Title

Measuring self-regulation in a physically active context: Psychometric analyses of scores derived from an observer-rated measure of self-regulation

Journal

Mental Health and Physical Activity, 6(3)

ISSN

1755-2966

Author

Lakes, Kimberley D

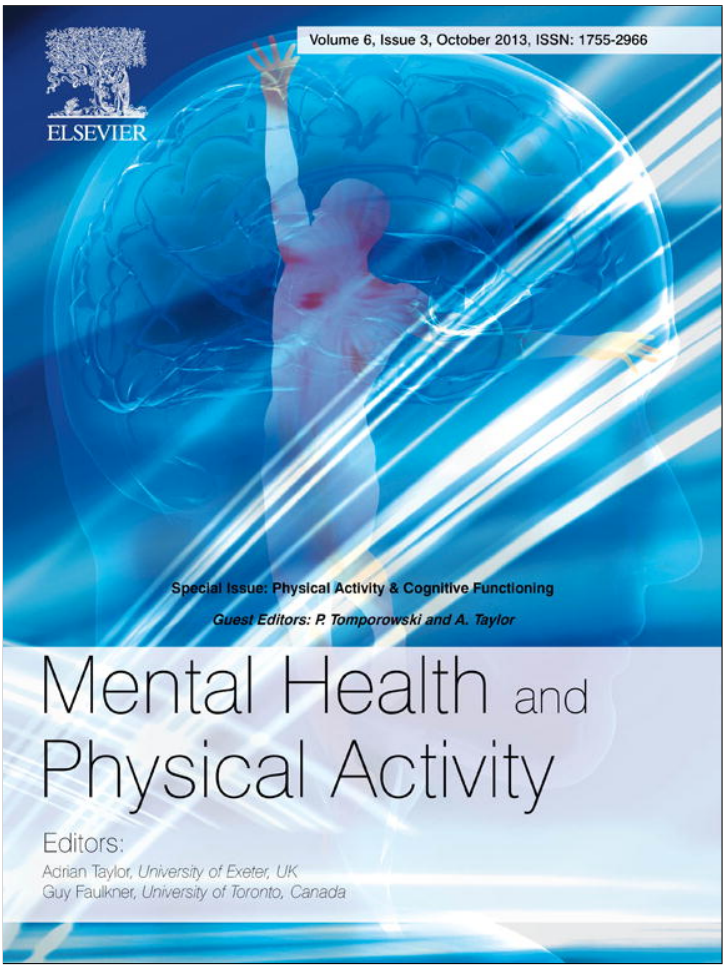
Publication Date

2013-10-01

DOI

10.1016/j.mhpa.2013.09.003

Peer reviewed



Mental Health and Physical Activity 6 (2013) 189–196

Contents lists available at ScienceDirect

Mental Health and Physical Activity

journal homepage: www.elsevier.com/locate/menpa

Measuring self-regulation in a physically active context: Psychometric analyses of scores derived from an observer-rated measure of self-regulation

Kimberley D. Lakes

Division of Development and Behavior, Department of Pediatrics, University of California, Irvine, 101 Academy Way, Suite 150, Irvine, CA 92612, United States

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form
9 September 2013
Accepted 10 September 2013

Keywords:
Self-regulation
Rating scale
Observation
RCS
Cognitive regulation
Affective regulation
Motor regulation
Executive function

ABSTRACT

The purpose of this study is to report psychometric properties of scores obtained using a novel observer-rated measure of children's self-regulation, the Response to Challenge Scale (RCS). The RCS was developed to rate children's self-regulatory abilities in a physically active context (e.g., while completing a physical challenge course). The RCS and other study measures were administered in a private school sample of 207 children. Analyses of score distributions indicated that the RCS was able to capture variance among children in self-regulatory abilities; the distribution was normal for the Affective, Cognitive, and Total Self-Regulation scales. Validity analyses revealed significant positive correlations between Cognitive, Affective, Motor, and Total Self-Regulation and executive function task performance; significant negative correlations between Cognitive Regulation and teacher-rated hyperactivity and inattention; significant negative correlations between Affective, Motor, and Total Self-Regulation and teacher ratings of peer problems; and significant positive correlations between Cognitive and Affective Regulation and parent ratings of prosocial behavior. Parent and teacher rated Total Difficulties scores were both negatively correlated with RCS Total Self-Regulation scores. Results suggest that it is possible for observers to rate self-regulatory abilities in the context of physical activities, and that these ratings correspond with performance on tasks requiring executive function as well as teacher and parent ratings of children's difficulties.

Self-regulation encompasses control over behavior, cognition, emotion, and motion (i.e., what is generally thought of as "self-control" or "inhibitory control," a core executive function) but also involves the maintenance of "levels of emotional, motivational, and cognitive arousal that are conducive to positive adjustment and adaptation, as reflected in positive social relationships, productivity, achievement, and a positive sense of self" (Blair & Diamond, 2008, p. 900). Self-regulation is crucial to positive outcomes as demonstrated by Moffit et al.'s (2011) 32-year observational study of 1000 children: children's self-control [which they equated with self-regulation (p. 2693)] predicted physical health, substance dependence, wealth, and criminal involvement in adulthood. As an important predictor of success in life, self-regulation is an important target for psychological measurement and intervention.

1. Self-regulation and physical activity

Physically active interventions can contribute to the promotion of self-regulation in children. There is growing evidence indicating that physical activity is important for optimal cognitive and behavioral functioning in children (e.g., Budde, Voelcker-Rehage, PietraByk-Kendziorra, Ribeiro, & Tidow, 2008; Davis et al., 2007; Diamond & Lee, 2011; Diamond, 2013; Tomporowski, Miller, Davis, Miller, & Naglieri, 2008; Tomporowski, Lambourne, & Okumura, 2011). Lakes and Hoyt (2004) demonstrated that a Taekwondo intervention promoted self-regulatory abilities (based on ratings from research observers and teachers) and executive function (based on performance on an executive function task) in school-age children. Tomporowski et al. (2011) reviewed several exercise interventions with children and concluded that there is compelling evidence suggesting that physical activity benefits children's cognitive functioning. They noted that that while there are surprisingly few studies examining the impact of physical activity on

E-mail address: klakes@uci.edu.

