work on a tangram puzzle on their own (the puzzles were harder than the appropriate level for their age and were different for each child). The experimenter's cell phone was set to ring (and vibrate at the same time) twice, once at the beginning of the task and once in the middle of the task as distractors. The task was coded for the length of time children were off task (either disrupted by distractors such as the phone, distruptive such as talking to other children or offtask for no reason) during the "puzzle time".

The achievement battery: two subsets of achievement tasks from the Woodoock Johnson Psycho-Educational Battery-III Tests of Achievement (Woodcock & Mather, 2000) were used: The Woodcock Johnson Letter-word Recognition (reading) and Applied Programs (math). The Applied Problems subtest measures early math and included questions about quantity, time, money, and word problems. The Letter-word Identification subtest measured early literacy, questions included naming letters and words.

Challenge tasks coding

A coding scheme was developed separately for each challenge task. For each task, points were given for targeted behavior per trial. Points were then averaged across trials for each child. Finally, final scores across different days per task per child were averaged and used as the final dependent variables. In sum, the challenge inhibition and working memory task scores were each computed from nine trials across different days.

For the *challenge inhibition task*, points were given based on the speed and accuracy of the child's stopping. Four points were given for immediate stop when the music stops, three for delayed stop with one more step, two points for delayed stop with two more steps, one point for delayed stop with three or more steps, zero point for forces stop (i.e. a child stops only be cause a child in front of him/her stops), non-stop or not participating in the game.

For the *working memory challenge task*, points were given based on the accuracy of the action. Two points were given for correct response (e.g. jumped three times), one point for attempt but failed response (e.g. clapped three times instead of twice), zero points for not responding or irrelevant or completely incorrect response.

Lastly, the *attention control task* was coded for the length of time students were off-ask during the puzzle task (including disrupted by distractors, actively off-task such as disturbing other children and passively off task such as staring at non-task related objects).

<u>*Coder training*</u>

Coders were four undergraduate research assistants who majored in psychology and were blind to hypotheses. Coders carefully reviewed code descriptions and viewed videotapes of classrooms before attending a coder's training workshop. During training, they were instructed with two detailed example videos for coding and questions that might arise during the coding process. After the workshop, coders conducted pilot observations and coded one to two additional videotaped cases. Finally, coders were required to pass a reliability test involving coding two to six children for the entire challenge task. Coders obtained an 80% ~85% match (Cohen' Kappa) with each other and maintained this level of reliability on a reliability retest with two children chosen randomly. All coders passed at these levels on a reliability test before being certified to code the videos used for data analysis.

Chapter 3

Results

Research Question 1, What is the nature, variability and reliability of the "Challenge tasks" as well as other tasks?

In this section, an overview of the distribution and variability of all variables will be provided; we will start with presenting individual performance, variability and reliability for all tasks, particularly those of the challenge tasks, floor effects of stop signal and factor analysis of teacher report of students' self-regulation. We will then move on to classroom variables and finally, to other background variables.

Table 2. Descriptive statistics for all self-regulation and achievement tasks.

	N	Mean (SD)	Minimum	Maximum	Skewness	Kurtosis	Reliability[a]
HTKS	125	17.89 (12.46)	0	37	-.28	-1.34	.95[c]
Sentence Completion	129	16.64 (10.33)	0	35	-.42	-1.12	.87[c]
Pair Cancelation	134	13.39(8.02)	0	37	.30	-.07	.81[f]
Operation Span	138	8.69(3.44)	0	16	-.46	.29	.72[c]
Cognitive Flexibility	138	16.36(4.08)	0	23	-2.03	5.51	.87[c]
Flanker							—
Congruent trials	104	.72(.25)	0	1	-.80	-1.25	
Incongruent trials	104	.57(.30)	0	1	-.18	-.47	
Stop signal[a]							—
Nonstop block nonstop trial accuracy	106	.63 (.25)	0	1	-.56	-.69	
Stop block nonstop trial accuracy	106	.38(.25)	0	1	.42	-1.09	
Step	106	438.7(2640)	-3300	8100	1.03	.49	
Number of times (sustained attention)	139	4.10(.94)	.1	7	.60	.20	—
Challenge-inhibition (Fixed trial)	139	3.67 (1.36)	0	7	-.12	.01	.56[c]
Challenge-working memory (Fixed trial)	139	6.71(2.99)	0	12	-.33	-.50	.50[c]
Challenge-attention control	80	2.06(1.16)	0	4.99	.74	.37	.91[e]
Teacher Rating of Self-regulation	139	4.96 (1.7)	1	7	-.70	-.73	.97[d]
Letter-word Identification	138	10.84 (4.54)	1	26	-.09	.42	.85[f]
Applied Problems	136	12.7(4.29)	1	20	-.63	.05	.85[f]

Note. HTKS = Head Toes, Knees, and shoulders. There was no consistent ceiling or floor effects across tasks, i.e. children who scored zeros on one task did not consistently scored zero or low on other tasks. [a]Showed floor effects (see analysis in the results section). [b]The original task scores were skewed and were transformed using square root transformation. [c]Reliability was calculated by adjusting split-half (odd-even)with the Spearman-Brown prophecy formula. [d]Reliability was calculated by Cronbach's Alpha. [e]Intraclass correlation (ICC) was calculated because the data were coded at interval level. [f]Official data from Woodcock Johnson. For challenge tasks, only data from fixed trials were used for final analysis.

