Studying children’s growth in self-regulation using changing measures to account for heterotypic continuity: A Bayesian approach to developmental scaling

Alexis Hosch¹ | Jacob J. Oleson² | Jordan L. Harris¹ | Mary Taylor Goeltz³ | Tabea Neumann⁴ | Brandon LeBeau⁵ | Eliot Hazeltine¹ | Isaac T. Petersen¹

¹Department of Psychological and Brain Sciences, University of Iowa, Iowa City, Iowa, USA
²Department of Biostatistics, University of Iowa, Iowa City, Iowa, USA
³Department of Psychology, University of Colorado Denver, Denver, Colorado, USA
⁴School of Medicine, University of Dundee, Dundee, Scotland, UK
⁵Department of Psychological and Quantitative Foundations, University of Iowa, Iowa City, Iowa, USA

Correspondence
Alexis Hosch, Department of Psychological and Brain Sciences, University of Iowa, G60 Psychological and Brain Sciences Building, Iowa City, IA 52242, USA. Email: alexis-hosch@uiowa.edu

Funding information
Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Grant/Award Number: HD098235; National Institute of General Medical Sciences (NIGMS), Grant/Award Number: T32GM108540; National Center for Advancing Translational Sciences (NCATS), Grant/Award Number: UL1TR002537

Abstract
Self-regulation is thought to show heterotypic continuity–its individual differences endure but its behavioral manifestations change across development. Thus, different measures across time may be necessary to account for heterotypic continuity of self-regulation. This longitudinal study examined children’s (N = 108) self-regulation development using 17 measures, including 15 performance-based measures, two questionnaires, and three raters across seven time points. It is the first to use different measures of self-regulation over time to account for heterotypic continuity while using developmental scaling to link the measures onto the same scale for more accurate growth estimates. Assessed facets included inhibitory control, delayed gratification, sustained attention, and executive functions. Some measures differed across ages to retain construct validity and account for heterotypic continuity. A Bayesian longitudinal mixed model for developmental scaling was developed to link the differing measures onto the same scale. This allowed charting children’s self-regulation growth across ages 3–7 years and relating it to both predictors and outcomes. Rapid growth occurred from ages 3–6. As a validation of the developmental scaling approach, greater self-regulation was associated with better school readiness (math and reading skills) and fewer externalizing problems. Our multi-wave, multi-facet, multi-method, multi-measure, multi-rater, developmental scaling approach is the most comprehensive to date for assessing the development of self-regulation. This approach demonstrates that developmental scaling may enable studying development of self-regulation across the lifespan.

KEYWORDS
changing measures, construct validity invariance, developmental scaling, heterotypic continuity, longitudinal, self-regulation
INTRODUCTION

Considerable research has demonstrated that children's ability to willfully regulate their thoughts, emotions, and behaviors holds important implications for their long-term outcomes. This ability, commonly referred to as self-regulation, involves the flexible control of attentional, cognitive, emotional, and behavioral processes in pursuit of a goal (Berger, 2011; Calkins & Fox, 2002). Self-regulation has shown concurrent and predictive associations with myriad outcomes in childhood, including internalizing and externalizing problems (Eisenberg et al., 2009; Espy et al., 2011; Martel & Nigg, 2006; Olson et al., 2005; Petitclerc et al., 2015; Rothbart & Bates, 2006), social and intellectual functioning (Blair & Razza, 2007; Kochanska, 1997; Padilla-Walker & Christensen, 2011; Spinrad et al., 2006), and school readiness (Blair & Diamond, 2008; Liew, 2012; Liew et al., 2018). Furthermore, longitudinal research has shown that children with poorer self-regulation tend to have worse health, less wealth, and more criminal involvement as adults (Moffitt et al., 2011). Thus, it is crucial to investigate how self-regulation develops.

Self-regulation has been conceptualized in many ways, due partly to a lack of consensus regarding accepted terminology and to differing emphases of various research traditions. For example, neuropsychologists have examined self-regulatory processes called executive functions—higher-order (“top-down”) processes that exert control over attentional, cognitive, and behavioral tendencies in pursuit of a goal (Zhou et al., 2012). Temperament researchers have proposed that self-regulation is the result of an executive attentional control system (Shallice, 1988), in which effortful control reflects the efficiency of the executive attentional control system's ability to inhibit prepotent responses, plan behavior, and detect errors (Posner & Rothbart, 2000; Rothbart & Bates, 2006; Tieto et al., 2020). Other developmental researchers have argued that there should be greater consideration of emotional processes within a framework of self-regulation (Cole et al., 2004; Eisenberg et al., 2001; Lewis & Stieben, 2004; Mischel & Ayduk, 2004). Consequently, considerable work has examined self-regulation as a regulatory system that involves distinct “hot” (i.e., motivationally or affectively mediated) and “cool” (i.e., cognitively mediated) processes (Backer-Grondahl et al., 2019; Bechara et al., 1994; Cameron Ponitz et al., 2008; Denham et al., 2012; Metcalfe & Mischel, 1999; Simpson & Carroll, 2019; Willoughby et al., 2011). For instance, some factor analysis research has demonstrated that tasks designed to assess the inhibitory control aspect of self-regulation—the ability to inhibit responses to irrelevant stimuli in pursuit of a cognitively represented goal—load onto separate latent “hot” and “cool” factors (Bridgett et al., 2015; Carlson & Moses, 2001; Murray & Kochanska, 2002; Simpson & Carroll, 2019).