1755-2966/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mhpa.2013.09.003

cognitive functioning, the growth in this field in the last decade has been noteworthy.

The quality of research in this growing field will be in part impacted by the quality of outcome measurement designs. Best practices in measurement design emphasize the use of a multimethod, multisource approach to the measurement of psychological characteristics (e.g., Holmbeck, Li, Schurman, Friedman, & Coakley, 2002; Lakes, 2012a). Lakes and Hoyt (2009) described common sources of error in self-report, parent, teacher, and performance measures and noted that ratings by trained observers should be used more frequently in research with children because they can reduce common measurement errors (e.g., for example, trained observers can be fully crossed with all participants to reduce the impact of rater variance that is present when each child is rated by a different parent or teacher; in the latter situation, differences between parents and teachers in terms of leniency or severity are unknown and contribute to measurement error). Thus, the purpose of this manuscript is to examine an observer-rated measure of self-regulation designed to be administered in a physically active context as part of a multimethod, multisource approach to measuring physical activity intervention outcomes. In addition to using a novel, physically active context for rating cognitive and affective regulation, the measure is also novel in its inclusion of a motor regulation scale, as described below.

1.1. Response to Challenge Scale (RCS)

As has been noted in previous research (Gerber et al., 2012; Lakes & Hoyt, 2004), demonstrating sustained focus, determination, and resilience while confronting challenges and the pressure of performing is an important skill that may be associated with improved mental health outcomes. This ability has been referred to as mental toughness (Gerber et al., 2012) as well as self-regulation (Lakes & Hoyt). Lakes (2012b) noted that most assessment instruments designed to measure school-age children's self-regulation have used parent and teacher reports or performance on specific tasks and described the Response to Challenge Scale (RCS) as the first published observer rating scale developed to measure children's self-regulation in response to a series of physical challenges.

The RCS (Lakes, 2012b) is a theory-derived measure of children's self-regulation designed to measure cognitive, affective, and motor regulation. Generalizability studies have shown that the three factors on the RCS are interrelated, yet distinguishable (Lakes, 2012b), which is consistent with Diamond's (2000) characterization of the relationship between motor and cognitive development as "fundamentally intertwined." This close relationship between motor and cognitive development has been demonstrated in prior research. For example, Piek et al. (2004) studied the relationship between motor coordination and executive functioning in children between the ages of six and fifteen and found significant associations between performance on executive function tasks and motor ability. Prior studies have used separate instruments to assess motor and cognitive abilities; a review of current literature suggests that the RCS might be the first published observer-rated measure of motor, cognitive, and affective regulation for school-aged children.

In the intervention study that produced the data used in these analyses, the RCS was administered in a physical context theoretically designed to present children with a series of challenges, increasing in difficulty, in order to provide the opportunity for trained observers to rate the child's ability to regulate emotional, cognitive, and motor responses to the challenge. Children individually completed a physical challenge course while seven observers independently rated the child's ability to regulate cognition (e.g., attend to the course components, focus on the tasks at hand), affect (e.g., remain persistent in spite of challenges), and motor behavior or physical skill (e.g., exhibit skillfulness in completing physical tasks). The challenge course included tasks designed to vary in the demands they placed on children cognitively, affectively, and physically; for example, the course began with a relatively simple task (jogging a lap around the gymnasium), moved to more challenging tasks (running between cones; crab-walking), and ended with a very challenging task (jumping over a hurdle and trying to touch a target in the air).

The RCS has 16 items and three subscales: Cognitive Self-Regulation (7 items, including "attentive − inattentive"), Affective Self-Regulation (6 items, including "Control over emotions − Uncontrolled emotions"), and Physical/Motor Regulation (3 items, including "Coordinated − Clumsy"). Bipolar adjectives (e.g., "attentive − inattentive") are used for each item, and raters were asked to rate the child using a 7-point scale. All raters were blind to the children's intervention assignment and received training on the use of the RCS and assessing children's self-regulation (see Lakes & Hoyt, 2004 for more details on study methods). All seven observers rated all children, and ratings were aggregated across raters to enhance the generalizability of scores. As reported by Lakes and Hoyt (2004) the interrater reliability in this study was strong: when averaged across two occasions, agreement between the seven raters yielded ICC = .92, .93, and .91 for the cognitive, affective, and motor scales, respectively [relative intraclass correlation coefficients for the composite scores based on seven raters].

Prior research on the RCS described its development as well as the generalizability of scores obtained in a primary school sample (Lakes, 2012b; Lakes & Hoyt, 2004, Lakes & Hoyt, 2009). In the first published study using the RCS, Lakes and Hoyt (2004) documented the utility of the RCS as a measure of self-regulation, with the sensitivity to detect significant improvements in self-regulation following a randomized school-based exercise intervention (Taekwondo taught in physical education at a primary school). They reported results of a factor analysis indicating that the scale produced three factors, or subscales (Cognitive, Affective, and Motor Regulation), which were distinguishable, but strongly correlated. In a subsequent study examining the generalizability of scores obtained from the RCS, Lakes and Hoyt (2009) reported that RCS scores yielded strong *g* coefficients (which are similar to more common reliability coefficients, but tend to be lower because *g* coefficients simultaneously consider multiple sources of variance in ratings). In a fully-crossed research design with five raters rating 181 children, the RCS yielded the following *g* coefficients for each of the three subscales of the RCS: .86 (Cognitive), .92 (Affective), and .88 (Physical/Motor). Lakes and Hoyt (2009) also examined rater bias and occasion as potential sources of error in RCS ratings and illustrated how highly generalizable scores could be obtained by aggregating ratings over multiple raters and/or occasions. Lakes and Hoyt (2009) further reported conventional test-retest reliability coefficients for the RCS when administered two times over a four-month period; the coefficients were .64, .84, and .80 for the Cognitive, Affective, and Motor scales, respectively. Prior research has not yet examined the distribution of RCS scores in a school-wide sample or the relationship between RCS ratings and ratings on other measures of executive function or behavior.

