



# Advancing Educational Research on Children's Self-Regulation With Observational Measures

Journal of Psychoeducational Assessment  
2023, Vol. 41(3) 267–282  
© The Author(s) 2022



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/07342829221143208  
[journals.sagepub.com/home/jpa](https://journals.sagepub.com/home/jpa)



Janina Eberhart<sup>1,†</sup> , Andrew E. Koepp<sup>2,†</sup>, Steven J. Howard<sup>3</sup> ,  
Rianne Kok<sup>4</sup>, Dana C. McCoy<sup>5</sup>, and Sara T. Baker<sup>6</sup>

## Abstract

Self-regulation is crucial for children's development and learning. Almost by convention, it is assumed that self-regulation is a relatively stable skill, and little is known about its dynamic nature and context dependency. Traditional measurement approaches such as single direct assessments and adult reports are not well suited to address questions around variations of self-regulation within individuals and influences from social-contextual factors. Measures relying on child observations are uniquely positioned to address these questions and to advance the field by shedding light on self-regulatory variability and incremental growth. In this paper, we review traditional measurement approaches (direct assessments and adult reports) and recently developed observational measures. We discuss which questions observational measures are best suited to address and why traditional measurement approaches fall short. Finally, we share lessons learned based on our experiences using child observations in educational settings and discuss how measurement approaches should be carefully aligned to the research questions.

## Keywords

self-regulation, observational measures, intra-individual differences, children, direct assessments

---

<sup>1</sup>University of Tübingen, Germany

<sup>2</sup>The University of Texas at Austin, TX, USA

<sup>3</sup>University of Wollongong, NSW, Australia

<sup>4</sup>Erasmus University Rotterdam, Rotterdam, Netherlands

<sup>5</sup>Harvard Graduate School of Education, Cambridge, MA, USA

<sup>6</sup>University of Cambridge, UK

†These authors contributed equally to this work.

## Corresponding Author:

Janina Eberhart, University of Tübingen, Schleichstraße 4, Tübingen 72076, Germany.

Email: [Janina.eberhart@uni-tuebingen.de](mailto:Janina.eberhart@uni-tuebingen.de)

## Introduction

Decades of research leave little doubt that individual differences in young children's self-regulation predict later learning and health (e.g., [Moffitt et al., 2011](#); [Robson et al., 2020](#)). However, individual children may display a range of self-regulation depending on the context and the supports available to them. Little work has examined this intra-individual variability, despite theories that conceptualize self-regulation as dynamic and continually influenced by the environment ([Cole et al., 2019](#); [Nigg, 2017](#)). Indeed, self-regulation is thought to emerge early in life through continual inputs from adults in the environment that support co-regulation until children can appropriately regulate on their own ([Sameroff, 2010](#); [Vygotsky, 1978](#)). Children's successful regulation during the early childhood period likely varies depending on the presence or lack of these supports, but this area remains understudied, perhaps due to a historical lack of appropriate measures ([McCoy, 2019](#)). Emerging observational measures that record children's behavior in their everyday environments are uniquely positioned to address empirical questions about intra-individual variability and the social-contextual influences on children's regulation. Investigating these questions is critical to improve our understanding of how to support the development of self-regulation in everyday contexts and to advance applied research.

In this paper, we briefly review the most commonly used measures of self-regulation (i.e., direct assessments and adult reports) and argue that they are best suited to research questions about individual differences and intra-individual change over time. We then review observational approaches and describe the investigations of intra-individual variability that they make possible. To facilitate the wider adoption of these observational methods, we describe our experiences using them, lessons learned, and advice for researchers who would also like to use them. We conclude by noting the specific investigations that each measurement approach is best-suited to address, arguing that observational measures are the best choice to understand intra-individual differences and social-contextual influences on self-regulation.

## Common Approaches to Measuring Self-Regulation

Self-regulation is the ability to flexibly adapt thoughts, behaviors, and emotions to the constantly changing demands of the environment ([Nigg, 2017](#)). It is important in school contexts and can help children to function in social situations ([Blair & Raver, 2015](#); [McClelland & Cameron, 2012](#)). Self-regulation involves the dynamic interplay of "bottom-up" reactivity and "top-down" processes of control ([Liew et al., 2019](#); [Nigg, 2017](#); [Ursache et al., 2012](#)). Children display this bottom-up reactivity when they respond to a situation on impulse or without thinking about it—in excitement about knowing the answer they may blurt it out instead of raising their hand, or their attention may shift to a distraction in the classroom rather than remaining focused on the activity at hand. Top-down processes can be engaged in an attempt to control responses to a reactive impulse and instead respond in a controlled, intentional way that is adaptive to the environment—for example, they raise their hand to answer a question despite their excitement and desire to share their thought immediately. Children's self-regulation is a dynamic interplay of the "bottom-up" and "top-down" process and an internal process that cannot be directly observed. To study it, researchers have developed paradigms to interpret how observed behaviors indicate the internal process of regulation ([Rabbitt, 1997](#)). The most commonly used of these paradigms are direct assessments and adult reports ([Duckworth & Yeager, 2015](#); [Obradović et al., 2018](#)).

## Direct Assessments

Direct assessments employ structured tasks that challenge a self-regulatory skill (e.g., inhibitory control) and measure participant performance. For example, the Head-Toes-Knees-Shoulders task (HTKS; Cameron Ponitz et al., 2009) asks children to reverse simple, intuitive instructions (e.g., “if I say touch your toes, I want you to touch your head”), requiring them to inhibit a prepotent response to instead direct their behaviors toward the requirements of the task (Cameron Ponitz et al., 2009). These tasks are typically administered in research labs or lab-like environments, such as a quiet room in children’s schools (Obradović et al., 2018). The well-defined demands of the task are intended to allow researchers to interpret participants’ responses as the successful or unsuccessful application of a specific skill or set of skills (Duckworth & Yeager, 2015). Direct assessments employ standardized procedures and are thus suitable for understanding differences between individuals (McCoy, 2019; Toplak et al., 2013), including how these differences predict development in other areas such as children’s academic achievement (Best et al., 2011; Blair et al., 2015; Blair & Razza, 2007; von Suchodoletz et al., 2015).