The “hot” and “cool” factor conceptualization of self-regulation is not universally supported, however. Some have argued that an integrated, single factor model of regulation more accurately represents the construct, particularly in early childhood (Allan & Lonigan, 2011; Cole et al., 2019; Lin et al., 2019; Sulik et al., 2010; Wiebe et al., 2008). Thus, researchers have called for an integrated model of self-regulation, highlighting meaningful conceptual and measurement overlap between regulation-related constructs, including effortful control and executive function (Bridgett et al., 2013; Zhou et al., 2012), metacognition and executive function (Roebers, 2017), executive function and self-regulation (Best et al., 2009; Hofmann et al., 2012; McCoy, 2019; Roebers, 2017), executive function and emotion regulation (Zelazo & Cunningham, 2007), and cognitive control and self-regulation (Mischel et al., 2011).

In response to calls for an integrated model of self-regulation, Nigg (2017) proposed a domain-general conceptualization of self-regulation that integrates diverse constructs, across emotion, action, and cognition, into a unified framework to provide consistency and prevent confusion within the field. This domain-general framework may also aid the development and improvement of measurement techniques and interpretation of results. We draw upon Nigg’s (2017) framework in our conceptualization of self-regulation and apply a domain-general model that encompasses various regulatory processes. We acknowledge that this is one of several empirically supported conceptualizations of self-regulation. Other approaches, such as a formative model of self-regulation, in which the construct is derived from the summation of a set of processes, may also reasonably operationalize the construct (Camerota et al., 2020; Willoughby et al., 2017). Alternatively, it is possible that what researchers have called “self-regulation” is merely a heuristic that describes a set of separate yet correlated abilities that do not reflect a common construct (Eisenberg et al., 2018, 2019). In sum, the structure of self-regulation is highly debated and remains an important empirical question. Research has supported several conceptualizations of self-regulation, including reflective, formative, or heuristic models. More research is needed to delineate how self-regulatory processes (e.g., inhibitory control, executive functions, etc.) are related within a broader developmental framework.

However, prior research has generally indicated that regulatory processes are inter-correlated and that there is considerable conceptual and measurement overlap between several components of self-regulation, including effortful control, executive functions, inhibitory

RESEARCH HIGHLIGHTS

- Self-regulation shows heterotypic continuity, but studies have not accounted for it when examining individuals' growth.
- This study used different measures across ages and developmental scaling to account for heterotypic continuity and chart children's self-regulation growth across ages 3–7 years.
- The developmentally scaled model of self-regulation placed age-differing measures onto the same scale and showed criterion validity in relation to externalizing problems and school readiness.
- Developmental scaling may promote studying individuals' self-regulation development across the lifespan.
control, and others (Berger, 2011; Bridgett et al., 2013; Carlson & Wang, 2007; Lin et al., 2019; Nigg, 2017; Reed et al., 2020; Zelazo & Cunningham, 2007; Zhou et al., 2012). Given the overlap between concepts, evidence suggests that there may be a general, over-arching factor (Allan & Lonigan, 2011; Espy et al., 2011; Lin et al., 2019; Sulik et al., 2010; Wiebe et al., 2008, 2011). Thus to prevent confusion across constructs and to facilitate more efficient communication across research groups (Cole et al., 2019; McClelland et al., 2010; Nigg, 2017; Zhou et al., 2012), we conceptualized self-regulation as a higher-order construct, reflecting cognitively and affectively mediated regulatory abilities, consistent with prior studies (Allan & Lonigan, 2011; Espy et al., 2011; Sulik et al., 2010; Wiebe et al., 2011). Cognitive regulatory processes include, for instance, sustained attention (i.e., the ability to maintain focus on a given task over prolonged periods), inhibitory control, and higher-order executive functions. Affective regulatory processes include, for instance, the ability to delay gratification (i.e., the ability to resist temptation in favor of long-term goals) and regulate emotions (Bridgett et al., 2015; Gagne et al., 2021; Metcalfe & Mischel, 1999).