2. The present study

In spite of the strong evidence for the generalizability and factor structure of the RCS, it is important also to examine the degree to which ratings of self-regulation in a physically active setting correspond with other measures of regulatory skills, including executive function task performance and global parent and teacher

ratings of children's abilities. Thus, several measures were selected as the standard against which to test the RCS; these included the Freedom from Distractibility Index of the Wechsler Intelligence Scale for Children (WISC III, which includes the mental math and digit span tasks) and the teacher and parent versions of the Strengths and Difficulties Questionnaire for school-age children (SDQ: Goodman, 1997). Moreover, it is important to examine the distribution of scores to evaluate the scale's effectiveness at measuring individual differences in a school population. Thus, the hypotheses for the current study were:

Hypothesis 1. Self-regulation is a psychological construct that theoretically occurs across a broad continuum in the population. As the RCS was designed to measure a full range of self-regulation (from poor to superior regulation), it was predicted that the distribution of RCS scores would approximate a normal distribution and that the mean (in a high-achieving, private primary school) would be close to, but slightly higher than, the midpoint on the scale.

Hypothesis 2. Because self-regulation is a broad construct that comprises a number of more specific constructs, such as attention, persistence, and emotional control, it was predicted that the expected associations with measures of more specific constructs that fall under the umbrella of self-regulation (e.g., executive function tasks; parent or teacher rated hyperactivity/inattention) would be significant, but modest. *Hypothesis 2a* – RCS ratings of self-regulation will correlate with performance on an executive function task. *Hypothesis 2b* – Total self-regulation will be associated with parent and teacher ratings of total difficulties. *Hypothesis 2c* – Observer-rated cognitive regulation will correlate with parent and teacher ratings of attention and hyperactivity. *Hypothesis 2d* – Observer-rated affective regulation will correlate with parent and teacher ratings of emotional problems. *Hypothesis 2e* – Observer-rated self-regulation will correlate with parent and teacher ratings of peer problems. *Hypothesis 2f* – Observer-rated cognitive and affective regulation will correlate with parent and teacher ratings of prosocial behavior. It was predicted that there would be a stronger relationship between RCS ratings and performance – based executive function measures, as both were based on task performance. Relationships with parent and teacher ratings of mental health problems and prosocial behaviors were predicted to be smaller as these constructs, while related to executive functioning or self-regulation, reflected specific categories of emotion and behavior that are affected by more than solely executive functioning capabilities.

2.1. Expected correlation coefficients

Decades of prior research have provided benchmarks that can be applied to aid in the interpretation of correlation coefficients derived from psychological assessment studies. In an extensive review of the psychological assessment literature (which included test validity research and more than 800 samples and 125 meta-analyses), Meyer et al. (2001) reported observed correlation coefficients that provide important context for the present study. They summarized 22 examples (from a much larger sample of studies) of cross-method convergent associations in research with children and adolescents; observed *r*'s ranged from .03 to .42. In research using the same executive function task used in the present study [WISC Freedom from Distractibility (FD) Index], relationships between the FD and teacher and parent ratings of attention yielded *r*'s of .10 and .03, respectively. Correlations between parent and other observer (including teachers and clinicians) ratings ranged from .16 to .42; in most of these examples, raters were using the same rating scale. Correlations between raters using different instruments designed to measure the same construct would be expected to be at the low end of this range. Piek et al.'s (2004) research examined the relationship between motor abilities (using the McCarron Assessment of Neuromuscular Development in which a high score reflects greater motor abilities) and executive functioning (in which lower scores reflect faster response times) and reported correlation coefficients ranging from about 0 to −.26; statistically significant coefficients ranged in size from −.13 to −.26 and were described as evidence of a "large" association between motor abilities and executive functioning. Thus, anticipated correlation coefficients for the associations between the RCS scales and FD tasks were expected to approximate those observed in prior research (i.e., an *r* between .13 and .26 for the relationship between motor regulation and tests of executive function and an *r* between .10 and .15 for the relationship between observer-rated cognitive and affective functioning and tests of executive function).

3. Method

3.1. Participants

Participants in this study were children who participated in a prior randomized intervention study (Lakes & Hoyt, 2004). An entire private primary school (Kindergarten through 5th grade) in the Midwestern United States participated in the intervention, and parents for all but one child consented to the child's participation in the research (yielding $N = 207$). Slightly more than half (51%) of the participants were female. Eighty-three percent of participants were Caucasian, and 73% were from families earning more than $100,000 per year. Assessments were conducted two times: the first week of the academic year (Time 1) and four months later (Time 2). Some students moved during the school year or were absent during one of the assessment periods; as a result, N's for the analyses vary and are reported in tables and figures.

3.2. Cross-method convergent validity instruments

3.2.1. Freedom from distractibility (FD)

This measure is a factor analytically-derived subscale of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) that includes the Arithmetic and Digit Span subtests. On both subtests, items increase in difficulty, and scores are calculated based on whether or not a response is correct or incorrect. There is evidence that the FD subscale focuses on a factor that has attention-concentration aspects (Kaufman, 1975), as well as additional executive functioning abilities (e.g., working memory, cognitive flexibility). The FD subscale is also thought to tap into sequencing ability (Bannatyne, 1974), short-term memory and auditory memory (Cohen, 1957), and executive processes (Wielkiewicz, 1990). In the sample used for this study, the correlation between the Digit Span and Arithmetic subtest scores for children in Kindergarten through fifth grade was $r = .63$. This measure was administered by evaluators who were trained using standard WISC-III training procedures and guidelines.

3.2.2. Strengths and Difficulties Questionnaire (SDQ)

The SDQ (Goodman, 1997) is a brief (25-item) rating scale that has five subscales: Emotional Symptoms, Conduct Problems, Inattention/Hyperactivity, Peer Problems, and Prosocial Behavior. The Inattention/Hyperactivity subscale includes items related to self-regulation, such as "Restless, overactive, cannot stay still for long" (Inattention/Hyperactivity). Items on the Prosocial subscale address positive behaviors that reflect self-regulation in social contexts, such as "Shares readily with other children." The Emotional Problems, Conduct Problems, and Peer Problems subscales contain items that address

Table 1
Descriptive statistics and the Shapiro–Wilk test of normality.