Although direct assessments allow researchers to study specific skills under controlled conditions with relative conceptual precision, they often require children to demonstrate skills in isolation (e.g., inhibitory control or shifting) and to demonstrate them out of context (McCoy, 2019). In fact, direct assessments aim to eliminate variability in social-contextual factors to decrease noise and increase comparability across assessments and across children (Obradović et al., 2018; Toplak et al., 2013). The same structure that makes the task standardized and objective makes it different from the everyday environments in which children must routinely engage these skills (e.g., Howard et al., 2021; McCoy et al., 2017). When researchers have found ways to make the assessment context more like children’s everyday settings, these assessments have been more predictive of academic results. For example, direct assessments administered in group settings in children’s regular classrooms were more predictive of their academic achievement than direct assessments applied to the same children in quiet, one-on-one settings (Obradović et al., 2018). Although children’s performance on direct assessments of self-regulation predicts performance on academic tasks, it is less clear whether these direct assessments index children’s everyday self-regulatory behaviors, such as following classroom rules or sharing with peers (e.g., Ahmed et al., 2021; Graziano et al., 2015; Jones et al., 2016; Koepp et al., 2021; Tamm & Peugh, 2019). Direct assessments are thus generally low in ecological validity (Obradović et al., 2018) and not ideal for examining social–contextual influences on self-regulation or for understanding children’s responding in everyday contexts.

## Adult Reports

One way that researchers can capture children’s behavior in everyday environments is via adult reports. These reports use responses from an adult who knows the child well, typically a parent or teacher, to index children’s self-regulatory behaviors in real-world contexts on rating scales (e.g., child has trouble with multistep activities; child finishes tasks/activities too fast; Gioia et al., 2003). A strength of adult reports is that they are ecologically comprehensive, describing children’s behavior across a variety of situations in which the adult observes the child (McCoy, 2019). They are also practical for many research applications because they tend to be quick to complete and inexpensive to use (McCoy, 2019).

However, there are also drawbacks to adult reports, which are often limitations of rating scales in general. Although their focus on behavior in everyday settings makes them ecologically valid, these reports are subject to bias (Garcia et al., 2018; McCoy et al., 2017). Teachers and parents must apply a reference frame to assess the frequency, extent or typicality of children’s skills,

leading to systematic variation across raters (Duckworth & Yeager, 2015). For example, teachers in a classroom where most children struggle to follow classroom rules might rate a child with moderate self-regulatory skills higher than would teachers of predominantly highly regulated children. Parents' ratings might be particularly biased as they lack a broad reference frame to compare their child's behavior to other children. This might explain why teacher reports are often more predictive of children's academic achievement than parent reports (e.g., Blair & Razza, 2007; von Suchodoletz et al., 2015).

Adult reports typically measure behaviors holistically, making it impossible to discern specific self-regulatory competencies underlying these skills (e.g., inhibitory control or working memory). Furthermore, teachers' or parents' ratings of self-regulatory skills may be conflated with non-regulatory skills (e.g., compliance) and child performance in other areas (e.g., academic performance; Obradović et al., 2018). Adult reports therefore may not be suitable for accurately and sensitively indexing child self-regulatory skills, beyond providing a rough estimate of children's within-context developmental position or rank in this ability.

Finally, as adult reports describe children's typical behavior over a reference period rather than behavior at a particular point in time, adult reports cannot capture short-term intra-individual differences in children's regulation arising from specific social-contextual factors (and are even questionable for capturing long-term change given evidence of stability in absolute scores; Howard et al., 2019).

## **Observational Measures**

Both direct assessments and adult reports excel at capturing differences between individuals at a given time point, as well as developmental change within individuals. However, both approaches provide little insight into shorter-term fluctuations in self-regulation over the course of a day or week, for example, or the role of specific contextual factors in explaining variability in children's scores. This is exacerbated by their other drawbacks, such as low ecological validity for direct assessments and subjectivity for adult ratings. For researchers seeking to understand children's behavior in everyday contexts and investigate how social-contextual factors influence fluctuation and incremental growth in self-regulation, child observations are an alternative to existing measurement approaches.

Observational measures create the opportunity to assess children's self-regulation in everyday contexts. Assessors or trained observers rate whether a child's behavior indicates the application of self-regulation and the sophistication of that application. By employing a standard method to code behaviors and a rater that does not know the child, observations can be more objective than adult reports. Many of these measures are also ecologically valid, directly observing children's authentic behaviors and responses in everyday settings or situations that require self-regulation. Researchers have long used direct observations of children's classroom behaviors to study inattentive, hyperactive, and impulsive behaviors associated with attention-deficit/hyperactivity disorder (e.g., Platzman et al., 1992). Measures that assess psychopathologies usually take specificities of the disorder into account, for example, to diagnose attention-deficit/hyperactivity disorder impairments must persist across situations and contexts (Toplak et al., 2013). In more recent years, researchers have also developed observational ratings for assessing self-regulation in typically developing populations (see Table 1 for examples).