Overall Individual reliability. There was individual variability in children's performance on self-regulation. Table 2 provides descriptive statistics of all variables including means, standard deviations and reliabilities. Participants showed fairly large variability on all variables (as expected from the diverse sample) and the distributions were close to normal in most cases. It's interesting to note that the number of times (usually on different days) it took children to finish the tasks varies (when children became inattentive, experimenters stopped testing after finishing one complete task and moved testing to another day). It ranged from one time to seven times with an average of 4.1 times (*SD*=1.8). 14% of children were unable to complete one or more tasks despite repeated attempts (they tried at least seven times thus giving a number seven in calculating mean scores).

Sustained attention. A closer look at the number of times it took children to complete that battery of tasks revealed that it negatively correlated with their performance on two of our attention measures in school A: pair cancelation (r (68) =0.25, p < .05) and teacher report of attention control (r (68) = 0.32, p < .05). This measure negatively correlated with family income (r (32) = .39, p< .05) and teacher/class type (r (65) =. 35, p<. 01) in school B (school B has three types of classes while school A has one so analyzed the schools separately). Specifically, in school B, the Head Start children spent significantly more sessions to finish the tasks compared to children in the tuition based classes and state funded readiness programs combined (t (2)= 2.54, p < .05); the tuition-based children spent the least number of times completing the tasks (t (2)= 3.4, p < .01).

Challenge tasks: The distributions of the challenge tasks were close to normal (the attention control task was transformed using square root transformations). The three-time measurements for inhibition and working memory tasks showed that the children's performance was reliable when sampling at different time points: the Cohen's Kappa among the three time points (these time points were taken at least a day apart) for the inhibition challenge task was .50 and .56 for the working memory and inhibition task. Possible confounding factors included timing (e.g. children were tested before or after a day's activity) and size of the location where challenge tasks took place (e.g. gym, which tend to be big or normal classroom, which tended to be smaller). The challenge attention task was videotaped once.

Stop signal: There was a possible floor effect on the stop signal task: the average accuracy rate for choosing the red heart (pressing 1) and green tree (pressing 4) in the stop block is .33, which was below chance (chance = .5 if only looking at visual stimuli) or equal to chance (chance = 0.33 if considering all responses: pressing 1, pressing 4 and not responding). In either case, the task might not be reflecting children's true performance on inhibition; it was therefore excluded from all analysis.

Teacher questionnaire: Questions in the teacher questionnaire reached an alpha reliability of .97. Exploratory Factor Analysis was used to further determine the components within teacher reported self-regulation. Factor analysis using direct oblimin rotation was performed. Direct oblimin rotation was used because it allowed correlation between factors (Gorsuch, 1997). Factor analysis revealed four factors, the first one was clearly working memory, items loaded on this factor included “this child utilizes multiple rules to complete a task”, “this child follows two-step instructions” etc.; the second factor was “attention” which included items such as” this child has a short attention span” and “this child only pays attention to things he/she is really interested in”, “this child is easily distracted”; the third factor was “inhibition ”, items included “this child interrupts or intrudes on others” and “this child runs about or is very active in situations where it is inappropriate”. Factor 2 and factor 3 were highly correlated with each other(r (138)= .84, p< .001). The first factor was also correlated with factor 2 (r (138)=. 70, p<. 001) and factor 3 (r (138) =. 70, p<. 001). Note that factor 2 and factor 3 can also be interpreted alternatively such as behaviors towards oneself (e.g. being inattentive) for factor 2 and destructive behavior towards others for factor 3 (e.g. intruding in other’s games). The fourth factor had two items (“this child waits to be called on before responding” and “this child waits for his/her turn”) and was labeled “compliance”. The first three factors, “working memory”, “attention” and “inhibition” were core components of self-regulation of theoretical interest to the current study. Separate scores for each factor were computed for further analysis.

Classroom variability. Despite the wide demographic differences in classrooms (i.e. head- start classrooms, state-funded school readiness programs and tuition based classrooms) and variability on achievement measures, there was no consistent classroom or teacher effect on self-regulation or achievement. For example, although the two head start classrooms scored lower on reading (t (103)= 2.34, p < .04) compared to other classes, there was no difference between head start children and other children on any other direct measures of self-regulation.

Other background variables: In addition, to determine the nature and variability in background variables, all background variables of theoretical significance were entered in models used to predict self-regulation performance and achievement respectively. Among these variables (i.e. age, gender, classroom, teacher, family income), age explained the most variance and significantly predicted both math (B= 3.8, t (5, 43) = 4.2, p <. 05) and reading achievement (B= 3.2, t (5, 43) = 3.5, p < .05). Although there was no age difference in most direct self-regulation measures (except that older children performed better on pair cancelation, B = .53, t (5, 43) =4.04, p <. 01)), teachers rated older children to have better overall self-regulation skills (B=. 42, t (5, 43) =2.94, p< .01). Because age's significant effects on achievement, it was controlled in subsequent models involving self-regulation and achievement.

Research Question 2, What are the relationships between challenge tasks and other self-regulation tasks (experimental tasks, teacher questionnaire)?

In this section, zero-order correlations between challenge tasks and experimental tasks are reported, followed by the correlations between challenge tasks and the teacher report of self-regulation. Finally, factor analysis with all the direct measures was performed to investigate the relation among all measures and components of self-regulation.