	Range of scores	Mean (*SD*)	Skewness	Kurtosis	*W* (*p*)
RCS scales (n = 178)					
Cognitive regulation	3.7–6.8	5.6 (.54)	−.40	.52	.99 (.07)
Affective regulation	3.3–6.9	5.3 (.74)	−.14	−.30	.99 (.64)
Motor regulation	2.4–6.9	5.2 (.84)	−.48	.15	.98 (.02)*
Total self-regulation	11.5–20.6	16.1 (2.0)	−.09	−.51	.99 (.39)
Executive function tasks (n = 178)					
Math	3.0–24.0	13.4 (4.2)	−.19	.14	.98 (.02)*
Digit span	4.0–24.0	12.5 (3.4)	.44	.76	.98 (.00)**
FD total	8.0–43.0	25.9 (6.9)	−.18	−.04	.99 (.53)
Teacher SDQ scales (n = 194)					
Emotional symptoms	.0–8.0	.96 (1.4)	1.98	4.91	.75 (.00)**
Conduct problems	.0–5.5	.69 (1.1)	2.06	4.35	.72 (.00)**
Hyperactivity/inattention	.0–9.5	2.53 (2.4)	1.02	.36	.88 (.00)**
Peer problems	.0–6.5	1.04 (1.2)	1.43	2.21	.84 (.00)**
Prosocial behavior	2.0–10.0	7.49 (1.8)	−.57	−.01	.93 (.00)**
Total difficulties	.0–21.5	5.22 (4.4)	1.07	.86	.91 (.00)**
Parent SDQ scales (n = 67)					
Emotional symptoms	.0–6.0	1.17 (1.4)	1.38	1.62	.80 (.00)**
Conduct problems	.0–4.0	.72 (.99)	1.40	1.33	.74 (.00)**
Hyperactivity/inattention	.0–9.0	2.02 (2.1)	1.33	1.85	.84 (.00)**
Peer problems	.0–4.0	.94 (1.1)	1.04	.31	.81 (.00)**
Prosocial behavior	4.0–10.0	8.69 (1.7)	−1.19	.39	.78 (.00)**
Total difficulties	.0–19.0	4.86 (3.8)	.96	1.48	.93 (.00)**

Note. SD = Standard deviation.
*$p < .05$, **$p < .01$.

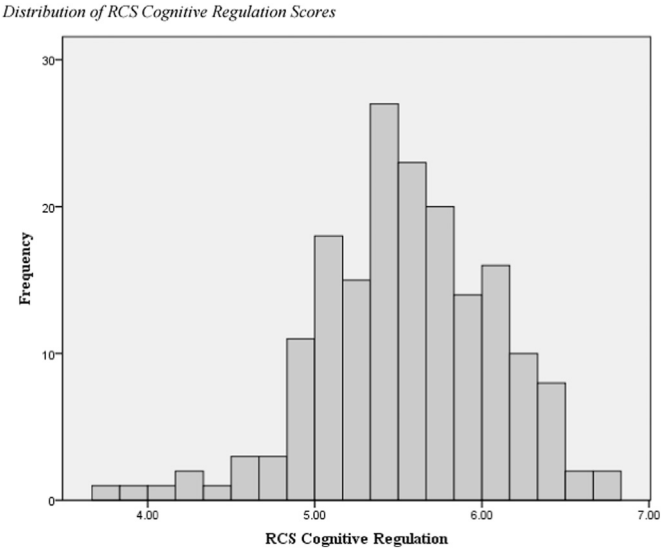


Fig. 1. Distribution of RCS Cognitive Regulation Scores. *Note.* $N = 178$. RCS Cognitive regulation scores are the average across items and raters on the Cognitive subscale. Possible scores range from 1 to 7. Rating Time 1.

problems that are likely affected by self-regulatory abilities, but are also affected by other factors. Items on the Emotional Problems subscale address depressive and anxiety symptoms (e.g., "Often complains of headaches, stomachaches"; "Many worries, often seems worried"). Items on the Peer Problems subscale include "Picked on or bullied by other children" and "Rather solitary, tends to play alone." Items on the Conduct Problems scale assess behaviors such as lying, cheating, and stealing; as this scale was not hypothesized to relate closely to observations of self-regulation, this was the only scale not considered for convergent validity analyses. The Total Difficulties score is derived from all subscales except the Prosocial Behavior subscale. The 25 items that comprise the SDQ are rated on a three-point scale (0 = not true, 1 = somewhat true, and 2 = certainly true).

In this sample, internal consistency reliabilities (Cronbach's alphas) for scores obtained on the SDQ-Parent and SDQ-Teacher versions were .85 and .85 (Total Scale), .84 and .88 (Inattention/Hyperactivity), .71 and .80 (Emotional Problems), .61 and .69 (Conduct Problems), .64 and .69 (Peer Problems), and .61 and .85 (Prosocial Behavior), respectively (Lakes & Hoyt, 2004). Teacher ratings on both the SDQ Conduct and Hyperactivity subscales are based on scores that aggregate two teacher ratings per child, in order to minimize variance due to teacher bias. As Lakes and Hoyt reported previously, agreement between the two teachers was limited and yielded a mean ICC across two rating occasions of .19, .64, .71, .33, and .16 for the emotional, conduct, hyperactivity/inattention, peer problems, and prosocial scales, respectively. Thus, while there was adequate consensus between teachers when rating hyperactivity and inattention, there was little agreement for most other subscales. Parent ratings are based on a single rating by one parent, as for most children, only one parent completed and returned the research forms.