Observations of self-regulation can include structured, semi-structured, or unstructured approaches. For example, the Preschool Self-Regulation Assessment report (PSRA; Smith-Donald et al., 2007) is a 28-item rating scale assessing children's emotions, attention, and behavior during a set of highly structured self-regulation tasks. This observational measure offers raters an opportunity to evaluate how children respond to situations which put demands on their attention

**Table 1.** Examples of Observational Measurement Approaches.

Observational measures	Preschool Self-Regulation Assessment (PSRA) assessor report (Smith-Donald et al., 2007)	Preschool Situational Self-Regulation Toolkit (PRISIST; Howard et al., 2019)	Regulation Related Skills Measure (RRSM; McCoy et al., 2022)
<b>Technical aspects of measures</b>			
Number of items	28 items (4-point Likert scale)	9 items (7-point Likert scale)	16 items (4-point Likert scale)
Example item	<i>“Remains in seat appropriately during test”</i>	<i>“Did the child sustain attention, and resist distraction, throughout the instructions and activity?”</i>	<i>“Follows classroom rules and routines independently”</i>
Subscales	- Attentive/Impulse control scale ( $\alpha = .89$ ) - Positive emotion scale ( $\alpha = .87$ )	- Cognitive self-regulation ( $\alpha = .95$ ) - Behavioral self-regulation ( $\alpha = .86$ )	- Attention control ( $\alpha = .82$ ) - Inhibitory control ( $\alpha = .70$ )
Who completes coding?	Trained observer conducts live coding following direct assessments	Trained observer conducts live coding following each activity	Trained observer conducts video coding
Reported inter-rater reliability	ICC = .82 for attention/impulse control	Individual item percent agreement = 80%, kappa = .69	Three criteria: Mean difference in average rating <.75, correlation between item ratings >.70, at least 80% of item ratings within one point
Validity	Concurrent validity with problem behaviors, academic achievement, and social competence	Concurrent validity with school readiness and with task-based assessment and adult rating of self-regulation	Concurrent validity with teacher ratings and task-based assessments of self-regulation
<b>Characteristics of the assessment situation</b>			
Typical application of measure	Assessment of children’s self-regulation during their completion of other structured and standardized direct assessments	Assessment of children’s self-regulation while playing a game with peers or playing a guessing game one-on-one with an adult	Assessment of children’s self-regulation in naturally occurring child-led and teacher-led classroom activities and the subsequent transitions
Level of structure during assessment	Highly structured (lab-like environment)	Semi-structured (activities that are typical of, and are conducted in, a classroom)	Unstructured (naturalistic classroom environment)
Familiarity of task	Tasks tend to be unfamiliar	Tasks tend to be familiar	Tasks are part of children’s daily classroom environment
Role of peers during assessment	No peer interaction	Peers as play partners in one of the activities	There may be peer interaction
Role of teacher during assessment	No teacher interaction	No teacher interaction	There may be teacher interaction

(e.g., “Pays attention during instructions and demonstrations”) and behavioral inhibition (e.g., “Lets examiner finish before starting task; does not interrupt”; Smith-Donald et al., 2007). Another measure, the Preschool Situational Self-Regulation Toolkit assessment (PRSIST; Howard et al., 2019), uses semi-structured group and individual activities, namely: a memory-card matching game completed in a group with four peers, as well as a curiosity box guessing game. The different activities used for the PRSIST have inherent self-regulation demands, such as turn taking, resisting impulsive behaviors, rules that should be followed, and internalization of self-regulation processes. A trained observer rates children’s ability to self-regulate on nine items tapping cognitive self-regulation (e.g., “Did the child sustain attention, and resist distraction, throughout the instructions and activity?”) and behavioral self-regulation (e.g., “Did the child control their behaviors and stay within the rules of the activity?”; Howard et al., 2020). A third measure, the Regulation Related Skills Measure (RRSM; McCoy et al., 2022), is an unstructured observation wherein children are observed during the last 5 minutes of naturally occurring teacher-led and child-led situations in early childhood classrooms, as well as transitions from one activity to the next. As such, children are observed during diverse activities which are not predetermined or controlled. The RRSM includes 16 items that assess children on similar constructs as PRSIST. Example items include “Pays attention to activity at hand” or “Follows classroom rules and routines independently”.

## Opportunities for Advancing Knowledge Using Observational Approaches

The three measures described above and in Table 1 demonstrate the range of observational approaches available to assess children’s self-regulation in different contexts. We argue that such approaches present opportunities to advance understanding of intra-individual differences in children’s self-regulation, its social-contextual influences, and how its constituent elements combine to affect self-regulatory success or failure.

### *Child Observations Can Provide New Insights Regarding Intra-Individual Differences of Self-Regulation*

Observational measures are uniquely positioned to capture repeated measurements of children’s behavior and thereby detect intra-individual variability in self-regulation. To date, little is known about the variability versus the “stability” of children’s self-regulation across situations, despite caregivers experiencing this firsthand and theoretical models hypothesizing that self-regulation is dynamic in response to external factors in the environment (Nigg, 2017).

The lack of studies on intra-individual variability may be due to the constraints inherent in direct assessments (administration in a fixed, standard situation, limiting contextual and environmental variability) and adult reports (broad measures providing a global estimate) that we reviewed above. Hence, adult reports are not precise or sensitive enough to capture intra-individual changes across contexts and indeed are intended to aggregate across them.

Observational measures may provide opportunities to overcome these limitations. For example, the PRSIST scale can be applied within two prescribed activities but also outside of them (using its items, elaborations and scoring scheme). Similarly, the RRSM was designed to observe children’s self-regulation across a variety of typical classroom activities, meaning that it is specifically designed to capture repeated observations of behavior during a range of activities, social contexts, and parts of the school day. Using the RRSM, McCoy et al. (2022) found substantial intra-individual differences in children’s attention and inhibitory control across classroom activities.