Relationships between challenge tasks and laboratory based self-regulation tasks

In general, challenge tasks correlated with most of the experimental tasks. There tended to be stronger correlations if the tasks were designed to measure the same components (table3). For example, the challenge working memory task correlated with Sentence completion (*r* =. 34, *p* < .001) and Operation Span (*r* =. 26, *p* < .05), they all measured working memory; the Challenge attention control task correlated with Pair Cancellation (*r* =. 27, *p* < .001), and flanker (*r* =. 29, *p*<. 05), all of which measure attention control. At the same time, some tasks that were designed to measure different components across contexts generally also showed moderate correlations. Finally, all challenge tasks correlated with cognitive flexibility (*r*=. 29, *r*=. 24, *r*=. 47, all *ps*<. 05). Challenge inhibition and challenge working memory correlated with each other (r = .21, *p*<. 05), but not with the challenge attention control task (all *rs*<. 10, *ns*).

Table 3. Correlations between Experimental tasks and Challenge Tasks

	HTKS	Sentence completion	Pair cancelation	Flanker (incongruent)[1]	Flanker (con-incon RT)[1]	Operation span	Cognitive flexibility
Challenge-inhibition	.29**	.33**	.32*	.24*	. 19^{+}	.18^{+}	.29*
Challenge-working memory	.27*	.34**	.28**	.31**	.14	.26*	.24*
Challenge-attention control	.12	.13	.37**	.29*	.18	.31*	.47**

Note. The correlation matrix was estimated with maximum likelihood and was adjusted for nonindependence and missing data by Mplus. The challenge attention task scores (time spent off task) were reversed so that larger values indicate better performance. HTKS = Head to Toes, Knees to Shoulders.

In addition, Experimental tasks were generally moderately correlated with each other (table 4), confirming previous results that components of self-regulation were moderately related. The correlation between congruent and incongruent trial of the flanker task[1] was *r (104)*=. 70, *p*<. 01.

Table4. Correlations between experimental tasks.

	Sentence Completion	Pair Cancelation	Operation Span	Cognitive Flexbility	Flanker (incongruent)	Flanker (con-incon RT)
HTKS	.39**	.37**	.31**	.36*	.21*	.14
Sentence Completion		.41**	.18^{+}	.22*	.24*	.08
Pair Cancelation			.30**	.35*	.36**	.14
Operation Span				.36**	.22*	.09
Cognitive Flexibility					.31**	.13
Flanker (incongruent)						.19^{+}

<u>Relationships between challenge tasks and teacher reported self-regulation</u>

In general, challenge tasks correlated moderately well with teacher report. The relationship was particularly strong for working memory tasks: the teacher reported working memory task was associated with all challenge tasks (r>.25, ps< .05) and the challenge working memory task was associated with all aspects of teacher reported self-regulation (r>.25, ps<. 05).

The teacher reported working memory task also correlated well with almost all experimental tasks (r>.22, p<. 05). The experimental working memory tasks (Operation Span and sentence completion) were less well correlated with aspects of teacher reported self-regulation and the challenge tasks.

Table 5. Correlations between challenge tasks and teacher report of self-regulation, experimental tasks and the teacher report.

Challenge tasks	TQ-Attention	TQ-Working Memory	TQ-Inhibition
Challenge-inhibition	.10	.26*	.12
Challenge-working memory	.37**	.26*	.26*
Challenge-attention control	.16**	.40**	.16
Experimental tasks			
HTKS	.31**	.44**	.25*
Sentence completion	.20^{+}	.22*	.19^{+}
Pair Cancelation	.37**	.53*	.31*

Operation Span	.27*	.44**	.11
Cognitive Flexibility	.15	.34**	.02
Flanker Incongruent Incon-con RT	.21*	.29**	.05
	.01	.10	.00

Note. TQ=Teacher Questionnaire. HTKS=Head to toes, knees to shoulders.

Confirmatory Factor analysis

Confirmatory factor analysis (CFA) was used to analyze the data. Model fit was evaluated by three criteria: the Chi-square value indicating the acceptance or rejection of the null hypothesis (null was that the model was a good fit), a significant Chi-Square value would indicate a poor fit; a Comparative Fit Index (CFI), a larger than .95 CFI indicated a good fit; and the Standardized Root Mean Residual (SRMR) of less than .08, indicating a good fit. Both the CFI and SRMR were suggested by Hu and Bentler (1998), All possible models of theoretical importance were tested, particularly “single versus multiple components of self-regulation” models and “single versus multiple contexts in which self-regulation was measured” models:

Single versus multiple components of self-regulation. A single factor compromising all the tasks was used to fit a model. The model also indicated a very good fit, χ^2 (23)=24, *p*=. 70; *CFI*=1.0; *SRMR*=. 05. All variables loaded significantly on the factor. A-three-factor model indicating the three factors of self-regulation (i.e. inhibition including HTKS, challenge_inhibition, working memory including operational span, sentence completion, challenge_working memory, and attention control including pair cancelation, challenge_attention, flanker) were then used to fit the data. χ^2 (24)=20 *p*=. 60; *CFI*=1.0; *SRMR*=. 04, but the correlations between the three factors are extremely high, *r* (inhibition and working memory)=. 98, *r* (working memory and attention control)=. 85 and *r* (inhibition and attention control)=. 83, indicating the three factors can probably be regarded as one factor.

Single versus multiple contexts in which self-regulation was measured. The single factor model would be the same as the single factor model in the previous section. Models measured self-regulation in different contexts were explored, latent variables included all the lab based behavioral self-regulation tasks as factor one (HTKS, cognitive flexibility, flanker, pair cancelation, sentence completion and operational span) and three challenge tasks as factor two were used to fit a two-factor model. The model did not converge, indicating changes required for revision. A third model was built moving the flanker task from factor one to factor two as it has a poor loading in the previous model. The revised model has a good fit, χ^2 (26)=22.4, *p*=. 67; *CFI*=1.00; *SRMR*=. 05. All variables loaded significantly on each factor respectively. However, the two factors were also very highly correlated (*r* = .93).