3.3. Analyses

SPSS 17 was used for all analyses. Data was first analyzed for outliers; four outliers were detected and removed from subsequent analyses. Descriptive statistics were calculated using all data from Time 1 (the baseline assessment prior to intervention), and the distributions of scores for all measures were examined and tested using the Shapiro–Wilk test of normality. Spearman's rho correlations (1-tailed) were computed to examine relationships between scores derived from the RCS, executive function tasks, and teacher and parent ratings. Correlations between executive function tasks and observer ratings (RCS) were computed for both assessments (Time 1 and Time 2). Correlations with the SDQ were computed for Time 2 only, as teacher ratings were more reliable after teachers had interacted with students for four months (versus during the first week of school when teachers' experiences with students were based on more limited observations). Parent SDQ analyses were also based on data collected at Time 2 to ensure that teacher and parent ratings occurred over the same time period.

4. Results

4.1. Hypothesis 1: in a school-wide sample, RCS scores will be normally distributed, with a mean slightly higher than the midpoint in this high-achieving private school sample.

Means for all three RCS scales were slightly higher than the midpoint on the scale. Table 1 reports descriptive and psychometric results, including statistics for skewness and kurtosis. These statistics fell within the normal range for RCS Cognitive (Fig. 1), Affective (Fig. 2), and Total Self-Regulation (Fig. 4) scores. Contrary to our hypothesis, the distribution of scores for the Motor regulation scale was not normal (Fig. 3). The distributions for scores obtained on two measures of executive function (mental math and freedom from distractibility) were normal, and the distribution for the third (digit span) was abnormal. Distributions for all of the parent or teacher rated scales were abnormal.

4.2. Hypothesis 2: convergent validity

Hypothesis 2a – RCS ratings of self-regulation will correlate with performance on an executive function task. As predicted, across both assessment occasions, executive function task performance was significantly correlated with RCS ratings of self-regulation (Table 2). Only one correlation (motor regulation and

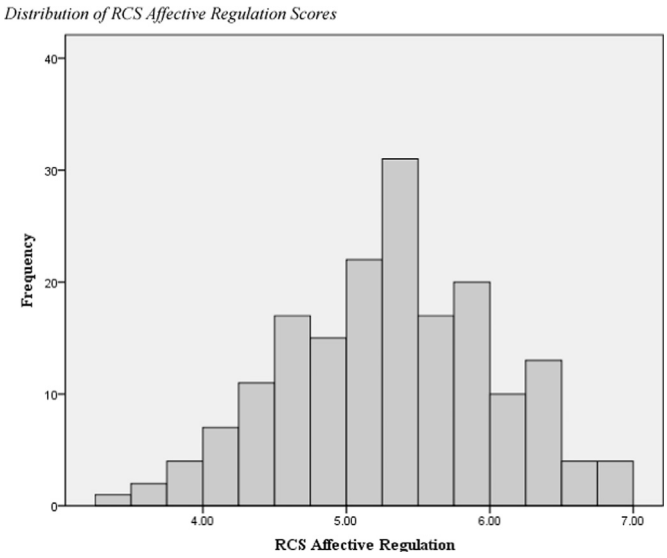


Fig. 2. Distribution of RCS Affective Regulation Scores. *Note.* $N = 178$. RCS Affective regulation scores are the average across items and raters on the Affective subscale. Possible scores range from 1 to 7. Rating Time 1.

digit span at Time 2) failed to achieve statistical significance ($r = .08$, $p = .14$).

Hypothesis 2b – total self-regulation will correlate with parent and teacher ratings of total difficulties. As predicted and reported in Table 3, there were significant, negative correlations between parent and teacher ratings of Total Difficulties and Total Self-Regulation (r's = –.21 and .14, p's = .04 and .02, respectively).

Hypothesis 2c – observer-rated cognitive regulation will correlate with parent and teacher ratings of attention and hyperactivity. As reported in Table 3, RCS Cognitive subscale scores were negatively correlated with teacher ratings of Hyperactivity/Inattention ($r = –.13$, $p = .04$). The correlation between RCS Cognitive subscale scores and parent ratings was not statistically significant ($r = –.11$, $p = .19$).

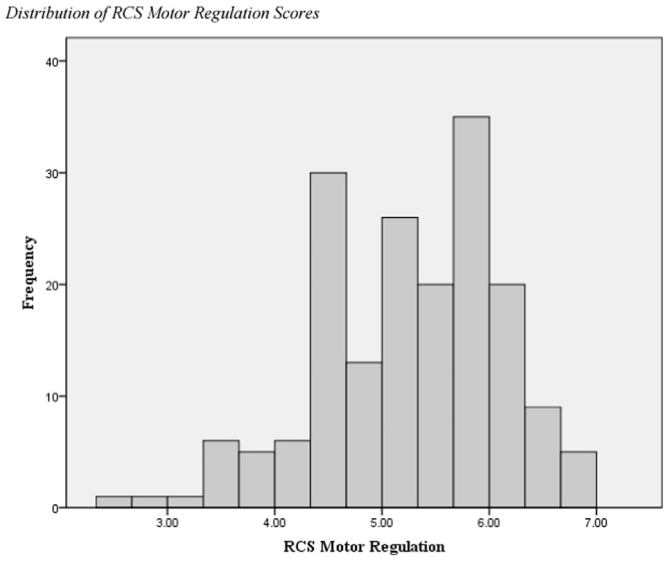


Fig. 3. Distribution of RCS Motor Regulation Scores. *Note.* $N = 178$. RCS Physical/Motor regulation scores are the average across items and raters on the Physical/Motor subscale. Possible scores range from 1 to 7. Rating Time 1.