## Investigating The Influence of Social-Contextual Factors on Children's Self-Regulation

Observational measures are also uniquely suited to investigating questions about social-contextual factors that help or hinder children to regulate their attention and behavior in the classroom. As noted above, children's regulation is thought to emerge through ongoing transactions between children and their social environments (Sameroff, 2010). As a result, external influences from the social environment—including interactions with adults and peers, social norms, classroom routines and expectations for behavior—are important influences on self-regulation (Nigg, 2017). However, the influence of social or regulatory partners is rarely taken into consideration in research (Finch et al., 2019). One reason might be that the child literature relies on the skill model of self-regulation, which considers self-regulation as a relatively stable individual competence (Eisenberg et al., 2004). After all, the most commonly used measures of self-regulation—direct assessments and adult reports—are designed to capture stable individual differences. Little is known about immediate social influences on children's enactment of regulation, despite its importance to developmental theory.

Furthermore, although self-regulation has been acknowledged to develop in interaction with social partners (e.g., Bernier et al., 2010; 2015), the primary focus of research is typically on adult caregivers and the quality of interactions are usually considered as *a longitudinal predictor* of children's self-regulation (Robson et al., 2020) rather than a social influence in the moment. However, it is well-recognized that these continual, varying social influences are what build children's capacity for self-regulation (Sameroff, 2010; Vygotsky, 1978). Caregivers and their infants influence one another moment-to-moment by responding to each other's emotional or behavioral cues and infants build internal mental representations based on the interactions that they experience (Beeghly & Tronick, 2011). These internal mental representations can include, for example, how emotional conflicts can be resolved or expectations of how social partners may respond (Beeghly & Tronick, 2011). Further, after children acquire measurable self-regulation skills in early childhood, social partners such as parents and peers are often no longer considered when explaining self-regulation scores. As a result, the role of social partners in childhood self-regulation is not well understood, although it might consistently be involved in regulation throughout and beyond childhood (Beckes & Coan, 2011; Coan et al., 2006).

Observational measures could help researchers to understand these social influences. For instance, in the PRSIST measure children play a group memory-card matching game, which requires turn taking, following rules, and managing frustration (Howard et al., 2020). Assessing children's self-regulation skills across time and different groups and analyzing how manifestations of self-regulation vary depending on the group composition may help understand how peers influence children's self-regulation responses in the moment and the upper- or lower-bounds of regulation we could expect to see from a child.

Observational measures can also help researchers to understand contextual influences beyond social factors. For example, observational measures could be used to understand how the level of structure in an activity could facilitate children's self-regulation. There is evidence that more structure at home (Berry et al., 2016; Hughes & Ensor, 2009) and more stable routines at school (Rimm-Kaufman et al., 2009) are associated with individual differences of self-regulation. Still, insufficient work has examined how different children (or groups of children) respond to various levels of structure and support in the classroom, despite clear relevance for practitioners. In one of the few available studies on this topic, using the RRSM, McCoy et al. (2022) found that children showed greater self-regulation during classroom transitions than in child- and teacher-led activities. Children also showed greater dysregulation during classroom story time, and those showing the greatest shift toward dysregulation were children rated by their teachers as typically demonstrating lower levels of self-regulation (Koepp et al., 2021). These findings suggest that

children's regulation can be influenced by the situations and activities in which they are engaged at the moment of the observation. By observing intra-individual variability, researchers can better understand how individual children respond to contextual factors and direct education and intervention efforts aiming to impact children's ongoing development in self-regulation.

However, a challenge of observational measures arises from the difficulty to control some environmental aspects of the observation, such as the level of structure in the classroom and the behavior of peers. In other words, the same variability in environmental aspects that can facilitate questions about contextual influences can present challenges for comparability when using unstructured observations. For example, children's classroom environments can be very different (e.g., highly controlled vs. chaotic), and may influence their ability or opportunity to demonstrate their self-regulation. Although the classroom environment is better controlled in semi-structured observations (e.g., when using the PRSIST), certain elements remain variable. Different peers, for example, could facilitate children's ability to demonstrate self-regulation or they could detract from it (although, after single observations, aggregate evidence from the PRSIST assessment indicated low intraclass correlation between self-regulation ratings within the group; Howard et al., 2019).

### *Investigating The Integrated Nature of Self-Regulation*

Direct assessments often fail to account for the integrated nature of self-regulation. That is, researchers frequently assess self-regulatory skills in isolation (e.g., working memory or inhibitory control), even though self-regulation constitutes an array of skills such as cognitive, behavioral, and emotion regulation that must be used in tandem to accomplish one's goals (Blair & Raver, 2015). To illustrate how complex skills emerge from a combination of simple skills, Jones et al. (2016) use the analogy of playing basketball. In basketball, foundational skills such as dribbling and shooting are necessary for playing the game. Still, evaluating only these isolated skills might not be as informative as observing players coordinating these skills, in interaction with their teammates, while adapting to challenging circumstances. Similarly, while direct assessments often aim to capture information about children's skills in isolation and out of context, observations in situ offer an opportunity to better understand the interaction between components of self-regulation and contextual factors. Levels of behavioral and cognitive aspects of self-regulation, for instance, are found to differ non-systematically (Howard et al., 2020), and may also be differentially susceptible to contextual stimuli. They also may not develop perfectly in concert, nuancing notions of when and how to intervene across self-regulation components. Thus, observational measures offer a more nuanced account of the varied intersections between children's cognitive, behavioral, and social-emotional self-regulatory skills (e.g., PRSIST and RRSM) in situations where it is necessary to navigate rules, expectations, and relationships.