Due to the high correlations between factors in multiple-component models and extremely good fit of the one-factor model, I chose the single-factor model as the best fitted model by rule of parsimony. Table 6 shows the factor loading and R^2 information of the single factor model; all tasks loaded on the factor significantly and explained decent variance.

Table 6. CFA Factor loadings.

	Single factor model	
	Factor	R^2
HTKS	.58***	34***
Sentence Completion	55***	30***
Pair Cancelation	.63***	39***
Operation Span	.43***	19*

Cognitive Flexibility	.56***	32***
Flanker[2]	47***	22*
Challenge-Inhibition	.48***	23**
Challenge-Working memory	.43***	19*
Challenge-Attention	.51***	26*

Note. $^{+}p<.10$ $*p<.05$ $***p<.001$.

Research Question 3: Do challenge tasks predict achievement?

In this section, joint contribution (contribution of challenge tasks to achievement without controlling variance explained by experimental tasks) of challenge tasks to achievement was first explored. The unique contribution of challenge tasks (controlling for experimental tasks and/or teacher questionnaire) was then investigated. In both cases, age was controlled.

Joint contribution of Challenge tasks (no control for other self-regulation tasks): Controlling for age, challenge inhibition significantly predicted math achievement (B = 1.76, p <. 01) but not reading (B = .92, ns); Challenge working memory predicted both math (B = 2.1, p < .001) and reading (B = 1.79, p < .001); challenge task attention control predicted reading (B = 2.1, p<. 05) but not math (B=. 01, ns).

Unique contribution of Challenge tasks: To explore the variance that challenge tasks added to the experimental tasks when predicting achievement, regression models were built with challenge and experimental tasks. Since there were multiple laboratory experimental measures and only one challenge task for each self-regulation component, combined measures of experimental tasks (e.g. combined z scores for working memory using sentence completion and operation span) were computed. Results showed that controlling for age and lab inhibition tasks (i.e. additional variances the challenge inhibition task explained for in addition to laboratory inhibition measures and age), the challenge inhibition task predicted math (B=. 14, p< .05) but not reading achievement (B=. 12, ns); the working memory challenge task added significant variance to both reading (B=1.0, p<. 01) and math achievement (B=1.70, p<. 05). The lab working memory task also predicted math (B =1.67, p<. 05). The challenge working memory task was a much better predictor of reading achievement (B=1.9, p<. 01). Contrary to the significant effect working memory and inhibition challenge tasks had, the attention control challenge task did not add variance to laboratory measures, although it predicted reading alone.

Finally, when controlling both experimental tasks and teacher report of self-regulation (the unique contribution of challenge tasks on top of both experimental measures of self-regulation and teacher questionnaire), the inhibition and working memory tasks still predicted math achievement (B=1.32, p<. 05 and B=. 96, p<. 1), and the working memory challenge task still predicted reading (B=1.5, p<. 01). Table 7 showed coefficients in the regression models.

Table 7. Regression coefficients.

		Reading									Math								
		Model 1 Joint Contribution[a]			Model 2 Unique Contribution[b]			Model 3 Unique Contribution[c]			Model4 Joint Contribution[a]			Model5 Unique Contribution[b]			Model6 Unique Contribution[c]		
		B	SE	β	B	SE	β	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Inhibition	Challenge Inhibition	.92	.71	.13	.12	.74	.16	1.2	.69	.17⁺	1.76	4.0	.28**	.14	.6	.20*	1.3	.62	.20*
	Inhibition experimental	--	--	--	.08	.04	.21*	.07	.04	.20⁺	--	--	--	.13	.03	.38**	.12	.03	.34**
	TQ-inhibition	--	--	--	--	--	--	.67	.26	.26*	--	--	--	--	--	--	.41	.22	.17⁺
Working memory	Challenge Working memory	2.1	.60	.32***	1.94	.68	.29**	1.53	.58	.23**	1.79	.54	.30**	1.70	.57	.19*	.96	.57	.16⁺
	Working memory experimental	--	--	--	.72	.63	.12	.18	.55	.03	--	--	--	1.67	.54	.31*	.14	.54	.27**
	TQ-Working memory	--	--	--	--	--	--	2.3	.41	.51***	--	--	--	--	--	--	.88	.40	.21*
Attention	Challenge Attention	.49	.50	.12	.05	.54	.01	.10	.49	.02	1.11	.44	.29*	.48	.44	.13	-.52	.44	-.14
	Attention control experimental	--	--	--	2.1	.9	.35*	1.78	.75	.30*	--	--	--	2.3	.73	.42**	2.47	.67	.44***
	TQ-Attention Control	--	--	--	--	--	--	1.16	.31	.44***	--	--	--	--	--	--	.74	.27	.29**

Note. [a]Age was controlled. [b]Experimental tasks and age were controlled. [c]Experimental tasks, Teacher Questionnaire and age are controlled.
⁺p<.10, * *p*<.05 ** *p*<.01 *** *p*<.001

Footnotes

[1]Two indicators from the Flanker task were reported, including the accuracy rate on the incongruent trials and a more commonly used indicator, the difference between reaction time in Congruent and Incongruent trials. If the congruent and incongruent trials were in separate blocks, the congruent trial block would serve as a baseline block, enabling one to separate the resistance to distracter effect from simply identifying the direction of the center fish. However, the two kinds of trials were mixed together in our task, a participant's performance on congruent trials was likely to be slowed down as a result of responding to a previous incongruent trial. Therefore, although reporting two indicators, we favor the inaccuracy rate of the incongruent trials and have used it in the all analysis. The correlations between congruent and incongruent trials are r (107) =. 73, p<. 001 while the correlation between incongruent and the difference scores is r (107)=. 19, p<. 10, the correlation between incongruent accuracy and the difference in accuracy is .56, p<. 001

Chapter 4

Discussion

The study designed a series of field-experimental task that were used to evoke students' self-regulation behavior in the classroom. It used a multi-dimensional approach incorporating measures that assess three components of self-regulation, inhibition, working memory and attention control. Results demonstrated that the challenge tasks were reliable and valid assessments in measuring self-regulation, outperforming the laboratory and naturalistic measurements in some aspects.