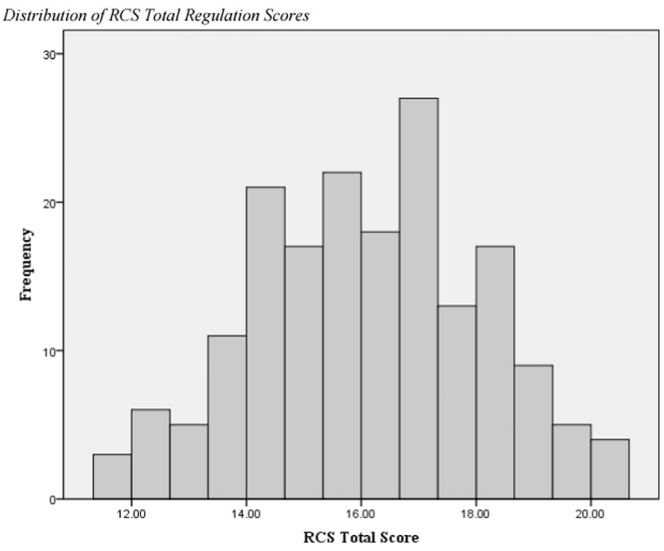


Fig. 4. Distribution of RCS Total Regulation Scores. *Note.* $N = 178$. RCS Total regulation scores are the total of 3 subscale scores. Possible scores range from 3 to 21. Rating Time 1.

Hypothesis 2d – observer-rated affective regulation will correlate with parent and teacher ratings of emotional problems. Correlations between parent and teacher ratings of emotional problems and affective regulation were not significant (Table 3). The relationship with teacher ratings was close to 0 ($r = .02$, $p = .42$), while the relationship between affective regulation and parent ratings of emotional problems approached significance ($r = –.21$, $p = .052$).

Hypothesis 2e – observer-rated self-regulation will correlate with parent and teacher ratings of peer problems. As reported in Table 3, there were significant, negative correlations between teacher ratings of peer problems and observer-rated affective regulation (e.g., the relationship between a child's Total RCS score and teacher rated peer problems yielded $r = –.16$, $p = .013$). The relationship between parent rated peer problems and the Total self-regulation scores was equal in size, but was not significant ($r = –.16$, $p = .105$).

Table 2
Correlations between executive function tasks and observer, teacher, and parent ratings.

	Executive function tasks					
	FD total		Mental math		Digit span	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
RCS ratings ($n = 176$)						
Cognitive regulation	**.35****	**.38****	**.39****	**.41****	**.26****	**.27****
Affective regulation	**.35****	**.25****	**.38****	**.30****	**.26****	**.13***
Motor regulation	**.17****	**.13***	**.20****	**.16***	**.12***	.08
Total self-regulation	**.28****	**.26****	**.32****	**.29****	**.21****	**.16***
Teacher ratings ($n = 181$)						
Emotional symptoms	–.07	–.09	–.06	–.07	–.10	–.10
Conduct problems	–.08	–.07	–.06	–.13*	–.11	–.13*
Hyperactivity/inattention	**–.24****	**–.18****	**–.20****	**–.22****	**–.26****	**–.24****
Peer problems	**.18****	**.23****	**.18****	**.12***	**.14***	**.20****
Prosocial behavior	–.10	–.07	–.11	.08	–.05	.02
Total difficulties	**–.13***	–.10	–.11	**–.14***	**–.16***	**–.15***
Parent ratings ($n = 60$)						
Emotional symptoms	–.06	.08	–.01	.18	–.10	.03
Conduct problems	–.07	–.12	–.07	–.02	–.05	–.12
Hyperactivity/inattention	–.18	.03	**–.22***	–.19	–.14	.12
Peer problems	–.09	–.04	–.08	.06	–.05	–.07
Prosocial behavior	.18	.13	.15	.03	.16	.14
Total difficulties	–.15	–.03	–.13	–.05	–.15	.01

Significant findings supporting hypotheses are indicated in bold.
**$p < .01$ *$p < .05$.

Table 3
RCS correlations with teacher and parent SDQ ratings.

	RCS ratings			
	Cognitive regulation	Affective regulation	Motor regulation	Total self-regulation
Teacher SDQ ratings ($n = 185$)				
Emotional	−.03	.02	.04	.03
Hyperactivity/inattention	**−.13***	.04	−.08	−.06
Conduct problems	−.11+	.02	−.05	−.03
Peer problems	−.10+	**−.12***	**−.20****	**−.16***
Prosocial behavior	−.04	**−.13***	.01	−.06
Total difficulties	**−.19****	−.05	**−.17***	**−.14***
Parent SDQ ratings ($n = 63$)				
Emotional	−.13	−.21+	−.10	−.14
Hyperactivity/inattention	−.11	−.12	−.05	−.05
Conduct problems	−.21+	−.20+	−.18+	−.20+
Peer problems	−.15	−.16	−.18+	−.16
Prosocial behavior	**.27***	**.28***	.09	.21+
Total difficulties	**−.23***	−.12	−.19+	**−.21***

Significant findings supporting hypotheses are indicated in bold.
Note. Reports results from data collected at Time 2.
$^{**}p < .01$ $^{*}p < .05 + p < .10$.

Hypothesis 2f − observer-rated cognitive and affective regulation will correlate with parent and teacher ratings of prosocial behavior. RCS Cognitive and Affective subscale scores were positively correlated with parent ratings of children's prosocial behavior (Table 3), indicating that children with higher scores on the RCS Cognitive and Affective subscales were rated by their parents as exhibiting more prosocial behaviors (r's = .27 and .28, p's = .016 and .012 for the cognitive and affective scales, respectively). There was no significant relationship between teacher ratings of prosocial behavior and RCS ratings of cognitive and motor regulation; however, teacher-rated prosocial behavior was unexpectedly negatively related to affective regulation ($r = -.13$ $p = .05$).

5. Discussion

In a school-wide sample of children, RCS Affective and Cognitive Regulation scores were normally distributed. This finding was particularly noteworthy given the restricted variance in the sample that was evidenced in parent and teacher ratings. The ability of an instrument to produce a normal distribution in a school-wide sample is an important consideration in instrument selection for at least four reasons. Cohen and Cohen (1983) described two reasons, including "the opportunity for valid discrimination of individual differences" and avoiding the "risk of underestimating the relationship between the conceptual variables" (p. 66). A third reason is that the lack of sufficient variance among children in a sample attenuates the generalizability of scores (Lakes, 2012c). Moreover, skewed measures decrease inter-item correlations and alpha coefficients (Greer, Dunlap, Hunter, & Berman, 2006). Thus, normally distributed scores are important for research. Although the distributions for Cognitive, Affection, and Total Regulation were normally distributed, the distribution for Motor Regulation scores was not; it is possible that in the sample studied for this research, there were fewer individual differences in motor skill or that the scale did not have the sensitivity to measure subtle individual differences. These possibilities should be investigated further in future research.