### **Lessons Learned**

All authors of this piece have applied observational measures of self-regulation in their research, and some were involved in the development of observational measures (Dana C. McCoy and Andrew E. Koepp were involved in the development of the RRSM; Steven J. Howard was involved in the development of PRSIST). Here we offer some lessons learned. The first lesson we have learned is the unique contribution that observational approaches add to the traditional point-in-time estimation of self-regulation. Evidence from the PRSIST tool showed that when the observational measure of self-regulation was combined with other sources of information about children's regulation (e.g., teacher reports and direct assessments), prediction of school readiness was improved over models that included these measures individually (Howard et al., 2021). The

RRSM has also been found to capture aspects of self-regulation that are distinct from what is captured via other measures (direct assessments and teacher reports), showing low to moderate correlations with existing measures despite adequate inter-rater reliability (McCoy et al., 2022). Thus, there is empirical evidence that observational measures capture aspects of self-regulation that other measurement approaches miss, such that adding them to other measurement approaches increases the prediction of child outcomes.

Yet a primary challenge for observational assessments is that they cannot provide direct information on fundamentally internal processes of self-regulation (Nigg, 2017). As a result, when using observational measures, one must use behavioral cues to infer the enactment (or lack of) self-regulation. However, behaviors across situations or children may not always look the same, so observational measures require a discerning eye and intensive training. One way we have collectively addressed this problem is by providing coders with anchor descriptors for each item. These anchors provide descriptions of behaviors that indicate, in the given setting, what it looks like when children are engaging their attention, ignoring distractions, inhibiting automatic impulses, as well as what it looks like when they are not (see [https://projects.iq.harvard.edu/files/rasm/files/rasm\\_observation\\_scoring\\_guide.pdf](https://projects.iq.harvard.edu/files/rasm/files/rasm_observation_scoring_guide.pdf) for examples of RRSM anchor items). Providing these anchor items, and elaborations to discern gradations between them, helps to standardize how observed behaviors, patterns or tendencies should be captured with ratings of self-regulation.

A related challenge arises in interpreting children's behavior. In particular, it can be difficult to differentiate between self-regulation and compliance (Duckworth & Yeager, 2015). Compliance implies following rules simply because an adult requests it and indicates a lower level of development in comparison to self-regulation (Kopp, 1982), which requires a more flexible, self-motivated adaptation of behavior where children guide themselves (Feng et al., 2017; Kochanska et al., 2001; Kopp, 1982). Similarly, there can be various reasons why children struggle to follow rules ranging from forgetting about rules, intentionally breaking rules, to not following rules because of lack of self-regulation. However, difficulties differentiating between self-regulation and compliance are not unique to observational measurement approaches. As noted above, teachers also struggle to differentiate between self-regulation and compliance when evaluating children's behavior using adult reports (Vasseleu et al., 2021). Adult reports often assess children's self-regulation in relation to class rules and children's independent regulation is considered less frequently, making adult reports particularly susceptible to compliance problems. In fact, teachers sometimes consider compliance and non-disruptive behavior seminal (or sometimes exclusive) indicators of self-regulation (Vasseleu et al., 2021). Direct assessments present similar challenges. In a telling example, children growing up in interdependent cultural contexts emphasizing obedience and parental control have been found to outperform on a delay of gratification task children from cultural contexts that emphasize child autonomy (Lamm et al., 2018). Similarly, being part of a group that delays behavior also seems to influence children's ability to wait for a reward (Doebel & Munakata, 2018). Although all self-regulation measurement approaches face challenges to differentiate compliance, observational approaches might be better suited to emphasize the internalization and independent enactment aspects of self-regulation. That is, in observational approaches, contextual factors such as verbal reminders or comparison with peers can be taken into consideration. For example, the contextual checklist of the RRSM asks if teachers provided behavioral redirection or feedback during the observed activity giving some indication if children had internalized rules and were able to enact them independently.

Besides the difficulty in interpreting some observed behavior, there are also challenges of unstructured naturalistic observations. Some skills such as emotion regulation or planning might only be called upon infrequently, and therefore may be difficult or impractical to measure. Our observational RRSM data suggested there were few instances of emotional dysregulation during regular classroom activities and many children demonstrated regulated behavior most of the time

(McCoy et al., 2022). This presents a challenge for generating holistic self-regulatory scores, because some children are not observed in situations that are equal in their challenge to self-regulation or their engagement of distinct skills. Further, highly internal contributors to self-regulation such as executive functions or planning are difficult to infer from behavior unless the activity presents sufficient demand that challenges that skill (Rabbitt, 1997). Children can also sometimes opt out of classroom situations that challenge self-regulation, and this self-selection may vary with self-regulation ability. Semi-structured or structured observations such as PRSIST or the PSRA assessor report address this challenge to an extent. A consistent challenge, however, is that emotion regulation can be difficult to observe and assess because if one regulates their emotion or does not become frustrated easily, this is not displayed, making it difficult to understand how children recover and regulate (Howard et al., 2020; Kok, 2017). These challenges are also not unique to observational measures, as adult ratings of behavior must also rely on observations of self-regulatory behaviors (Duckworth & Yeager, 2015). Direct assessments are even more problematic in this regard, as they often intentionally forego emotional challenge in the testing situation (Diamond, 2013).

Another challenge is that observational, ecologically valid measures are more difficult to implement than collecting data that relies on direct assessments or adult reports. Observational measures are time- and resource-intensive (McCoy et al., 2017). Data collection, training of coders, and establishing inter-rater reliability requires time and, where video coding is involved, this process is also labor intensive. However, several strategies have been used to mitigate these challenges. Developing anchor items, which provide examples of regulated and dysregulated behavior for each item, has helped establish inter-rater agreement. Further, training modules with in-built inter-rater reliability checks (as in the PRSIST assessment) can be a useful support prior to in-field training and data collection (Early Years Toolbox, 2019). Finally, in the PSRA children are observed by trained observers while completing tasks and no video coding is required.