1) The challenge tasks correlated with both experimental measures and naturalistic measures of self-regulation sharing variance with tasks from two different contexts, in particular, the challenge working memory task was shown to have good associations with most other tasks;

2) All the challenge tasks predicted academic achievement by themselves, particularly math achievement; in addition, controlling for both experimental and naturalistic measurements, the challenge inhibition and working memory tasks still predicted achievement, capturing variances traditional laboratory tasks did not capture.

3) The challenge tasks were easy to administer and analyze, drawing from multiple days of observations thus increasing the reliability of traditional experimental and naturalistic tasks.

4) As the first study aiming to develop quick and reliable tools for teachers, the design of the challenge tasks may be a starting point for developing further measures.

The discussion section is organized around the following three themes, First, the nature and variability of the sample and the reliability of challenge tasks; Second, the relationships between challenge tasks and existing self-regulation measures; Third, challenge tasks and academic achievement (i.e. ecological validity); finally, the advantages, implications and limitations of the challenge tasks and the study.

Nature and variability of the sample and Challenge tasks

We studied children who were 4-5 years old and designed challenge tasks for them. We were interested in this age group (approximately 3-6 years of age) because children undergo dramatic development of self-regulation skills at this age (e.g. Diamond & Taylor, 1996). Further, the sample was chosen in the Midwestern suburban area and is representative of the U.S. population (the average family income in our sample is very close to the average U.S. house hold income).

We found individual variability in self-regulation and achievement in this sample. In addition to varying in core inhibition, attention control and working memory measures, children also differ in persistence/sustained attention. It is interesting to see that in school B, the lower family income parents have, the worse sustained attention/persistence children have, despite the fact that other direct measures did not show such correlation. This might imply that children from low-income families might especially need training in sustained attention as such ability increase at this age (4-6 years of age) (Levy, 1996). Some head-start programs have been shown to increase children's sustained attention (Love et al., 2005) but in general very few studies have not systematically studied this problem.

Despite the high variability of the sample, the challenge tasks proved to be reliable measures of self-regulation and considerably increased reliability of traditional video observations. Landis & Koch (1977) argued that a reliability of .6 - .8 was sufficient agreement for inter-coder reliability (two coders that code the same instances of observations) and .8+ is "almost perfect". Due to the highly structured nature of the challenge tasks, coding was easy. All inner- and inter- coder reliabilities reached .8 and above requiring only minimum adjusting and re-adjusting between coders. Additionally, the challenge tasks were one of the few studies to videotape children behaviors in the classroom repeatedly on different days. Despite the fact that the tasks were used and videotaped on different days which often vary in time, location, prior activities, even peers, the challenge task still produced reliable data: the Cohen's kappa among three different days of observations reached .5 to .6 level, comparable to sufficient classic coder reliability that were aimed to code the same instances.

The relationships between challenge tasks and existing self-regulation measures.

The three components of self-regulation were found to be correlated (Miyake et al., 2000), the challenge tasks generally correlated well with both experimental self-regulation tasks and teacher report of self-regulation across all components, confirming that the tasks shared common variances with those used in very different contexts. In particular, the challenge working memory task was related to all experimental self-regulation tasks and all aspects of teacher reported self-regulation.

Working memory was found to be one of the most critical components of self-regulation perhaps (Engle, 2002). A child has to first remember the instructions to be able to pay attention and participate. Unfortunately, traditional measurements only test children in decontextualized situations with abstract tasks such as asking them to recall digits backwards or remembering words or numbers. Sometimes children have trouble understanding instructions and concentrating on the tasks. Developing naturalistic/semi-naturalistic measurements of working memory is therefore critical in measuring children's true performance in and outside of the classroom. In the current study, the two experimental working memory tasks underperformed the teacher report of working memory (which correlated with most inhibition and attention control measures and was a better predictor of achievement) as well as the working memory challenge task. The challenge task is likely a better measurement of working memory as it allows children to behave in real world situations they are familiar with thus reducing the difficulty in understanding. Further, unlike traditional working memory tasks (e.g. the Working Memory Index of the Wechsler Intelligence Scales for Children, 4th edn, Wechsler, 2004) which are almost exclusively verbal in nature, the challenge working memory task does not rely on verbal or digital process thus assessing working memory abilities not restricted by verbal abilities or familiarities with numbers. Finally, the challenge working memory task might rely heavier on other self-regulation abilities such as resistance to distracters (e.g. peers) compare to traditional working memory tasks.

Additionally, confirmatory factor analysis confirmed that like other self-regulation tasks, the challenge tasks can be regarded as self-regulation measures that closely relate to existing measures (consistent with Wiebe et al., (2008) study).