5.1. Cross-method convergent validity: associations between the RCS and tests of executive function

Associations between the RCS and tests of executive function were significant and in the expected range; results indicated that children rated by observers as exhibiting greater self-regulation scored higher on tests of executive function. According to empirical guidelines for interpreting correlation coefficients (Hemphill, 2003), the magnitude of most of these observed correlations fell in the middle third ($r > .20$) or upper third ($r > .30$) of observed correlations in psychological research. The observed associations were strong when compared to prior research using the same tasks in conjunction with parent or teacher ratings: Meyer et al. (2001) reported that relationships between the task (FD) and teacher and parent ratings of attention yielded r's of .10 and .03, respectively. In the present study, the association between a child's FD score and teacher-rated hyperactivity/inattention ranged from $r = -.18$ to $-.24$; similar to the findings reported by Meyer et al., the relationship between a child's FD score and parent-rated hyperactivity/inattention ranged from $r = .03$ to $-.18$ (neither was significant in the present study). Meyer et al. did not report prior studies examining the relationship between research observer ratings of attention and FD scores; in the present study, the observed associations between observer-rated cognitive regulation and FD scores on two occasions about four months apart were larger (r's = .35 and .38, $p < .01$) than the relationships between teacher ratings and the FD. Moreover, the associations between scores for Motor Regulation and executive function tasks were similar in size and significance to those reported by Piek et al. (2004). These results support the validity of scores obtained using the RCS.

5.2. Cross-method convergent validity: associations between the RCS and the SDQ

Children rated on the RCS as exhibiting greater cognitive and total self-regulation were rated by teachers and parents as having fewer difficulties. Teacher ratings of children's total difficulties were also associated with observer ratings of motor regulation. Children with higher ratings on the RCS Cognitive subscale tended to have lower scores on SDQ Teacher ratings of hyperactivity and inattention. Moreover, there were significant, negative correlations between teacher ratings of peer problems and observer-rated affective, motor, and total self-regulation, indicating that when observer-rated self-regulation in these domains was high, teacher-reported peer problems were low. RCS Cognitive and Affective subscale scores were positively correlated with parent ratings of children's prosocial behavior, indicating that children with higher scores on the RCS Cognitive and Affective subscales were rated by their parents as exhibiting more prosocial behaviors.

There was an unexpected finding indicating a significant, but small negative correlation between teacher-rated prosocial behavior and observer-rated affective regulation. This finding indicates that students rated higher by observers in affective regulation were rated lower in prosocial behavior by teachers. The direction of the relationship was opposite the relationship observed between parent ratings of prosocial behavior and observer-rated affective regulation. It is possible that teachers differed in their interpretation of items on this particular subscale or perhaps had limited opportunity to observe the behaviors assessed; this hypothesis is supported by the fact that the agreement between teachers was weakest for this subscale as noted previously ($ICC = .16$). In fact, teacher agreement was so low on this particular scale, that it raises serious questions about the reliability of teacher ratings of prosocial behavior in this particular study, and the finding should be interpreted with caution. Future studies should further address this issue, perhaps using an alternative measure of prosocial behavior.

5.3. Post-hoc analyses

Although the size and significance of most correlations between the RCS and SDQ parent and teacher ratings were within the expected range (see Meyer et al., 2001), post-hoc analyses were conducted to further aid in interpreting findings. It is possible that because the different assessment situations (observer-rated performance on a challenge course, parent ratings based on overall experience with child, and teacher ratings based on child's classroom behavior) measure a child's behavior in different contexts, variability in a child's behavior in a given context may contribute to the lack of correlation between measures. As Table 4 indicates, there was no agreement between teacher and parent ratings of behavioral and emotional problems; in fact, the only significant correlation (between teacher and parent rated total difficulties) was negative, indicating that children rated as having more difficulties by teachers were rated as having fewer difficulties by parents. Given the striking lack of agreement between parent and teacher scores on the same questionnaire (SDQ), it is plausible that some combination of rater bias and contextual differences in children's behavior impacted SDQ scores, thereby affecting the size and significance of correlations with the RCS. Poor agreement between parent and teacher ratings of children's behavior is not unique to this study; as described by Meyer et al. (2001), across numerous prior studies, agreement between teachers and parents has produced *r*'s in the range of .13–.29. Moreover, findings in the current study are limited by the low response rate among parents; perhaps parents who returned the questionnaires differed in some respect from the full potential sample of parents (for example, it is possible that parents who returned the questionnaire were the most likely to underestimate their child's difficulties and to present them in the most favorable manner possible and that this tendency prompted them to complete and return the forms, which would explain the negative relationships between parent and teacher ratings).

Another likely explanation for the lack of correlation between teacher and parent ratings (SDQ) and RCS scores is that due to the lack of variance in SDQ ratings, there could be little covariance (i.e., only very weak validity coefficients were possible). Table 1 shows that parent and teacher ratings were significantly skewed, with parents and teachers tending to rate children very positively. On the SDQ Conduct subscale, teachers rated 62% of children between a 0 and a 1, and parents rated 85% of children between a 0 and a 1 on a 10-point scale. On the Hyperactivity subscale, parents and teachers rated 50% of children between a 0 and a 1 on a 10-point scale. It is possible that these consistently positive ratings are reflective of characteristics and circumstances associated with the high-achieving, socioeconomically advantaged sample used for this study. Few behavioral problems were noted, and most children tended to be viewed very favorably by parents and teachers. Although scores for the RCS were normally distributed, scores for the SDQ were severely skewed, and strong correlations between the two measures were unlikely to be obtained. Therefore, it is not surprising that the internal consistencies reported for the SDQ were also low. The scale with one of the highest alphas (Total Difficulties) yielded one of the few consistently significant, observed relationships between the SDQ teacher and parent ratings and the RCS.