Another consideration for applied researchers may be how many observations they need to collect for the children in their studies. Researchers might think about this in two ways. They might first consider the range of situations or settings in which they want to observe the child (e.g., child- versus teacher-led activities) and ensure that they observe children in each of the situations or settings that are important for the research question. The second consideration is how many observations are needed to determine a child's range or typical level of self-regulation. This is an area for future empirical work, where researchers establish the number of observations that maximize concurrent or predictive validity while balancing the demands of conducting additional observations.

Despite these challenges, some aspects of the observational measures also make them easier to implement. Although the observational measures require time and energy on the part of researchers, there is very little burden on participants because they are simply observed, either as they complete regular classroom activities (RRSM) or participate in semi-structured, game-like assessments with their peers (PRSIST). Because of this lower burden, educators have been generally willing to work with us on our research projects and allow us in the classroom. Many educators were also interested in learning about children's self-regulation and educators sometimes continued to use the tool to inform their own planning and practice (Vasseleu et al., 2021). In this sense, with additional validation work, observational measures could become useful formative tools for educators (Howard et al., 2020). This would necessitate, for instance, research into the viability of educator use (inter-rater reliability between educators and researchers), as well as the establishment of suitable training, educational support, and linked practices (including what to do with and about the data). In learning about these measures and how to apply them, there may be an opportunity for educators to also learn about children's self-regulation and what it looks like in their classrooms. With this information, they may be better able to tailor practices to the needs of

the children in their classrooms. It may also be that, if educators can attain reliability, research studies could use educators to help collect data, and thus break down geographic constraints and reach hard-to-reach children and communities. Although this would offer an opportunity to include specific populations, inter-rater reliability would need to be established carefully to avoid subjectivity.

In sum, when applying observational measures, researchers may experience challenges concerning the interpretability of children's behavior, the unstructured environment, and implementation. Although observational measurement approaches have great potential, measurement approaches should be carefully chosen based on the research question. The next section describes research questions and their fit to different measurement approaches.

## Measurement Approach Selection—There Is No One Size Fits All

Observational measures can enhance our understanding of how children self-regulate their behavior in highly complex and unpredictable situations with differing rules, hidden expectations, and emotional demands that cannot be mimicked in direct assessments. However, when choosing a measurement tool, we recommend researchers evaluate the benefits and costs of each approach (i.e., direct assessments, adult reports, and child observations), as there is no “gold standard” tool or approach for all questions we must ask about self-regulation.

For some applications, observational measures may not be the best choice. For example, for research where a single index is sufficient, longitudinally predictive measures such as the Head-Toes-Knees-Shoulders task (Cameron Ponitz et al., 2009), or adult reports, may do as well or better than observational approaches, with fewer resources required for training and data collection.

In contrast, if researchers aim to test the extent to which an intervention improved adaptive classroom behaviors (e.g., standing in line, taking turns, and not shouting out), direct assessments might be less appropriate. It is not yet clear if children's performance on direct assessments reflects their self-regulation behaviors and choices in everyday situations (Jones et al., 2016). Indeed, there are cases of observational measures of self-regulation being successfully used to detect outcome differences in intervention studies (Howard et al., 2020; Raver et al., 2011). For example, Raver et al. (2011) used the PSRA assessor report in combination with direct assessments in an evaluation study of the Chicago School Readiness Program (CSRP). The assessor report and direct assessments of executive functions detected intervention differences. Howard et al. (2020) used the PRSIST scale along with adult reports and direct assessments to evaluate a program designed to enhance children's self-regulation. However, significant effects of the intervention were only found with direct assessments. The authors suggested that this was probably related to the intervention (e.g., intensity, duration, and implementation quality) and not the measures. Relatedly, interventions could benefit from observational measures, using them for formative (in addition to summative and evaluative purposes) to tailor the intervention to specific individuals, situations, conditions, and circumstances. Direct observations are also well suited to within-person study designs, including single-subject designs. For example, researchers might observe how an individual child responds in different conditions and contexts, such as receiving instructional supports during some activities but not others.

Moreover, observational measures' ecological validity, easy applicability in school settings, consideration of context, and ability to detect intra-individual differences make them valuable for researchers who are interested in applied self-regulation in the classroom. As described before, observational measures can be used to better understand the role of social-contextual factors in children's self-regulation development. Social interaction styles (e.g., role of peers or teachers) or concrete situational supports (e.g., routines) that appear to benefit children's self-regulatory skills

can be identified and policy makers and practitioners could make use of these findings. As more researchers use observational approaches, we will be able to build on the experiences of others, for example, regarding which circumstances and for which questions this approach is useful. Feasibility of these observational tools can be discussed amongst researchers and adjustments can be made, leading to a virtuous cycle of refinements. This process can lead to improvement of ecologically valid measurement approaches and proliferation of these tools in the research community.

## Conclusion

Different measurement approaches such as direct assessments, adult reports, and child observations can each contribute to our understanding of children's self-regulation. The application of multiple measurement approaches would be ideal but can be too costly for researchers or burdensome for participants. Hence, researchers are often forced to opt for one measurement approach over another. We argue that, depending on the nature of the research question, observational measures may be an under-utilized resource for capturing children's self-regulation in real-world contexts (Robson et al., 2020). Rather than favoring one measurement approach over another, we want to encourage a discussion and optimized decisions regarding the situations for which each measurement approach is best suited. Observations open new possibilities for applied research that can enhance our knowledge of children's behavior in everyday environments beyond unfamiliar, decontextualized, highly structured tasks that have unclear parallels to the everyday situations in which children are expected to self-regulate. This knowledge is highly valuable for researchers and practitioners, increasing our understanding of children's self-regulation development and catering interventions more precisely towards children's needs.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: AK is supported by the National Academy of Education/Spencer Dissertation Fellowship. RK is supported by an EUR Fellowship Grant from the Erasmus University Rotterdam. STB is supported by the LEGO Foundation and Cambridge Trust.