Finally, it's worth noting that the cognitive flexibility task correlated well with tasks used in this study. It is consistent with previous findings, which showed that cognitive flexibility correlated with inhibitory control and working memory (Miyake et al., 2000). The results also suggested that cognitive flexibility may be the third component of self-regulation/executive function. Regardless of the theoretical debate over the components of self-regulation, factor analysis suggested that all these components are likely to be highly correlated for young children.

Challenge tasks and academic achievement

The challenge tasks were also demonstrated to have good predictability for math and reading achievement. Each challenge task individually predicted math and/or reading achievement; when controlling for experimental tasks, the challenge inhibition still predicted math and the working memory task predicted both reading and math. Such effects remained when controlling for teacher report of self-regulation. The strong effect of the challenge tasks, particularly the inhibition and working memory to academic achievement revealed that 1), the challenge tasks are ecologically valid measurement by themselves, and 2) the challenge working memory and inhibition tasks added unique contributions to existing self-regulation measurements in predicting academic achievement.

The challenge tasks overlap with traditional self-regulation measures yet explain additional variances traditional measures might not measure, some possible reasons could be: 1. The challenge tasks elicit children's true performance by engaging them in familiar games in their everyday environment. Traditional experimental tasks tend to be abstract (e.g. counting digits backwards) and isolated (i.e. often measured a quiet place by a stranger – an experimenter) thus introducing obstacles for understanding. 2. The tasks capture reliable performance by taking multiple measurement points as oppose to single time measure used by traditional self-regulation tasks; 3. Each challenge task not only tests one basic cognitive skill but also a wide rang of skills involved to successfully carry out the behaviors in multi-faced environments, such as avoid being distracted by peers, focusing on the teachers and sustained their attention.

Implication

The multidimensional challenge tasks were proven to be reliable (had good test-re-test reliability) and valid (predicted academic achievement). They also overlapped with both experimental and naturalistic tasks in variance. In addition to the theoretical and methodological implications, challenge tasks also have implications for assessing children's self-regulation skills at school by teachers and teacher professional training.

Since the tasks are conducted in school in a very short period of time, administered by teachers and are very similar to what teachers normally do, they can serve as a useful tool for teachers to assess children's self-regulation skills on any normal day. Traditional experimental tasks needed to be carried out by trained experimenters performing individual assessment; traditional naturalistic tasks require large amount of coding. These challenge tasks can be easily used, videotaped and observed by teachers to identify children who are lacking self-regulation skills. Since the tasks are highly structured, teachers can simply observe the videotape and take notes about children's self-regulation skills, which can be used as unbiased indicators for identifying low self-regulation children at school entry. Second, the challenge tasks can also be used for professional teacher training in identifying low-self-regulated children or train novice teachers to lead classroom activities. Although making testing natural and easy, having teachers lead experimental tasks might inevitably complicated the tasks by introducing individual differences among teachers.

Additionally, the teacher questionnaire developed in this study was proven to be a reliable (very high reliability within the items) and valid measurement (all aspects of it predicted achievement and most of the aspects correlated with existing measures) of self-regulation that measured each component of self-regulation. It can be used as an indirect, naturalistic measurement of self-regulation.

Limitations and future studies

The challenge task – attention control. We should point out that the challenge-inhibition and working memory tasks are better predictors of math and reading achievement compared to the attention control challenge task. This could be because inhibition and working memory are themselves better predictors (Blair et al., 2005); it is also possible that the attention control challenge task was added with more errors when used in small group settings (in comparison to other challenge tasks which were used with all children in a classroom) or have the most missing data in the current dataset. Future studies should build on the current attention control challenge task and investigate the contribution of other "attention control" challenge tasks to achievement.

Floor effect for the stop signal task. The stop signal task used in this study has a floor effect. Some possible reasons are: 1). The stop signal task is not developmentally appropriate. Previous studies have found that 3-4 year old children have difficulty with the stroop like task (Gerstadt, Hong, & Diamond, 1994) comparing to order children. 2). Children may find remembering three step instructions at the same time (press the red button when see the red heart; press the green button when see the green tree; and do not press anything if hear a beep) extremely challenging. 3). The stop signal task is relatively more abstract and uninteresting. The flanker task works well where the stimuli are fish with names and has feedback sounds when children successfully catch a fish; in comparison, the stop signal task does not provide a story, interesting stimuli or sound feedback. In either case, adapting a new stop signal task and addressing the above possibilities would be helpful to eliminate the flooring effect on this task in future studies.

In the current study, the challenge tasks were standardized by providing all teachers with standard instructional scripts, using the same marching music, and controlling the length of each trial; future studies should further perfect the tasks by choosing locations and testing time points that are most comparable across all classes (e.g. test in the morning after snack time in the gym).

Despite the complications, the challenge tasks are the first stop in developing field-experimental tasks and could be a great tool in understanding and assessing children's self-regulation.

Appendices

A: Teacher Questionnaire

This child:	Strongly Disagree	Disagree	Somewhat Disagree	I don't know	Somewhat Agree	Agree	Strongly Agree
1. waits patiently for her/his turn	1	2	3	4	5	6	7
2. follows one-step instructions	1	2	3	4	5	6	7
3. is easily distracted	1	2	3	4	5	6	7
4. is prone to disturb other children	1	2	3	4	5	6	7
5. has a short attention span	1	2	3	4	5	6	7
6. follows two-step instructions	1	2	3	4	5	6	7
7. leaves seat in classroom in which remaining seated is expected	1	2	3	4	5	6	7
8. only pays attention to things he/she is really interested in	1	2	3	4	5	6	7
9. delays gratification on appropriate occasions (e.g. waits to eat his/her treat)	1	2	3	4	5	6	7
10. follows multiple step instructions (e.g., first, wash your hands; second, get some wather; third, eat your snack)	1	2	3	4	5	6	7
11. utilizes multiple rules to complete a task	1	2	3	4	5	6	7
12. fidgets with hands or feet or squirms in seat	1	2	3	4	5	6	7
13. waits to be called on before responding	1	2	3	4	5	6	7
14. have difficulty remaining still	1	2	3	4	5	6	7
15. runs about or is very active in situations where it is inappropriate	1	2	3	4	5	6	7
16. interrupts or intrudes on others (e.g., butts into others' conversations or games)	1	2	3	4	5	6	7
17. has difficulty playing or engaging in leisure activities quietly	1	2	3	4	5	6	7
18. restless, always up and on the go	1	2	3	4	5	6	7