5.4. Limitations and directions for future research

Finding an ideal criterion measure for validating RCS scores is difficult. Parent and teacher ratings are beneficial as they are global and relevant to real life outcomes, but they are not perfectly valid measures of the underlying constructs themselves. The weak associations between the RCS and parent and teacher ratings is due, at least in part, to the criterion measure itself and differences in the constructs assessed by each measure. Moreover, interpretation of the findings is limited by the fact that fewer parents returned the questionnaire, and as a result, parent ratings were available for a limited number of children. In future research, a distributional parent or teacher rated scale of attention and behavior could be administered along with the RCS to address the limitations of the SDQ in this study, and parent incentives could be offered to increase the response rate. In addition, future studies could examine the relationship between the RCS and other established measures of motor abilities and experimental tasks (such as computerized tests of executive function) used to measure self-regulation. Moreover, future research could examine the predictive validity of the RCS, which was beyond the scope of the current study.

5.5. Conclusion and practical implications

This research provides preliminary support for the validity of self-regulation scores derived from observation of children in a physically active context using the RCS. There is potential for this type of measurement to enhance methods in studies examining the relationship between mental health and physical activity. For example, researchers investigating the impact of physical activity and exercise interventions on children's motor skills and executive functioning may find that the RCS is a useful supplement to parent and teacher reports and executive function tasks. The RCS allows researchers to assess motor, cognitive, and affective regulation with one instrument, and when used in a rating design where all raters rate all children, researchers can control for rater bias, which is not possible to do in research designs dependent on parent or teacher ratings. While it will be important to continue to collect data using common measures of mental health and cognitive abilities (i.e., symptom rating scales, performance tasks), the addition of standardized observations of children during participation in physical activity contributes to multi-method, multisource research design and will enhance the study of physical activity and mental health.

Table 4
Correlations between the teacher and parent SDQ ratings.

Parent ratings	Teacher ratings					
	Emotional problems	Conduct problems	Hyperactivity inattention	Peer problems	Prosocial behavior	Total difficulties
Emotional problems	.01	–	–	–	–	–
Conduct problems	–	–.16	–	–	–	–
Hyperactivity	–	–	–.16	–	–	–
Peer problems	–	–	–	–.17+	–	–
Prosocial behavior	–	–	–	–	–.14	–
Total difficulties	–	–	–	–	–	**–.22***

Significant findings supporting hypotheses are indicated in bold.
Note. SDQ = Strengths and Difficulties Questionnaire. $N = 65$. Reports results from data collected at Time 2.
*$p < .05$ + $p < .10$.

References

Bannatyne, A. (1974). Diagnosis: a note on recategorization of the WISC scaled scores. *Journal of Learning Disabilities, 7*, 272–274.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*, 899–911.

Budde, H., Voelcker-Rehage, C., PietraByk-Kendziorra, S., Ribeiro, P., & Tidow, G. (2008). Acute coordinative exercise improves attentional performance in adolescents. *Neuroscience Letters, 441*, 219–223.

Cohen, J. (1957). The factorial structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology, 21*, 283–290.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Davis, C. L., Tomporowski, P. D., Boyle, C. A., Waller, J. L., Miller, P. H., Naglieri, J. A., et al. (2007). Effects of aerobic exercise on overweight children's cognitive functioning: a randomized controlled trial. *Research Quarterly for Exercise and Sport, 78*, 510–519.

Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Development, 71*, 44–56.

Diamond, A. (2013). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science, 21*, 335–341.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*(6045), 959–964.

Gerber, M., Kalak, N., Lemola, S., Clough, P. J., Puhse, U., Elliot, C., et al. (2012). Adolescents' exercise and physical activity are associated with mental toughness. *Mental Health and Physical Activity, 5*, 35–42.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Greer, T., Dunlap, W. P., Hunter, S. T., & Berman, M. E. (2006). Skew and internal consistency. *Journal of Applied Psychology, 91*, 1351–1358.

Hemphill, J. F. (2003). Interpreting the magnitude of correlation coefficients. *American Psychologist, 58*, 78–80.

Holmbeck, G. N., Li, S. T., Schurman, J. V., Friedman, D., & Coakley, R. M. (2002). Collecting and managing multisource and multimethod data in studies of pediatric populations. *Journal of Pediatric Psychology, 27*, 5–18.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology, 43*, 135–147.

Lakes, K. D. (2012a). Psychological and educational assessment in child and adolescent research: introduction to the special issue. *The International Journal of Educational and Psychological Assessment, 10*, 1–5.

Lakes, K. D. (2012b). The Response to Challenge Scale (RCS): the development and construct validity of an observer-rated measure of children's self-regulation. *The International Journal of Educational and Psychological Assessment, 10*, 83–96.

Lakes, K. D. (2012c). Restricted sample variance reduces generalizability. *Psychological Assessment, 25*, 643–650.

Lakes, K. D., & Hoyt, W. T. (2004). Promoting self-regulation through school-based martial arts training. *Journal of Applied Developmental Psychology, 25*, 283–302.

Lakes, K. D., & Hoyt, W. T. (2009). Applications of generalizability theory to clinical child and adolescent psychology research. *Journal of Clinical Child and Adolescent Psychology, 38*, 1–22.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: a review of evidence and issues. *American Psychologist, 56*, 128–165.

Moffit, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *PNAS, 108*, 2693–2698.

Piek, J. P., Dyck, M. J., Nieman, A., Anderson, M., Hay, D., Smith, L. M., et al. (2004). The relationship between motor coordination, executive functioning, and attention in school aged children. *Archives of Clinical Neuropsychology, 19*, 1063–1076.

Tomporowski, P. D., Lambourne, K., & Okumura, M. S. (2011). Physical activity interventions and children's mental function: an introduction and overview. *Preventive Medicine, 52*, 53–59.

Tomporowski, P. D., Miller, P. H., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Psychology Review, 20*, 111–131.

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R third factor: it's more than distractibility. Psychological Assessment. *A Journal of Consulting and Clinical Psychology, 2*, 91–97.