## ORCID iDs

Janina Eberhart  <https://orcid.org/0000-0002-3650-5452>

Steven J. Howard  <https://orcid.org/0000-0002-1258-3210>

## References

- Ahmed, S. F., Grammer, J., & Morrison, F. (2021). Cognition in context: Validating group-based executive function assessments in young children. *Journal of Experimental Child Psychology*, 208(2021), 105131. <https://doi.org/10.1016/j.jecp.2021.105131>
- Beckes, L., & Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Social and Personality Psychology Compass*, 5(12), 976–988. <https://doi.org/10.1111/j.1751-9004.2011.00400.x>

- Beeghly, M., & Tronick, E. (2011). Early resilience in the context of parent-infant relationships: A social developmental perspective. *Current Problems in Pediatric and Adolescent Health Care*, 41(7), 197–201. <https://doi.org/10.1016/j.cppeds.2011.02.005>
- Bernier, A., Beauchamp, M. H., Carlson, S. M., & Lalonde, G. (2015). A secure base from which to regulate: Attachment security in toddlerhood as a predictor of executive functioning at school entry. *Developmental Psychology*, 51(9), 1177–1189. <https://doi.org/10.1037/dev0000032>
- Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development*, 81(1), 326–339. <https://doi.org/10.1111/j.1467-8624.2009.01397.x>
- Berry, D., Blair, C., Willoughby, M., Garrett-Peters, P., Vernon-Feagans, L., & Mills-Koonce, W. R. (2016). Household chaos and children's cognitive and socio-emotional development in early childhood: Does childcare play a buffering role? *Early Childhood Research Quarterly*, 34, 115–127. <https://doi.org/10.1016/j.ecresq.2015.09.003>
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, 21(4), 327–336. <https://doi.org/10.1016/j.lindif.2011.01.007>
- Blair, C., & Raver, C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, 66(1), 711–731. <https://doi.org/10.1146/annurev-psych-010814-015221>
- 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(2), 647–663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter-word knowledge in the early elementary grades. *Developmental Psychology*, 51(4), 459–472. <https://doi.org/10.1037/a0038813>
- Cameron Ponitz, C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, 45(3), 605–619. <https://doi.org/10.1037/a0015365>
- Coan, J. A., Schaefer, H. S., & Davidson, R. J. (2006). Lending a hand: Social regulation of the neural response to threat. *Psychological Science*, 17(12), 1032–1039. <https://doi.org/10.1111/j.1467-9280.2006.01832.x>
- Cole, P. M., Ram, N., & English, M. S. (2019). Toward a unifying model of self-regulation: A developmental approach. *Child Development Perspectives*, 13(2), 91–96. <https://doi.org/10.1111/cdep.12316>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Doebel, S., & Munakata, Y. (2018). Group influences on engaging self-control: Children delay gratification and value it more when their in-group delays and their out-group doesn't. *Psychological Science*, 29(5), 738–748. <https://doi.org/10.1177/0956797617747367>
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. <https://doi.org/10.3102/0013189X15584327>
- Early Years Toolbox. (2019). *EYT training and resources*. <http://www.cytoolbox.com.au/eyt-training>
- Eisenberg, N., Smith, C., & Spinrad, T. (2004). Effortful control. In K. D. Vohs & R. F. Baumeister (Eds), *Handbook of self-regulation: Research, theory, and applications* (3rd ed.). Guilford Press.
- Feng, X., Hooper, E. G., & Jia, R. (2017). From compliance to self-regulation: Development during early childhood. *Social Development*, 26(4), 981–995. <https://doi.org/10.1111/sode.12245>
- Finch, J. E., Garcia, E. B., Sulik, M. J., & Obradović, J. (2019). Peers matter: Links between classmates' and individual students' executive functions in elementary school. *AERA Open*, 5(1), 2332858419829438. <https://doi.org/10.1177/2332858419829438>