B. Coding Scheme

(Note. Not all the codes were used in the analysis)

Inhibition

The Inhibition game: In this game, children were instructed to freeze and become a statue when the music stops.

For each trial:

Items	Codes	Definition
Inhibition trial	4	Immediate stop (no more step)
	3	Stop (1 more step)
	2	Delayed stop (2 steps)
	1	Delayed stop (3 or more steps)
	0	Forced stop (can not walk anymore since other children stopped, did not stop or stopped but start running around
	Time1:	Reaction Time (time from music stops to the child stops)
	Time2:	Time spent frozen(until music starts or teacher says 'unfreeze' or music starts)

Working Memory

The Working Memory game: In this game, children were instructed to follow multiple instructions. Trials 1 to Trial 3 are fixed trials, in which children were asked to follow one–step-instruction (jump three times), two-step-instruction (jump three times and clap twice), and three-step-instruction (jump three times, clap twice and walk one step back) respectively. Trial 4 to Trial 6 were varied two-step-instruction trials in which teacher specified activities (e.g. swim twice and jump once) that were familiar to these children. In all the trials, both accuracy and reaction time is coded. Only fixed trials were analyzed in the current study.

One-Step Instruction-Jump three times

Coding Items	Codes	Definition
One step instruction	2	Correct response
	1	Attempt, but failed to respond
	0	No response
	Reaction time 1	Reaction time until stops

	Reaction time 2	Reaction time until start responding (no reaction time if did not respond at all)

Two- Step Instruction-Jump three times and clap twice

Coding Items	Codes	Definition
WM-two step instructions	1	Jump three times
	2	Correct response
	1	Attempt, but failed to respond
	0	No response
	Reaction time 1	Reaction time until stops
	Reaction time 2	Reaction time until start responding (no reaction time if did not respond at all)
	Reaction 2	Clap twice
	2	Correct response
	1	Attempt, but failed to respond
	0	No response
	Reaction time3	From stop jumping to start clapping
	Accuracy	Order of Actions
	1	Jump, Clap
	2	Clap, Jump
	0	By itself: Both actions not attempted Behind 1 or 2 (i.e. 10, 20): The last action of the sequence not attempted.
	NRA	Non-related Action (attempted an action that was required in a previous trial)

Three-step Instruction: Jump three times and clap twice and go one step backwards

Coding Items	Codes	Definition
WM-three step instructions	1	Jump three times
	2	Correct response
	1	Attempt, but failed to respond
	0	No response
	Reaction time 1	Reaction time until stops

Reaction time 2	Reaction time until start responding (no reaction time if did not respond at all)
Reaction 2	Clap twice
2	Correct response
1	Attempt, but failed to respond
0	No response
Reaction time3	From stop jumping to start clapping
Reaction3	One step backwards
2	Correct response
1	Attempt, but failed to respond
0	No response
Reaction time4	From stop clapping to start stepping back
Accuracy	Order of Actions
1	Jump, Clap, Step
2	Clap, Jump, Step
3	Clap, Step, Jump
4	Step, Jump, Clap
5	Step, Clap, Jump
6	Jump, Step, Clap
0	By itself: Both actions not attempted Behind a number (i.e. 10, 20): The last action of the sequence not attempted.
00	Behind a number: Last two actions in the sequence not attempted.
NRA	Non-related Action (attempted an action that was required in a previous trial)

<u>Attention Control</u>

Off-task behavior includes off-task non-distruptive and off-task disruptive.

<u>Off-Task – Non-Disruptive</u>

Off-Task means that the child is focused on some object/event other than the assigned task; his/her visual, behavioral, conversational focus is off-task for 2 seconds or more. This can cover a range of behaviors, including distracted behavior. Examples include passive lack of focus such as looking around the room in a distracted manner for 2 seconds or more, lack of visual attention to experimenter's demonstration, or not responding to the experimenter. This may also include being redirected by the experimenter to pay attention to the task; involved nose picking; or playing with clothes instead of doing task. Consider immediate expectations for child (e.g., if the child has finished and is waiting for the experimenter, s/he might look around while waiting). Don't count fleeting glances away or engagement with another child's puzzle as Off-Task.

Off-Task - Disruptive

Off-Task means that the child is focused on some object/event other than the assigned task; his/her visual, behavioral, conversational focus is off-task for 2 seconds or more.

Off-Task Disruptive behavior meets all of the criteria for Off-Task behavior and there is also evidence that the experimenter is distracted or involved in the episode (e.g., gives verbal correction) or evidence that other children are distracted by the behavior (their visual attention is on the off-task child during the disruption).

This can cover a range of behaviors. Examples include attempting to engage experimenter in discussion or making comments that do not pertain to the task at hand (e.g., "It's my birthday next month") while discontinuing work on the puzzle and directing visual attention to the experimenter; disruptive movement (e.g., flicking puzzle pieces), or engaging socially with other children (e.g., poking them). Consider immediate expectations for child (e.g., getting out of seat to pick up puzzle piece is not necessarily Off-Task).

Bibliography

Acock, A. C. (2007). 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,* 364-373.

Aron A. R., Fletcher, P. C., Bullmore, E. T., Sahakina, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience, 6,* 115-116.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65-94.

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

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647-663.

Bodnar, L.E., Prahme, M. C., Cutting, L.E., Denckla, M.B., & Mahone, E. M. (2007). Construct validity of parent ratings of inhibitory control. *Child Neuropsychology, 13*, 245-362

Bronson, M. B., Tivnan, T., & Seppannen, 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*, 253-282.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology*. *19*, 273-293

Calson, S. M (2005) Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology, 28,* 595-616.

Cameron, C. E. , McClelland, M. M., Jewkes, A.M., Connor, C. M., Farris, C. L. & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly. 28*. 141-158

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early Years: Algorithm-Guided Individualized Reading Instruction. *Science, 315,* 464-465.

Deak, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition Failulre: Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology, 86,* 194-222.

Derks, E., Hudziak, J., & Boomsma, D. (2007). Why more boys than girls with ADHD receive treatment: A study of Dutch twins. Twin Research and *Human Genetics*, *10,* 765-770.

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*, 352–362.

Engle (2002). Workig memory capacity as executive function. *Current Directions in Psychological Science, 11*, 19-23

Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function: A summary. In G. R. Lyon, & N. A. Krasnegor (Eds.), *Attention, Memory, and Executive Function (pp. 367-395).* Baltimore, MD: Paul H. Brookes Publishing.

Espy, K. A. (2004). Introduction: Executive control in preschool children. *Developmental Neuropsychology, 26*, 379–384.

Gathercole, S. E., &Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7yearsofage. *British Journal of Educational Psychology, 70*, 177–194.

Gerstadt, C. L., Hong, Y. J., &Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5–7 years old on a Stroop-like Day-Night test. *Cognition, 53*, 129–153.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68,* 532–560.

Grabowsk, L., (2007). The writing superiority effect in the verbal recall of knowledge: sources and determinants. In Rijlaarsdam. G. (Series Ed.) and M. Torrance. L. van Waes & D. Galbraith (Volume Eds.), *Writing and Cognition: Research and Applications* (Studies in Writing, Vol. 20, pp. 165-179). Amsterdam: Elsevier.

Hu, L., Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

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

Klein, C. (2003). Assessing the development of executive functions. *International Society for the Study of Behavioural Development Newsletter, 43*, 8-11.

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20,* 407– 428.

Kochanska, G., Murray, K. T., & Harlan, E.T. (2000).Effortful control in early childhood : Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36,* 220–232.

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

Landis & Koch (1997). The measurement for observer agreement for categorical data. *Biometrics. 33.*159-174.

Lan, X., & Morrison, F. (2008, July), Inter-correlations among components of behavioral regulation and their relationships with academic outcomes in China. Paper presented at the 20th Biennial Meeting of the International Society for the Study of Behavioral Development, Wurzburg, Germany.

Lan, X., Cameron, C. E., Miller, K.F., Li, S., Cortina, K., Perry, M., & Fang, G. (2009). Keeping their attention: Classroom practices associated with behavioral engagement in first grade mathematics classes in China and the United States. *Early Childhood Research Quarterly. 24. 198-211*

Levitt, H. (1970). Transformed up-down methods in psychoa-coustics. *The Journal of the Acoustical Society of America , 49*, 467-477.

Levy, F. (1980). The development of sustained attention (vigilance) in children: some normative data. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *21*, 77-84.

Love, J. M.; Kisker, E. E., Ross, C., Raikes, H.. Constantine, J., Boller, K. et al. (2005) *Developmental Psychology, 41*, 885-901.

McCabe, L., & Brooks-Gunn, J. (2007). With a little help from my friends?: self-regulation in groups of young children. *Infant Mental Health Journal, 28,* 584-605

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.

Mesulam, M. M. (1985). *Principles of behavioral neurolog**y.*** Philadelphia, PA: **E** A. Davis.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental and socio-demographic properties in an epidemiological sample of young, urban children*. Child Development, 75*, 1373-1386.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Muthén, L.K. and Muthén, B.O. (1998-2007). Mplus User's Guide. Fifth Edition. Los Angeles, CA: Muthén & Muthén

Müller, U., Zelazo, P. D., Hood, S., Leone, T., & Rohrer, L. (2004). Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development, 75*, 1594-1609.

Ponitz, C. C., McClelland, M. M., Matthews, J. M., & Morrison, F. J. (in press). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*.

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

Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, *16,* 228–240

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

Towse, J. N., Hitch, G. J., & Hutton, U. M. Z. (2002). On the nature of the relationship between processing activity and item retention in children. *Journal of Experimental Child Psychology*, *82*, 156–184.

Wechsler, D. (2004). *Wechsler Scale of Intelligence for Children* (4th ed).London: Harcourt.

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,* 131–149.

Wiebe, S. A., Espy, K. A., & Charak D., (2008) Using confirmatory factor analysis to understand executive control in preschool children: I. latent structure. *Developmental Psychology. 44,* 575-587.

Woodcock, R. W., & Mather, N. (2000). Woodcock Johnson Psycho-Educational Battery-III. Itasca, IL: Riverside Publishing Company.

Zelazo, P. D., Frye, D. & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37-63.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 63(3)*. *Serial No. 274.*