- Garcia, E. B., Sulik, M. J., & Obradović, J. (2018). Teachers' perceptions of students' executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology, 111*(5), 918–931. <https://doi.org/10.1037/edu0000308>
- Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *Behavior rating inventory of executive function—preschool version (BRIEF-P)*. PAR.
- Graziano, P. A., Slavec, J., Ros, R., Garb, L., Hart, K., & Garcia, A. (2015). Self-regulation assessment among preschoolers with externalizing behavior problems. *Psychological Assessment, 27*(4), 1337–1348. <https://doi.org/10.1037/pas0000113>
- Howard, S. J., Neilsen-Hewett, C., de Rosnay, M., Vasseleu, E., & Melhuish, E. (2019). Evaluating the viability of a structured observational approach to assessing early self-regulation. *Early Childhood Research Quarterly, 48*(12), 186–197. <https://doi.org/10.1016/j.ecresq.2019.03.003>
- Howard, S. J., Vasseleu, E., Batterham, M., & Neilsen-Hewett, C. (2020). Everyday practices and activities to improve pre-school self-regulation: Cluster RCT evaluation of the PRSIST program. *Frontiers in Psychology, 11*, 137. <https://doi.org/10.3389/fpsyg.2020.00137>
- Howard, S. J., Vasseleu, E., Neilsen-Hewett, C., de Rosnay, M., & Williams, K. E. (2021). Predicting academic school readiness and risk status from different assessment approaches and constructs of early self-regulation. *Child and Youth Care Forum, 51*(2), 369–393. <https://doi.org/10.1007/s10566-021-09636-y>
- Hughes, C., & Ensor, R. (2009). How do families help or hinder the emergence of early executive function? In C. Lewis & J. I. M. Carpendale (Eds), *Social interaction and the development of executive function*. New Directions in Child and Adolescent Development (123, pp. 35–50).
- Jones, S. M., Bailey, R., Barnes, S. P., & Partee, A. (2016). Executive function mapping project: Untangling the terms and skills related to executive function and self-regulation in early childhood. *OPRE report # 2016-88*. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development, 72*(4), 1091–1111. <https://doi.org/10.1111/1467-8624.00336>
- Koeppe, A. E., Gershoff, E. T., Castelli, D. M., & Bryan, A. E. (2021). Measuring children's behavioral regulation in the preschool classroom: An objective, sensor-based approach. *Developmental Science, 25*(5), Article e13214. <https://doi.org/10.1111/desc.13214>
- Kok, R. (2017). Emotion regulation. In V. Zeigler-Hill & T. K. Shackelford (Eds), *Encyclopedia of personality and individual differences*. Springer International Publishing. [https://doi.org/10.1007/978-3-319-28099-8\\_811-1](https://doi.org/10.1007/978-3-319-28099-8_811-1)
- Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*(2), 199–214. <https://doi.org/10.1037/0012-1649.18.2.199>
- Lamm, B., Keller, H., Teiser, J., Gudi, H., Yovsi, R. D., Freitag, C., Poloczek, S., Fassbender, I., Suhrke, J., Teubert, M., Vöhringer, I., Knopf, M., Schwarzer, G., & Lohaus, A. (2018). Waiting for the second treat: Developing culture-specific modes of self-regulation. *Child Development, 89*(3), Article e261-e277. <https://doi.org/10.1111/cdev.12847>
- Liew, J., Valiente, C., Hernández, M. M., & Abera, D. (2019). Emotional self-regulation and reactivity, school-based relationships, and school engagement and achievement. In D. Whitebread, V. Grau, K. Kumpulainen, M. M. McClelland, N. E. Perry, & D. Pino-Pasternak (Eds.), *The SAGE handbook of developmental psychology and early childhood education* (pp. 42–62). SAGE Publications Ltd. <https://doi.org/10.4135/9781526470393.n3>
- McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6*(2), 136–142. <https://doi.org/10.1111/j.1750-8606.2011.00191.x>

- McCoy, D. C. (2019). Measuring young children's executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*, 22(1), 63–74. <https://doi.org/10.1007/s10567-019-00285-1>
- McCoy, D. C., Jones, S. M., Leong, D., Bodrova, E., Koepp, A., & Hemmenway, A. (2017). *The regulation-related skills measure*. Harvard University.
- McCoy, D. C., Koepp, A. E., Jones, S. M., Bodrova, E., Leong, D. J., & Deaver, A. H. (2022). An observational approach for exploring variability in young children's regulation-related skills within classroom contexts. *Developmental Science*, 25(6), Article e13250. <https://doi.org/10.1111/desc.13250>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698. <https://doi.org/10.1073/pnas.1010076108>
- Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 58(4), 361–383. <https://doi.org/10.1111/jcpp.12675>
- Obradović, J., Sulik, M. J., Finch, J. E., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology*, 55, 4–13. <https://doi.org/10.1016/j.appdev.2017.03.003>
- Platzman, K. A., Stoy, M. R., Brown, R. T., Coles, C. D., Smith, I. E., & Falek, A. (1992). Review of observational methods in attention deficit hyperactivity disorder (ADHD): Implications for diagnosis. *School Psychology Quarterly*, 7(3), 155–177. <https://doi.org/10.1037/h0088258>
- Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–37). Psychology Press.
- Raver, C. C., Li-Grining, C., Bub, K., Jones, S. M., Zhai, F., & Pressler, E. (2011). CRSP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. *Child Development*, 82(1), 362–378. <https://doi.org/10.1111/j.1467-8624.2010.01561.x>
- Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, 45(4), 958–972. <https://doi.org/10.1037/a0015861>
- Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, 146(4), 324–354. <https://doi.org/10.1037/bul0000227>
- Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, 81(1), 6–22. <https://doi.org/10.1111/j.1467-8624.2009.01378.x>
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the preschool self-regulation assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22(2), 173–187. <https://doi.org/10.1016/j.ecresq.2007.01.002>
- Tamm, L., & Peugh, J. (2019). Concordance of teacher-rated and performance-based measures of executive functioning in preschoolers. *Child Neuropsychology*, 25(3), 410–424. <https://doi.org/10.1080/09297049.2018.1484085>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, 54(2), 131–143. <https://doi.org/10.1111/jcpp.12001>
- Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure: Promotion of self-regulation in school. *Child Development Perspectives*, 6(2), 122–128. <https://doi.org/10.1111/j.1750-8606.2011.00209.x>

- Vasseleu, E., Neilsen-Hewett, C., Cliff, K., & Howard, S. J. (2021). How educators in high-quality preschool services understand and support early self-regulation: A qualitative study of knowledge and practice. *The Australian Educational Researcher*. <https://doi.org/10.1007/s13384-021-00466-4>
- von Suchodoletz, A., Uka, F., & Larsen, R. A. A. A. (2015). Self-regulation across different contexts: Findings in young Albanian children. *Early Education and Development*, 26(5–6), 829–846. <https://doi.org/10.1080/10409289.2015.1012189>
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Harvard University Press. (Original work published in 1930, 1933, 1935).