Previous research has also not precisely delineated how regulatory abilities develop across the lifespan. Prior literature generally supports a developmental model in which lower-level processes develop in early childhood and are followed by higher-level processes in later childhood. Montroy et al. (2016) described a hierarchical differentiation framework, in which children develop separate skills that enable self-regulation in infancy and later integrate these processes into a hierarchically organized regulatory system. Indeed, research has shown that there is a qualitative shift in regulatory skills beginning at age three, in which rapid growth occurs and then shows marked deceleration around age seven (Cameron Ponitz et al., 2008; Diamond, 2002; Montroy et al., 2016; Wiebe et al., 2011). Similarly, inhibitory control and working memory processes manifest in the first years of life and increase in capacity across the preschool years (Geeraerts et al., 2021; Greene, 2017; Kopp, 1982). As children enter formal schooling around age five, their executive function capacity increases, which supports early manifestation of higher-order processes such as cognitive flexibility and active self-regulation of cognition, emotion, and behavior (Anderson, 2002; Berger, 2011; Greene, 2017).

Individual trajectories may differ from this prototypical developmental timeline. For example, Montroy et al. (2016) examined 1386 children aged 3–7 using an inhibitory control task and found that, while most children demonstrated a pattern of rapid development of self-regulation followed by a deceleration period, child-specific factors (e.g., gender and language ability) predicted when and how quickly this growth occurred. Thus, behavioral manifestations of self-regulation may change across development due to non-linear development of self-regulation, as well as child-specific individual differences.

Persistence of a construct, such as self-regulation, with behavioral manifestations that change across development is called heterotypic continuity (Cicchetti & Rogosch, 2002). Self-regulation is thought to manifest differently in preschool-aged children compared to those in later childhood (Greene, 2017; Kopp, 1982; Petersen et al., 2016). In early childhood, self-regulation is thought to reflect a gradual transition from external sources of control to internal self-control (Berger, 2011; Kopp, 1982). Infants are reliant on caregivers to provide regulation, such as soothing through feeding, diaper changing, or holding (Kopp, 1982). Infants are also able to reduce excessive arousal or stimulation by turning away or self-soothing (Kopp, 1982). Between ages 2 and 3 years, children begin to develop more sophisticated forms of cognition, such as language and representational thinking, which allow them to act intentionally and comply with external commands (Berger, 2011; Kochanska, 2002; Kochanska et al., 2001; Kopp, 1982). However, 2- and 3-year-old children are still largely reliant on caregivers and more likely to react with physical aggression and have emotional outbursts during this time (Kopp, 1982). Kopp refers to this phase as “self-control”, a more limited form of self-regulation, characterized by the development of autonomy and self-awareness. “Real” forms of self-regulation begin to emerge between 3 and 4 years of age, in which children become increasingly able to use rules, strategies, and plans to guide behavior (Berger, 2011; Kopp, 1982). During this time, children may use private (self-directed) speech to guide thoughts and actions during challenging tasks (Berger, 2011; Berk, 1999; Bivens & Berk, 1990). Initially, private speech functions as a planning instrument, occurring before action, in which children regulate their actions verbally. Eventually, private speech is thought to become internalized between ages 6 and 8 years, and it serves as an internal regulatory mechanism (Berger, 2011; Berk, 1999; Bivens & Berk, 1990). Internalized private speech is considered critical for self-regulation (Berger, 2011). Language achievements and concomitant growth in self-control (Whedon et al., 2021) are paralleled by development in the prefrontal cortex (e.g., anterior cingulate cortex and dorsolateral prefrontal cortex) and executive functions. This growth, which typically occurs between ages 3 and 7 years, supports more sophisticated forms of self-regulation as children get older (Berger, 2011; Diamond, 2002; McClelland et al., 2010).

With age, an increase in developmental capacity, paired with environmental changes (e.g., school entry), leads to heterogeneous manifestations of self-regulation. For instance, verbally requesting a toy rather than employing an automatic response, such as physical aggression, may indicate overt self-regulation in younger children, whereas similar behavior in older childhood may not reflect the same degree of inhibition. Among older children, self-regulation may instead appear as inhibition of a prepotent behavioral response despite a concrete command (e.g., “Simon Says”) or social pressure (e.g., invitation by a peer to participate in a rule-breaking action), or as completion of a homework assignment that requires integration of planning, working memory, and control. In general, elementary school children tend to be more responsible and conscious of their behavior compared to preschool children (Berger, 2011).

Empirical work supports the notion that self-regulation shows heterotypic continuity. Studies have examined the heterotypic continuity of specific components of self-regulation, including inhibitory control (Geeraerts et al., 2021; Petersen et al., 2016; Petersen, Bates, et al., 2021), effortful control (Putnam et al., 2008), and emotional/behavioral control (Chang et al., 2015; Zimmermann & Iwanski, 2014). However, no studies have examined the heterotypic continuity...
of the higher-order self-regulation construct. A meta-analysis found that the behavioral manifestations of inhibitory control changed across time in children between 1 and 8 years of age (Petersen et al., 2016). More specifically, findings suggested that perceptual inhibition may develop earlier than other forms of inhibition, such as performance and association inhibition, which in turn may develop earlier than motivational inhibition. Similar patterns have also been found for other components of self-regulation. For example, Chang et al. (2015) found that children displayed different forms of emotional and behavioral control in early childhood, in which the inability to master early regulatory skills hindered the development of more advanced regulatory skills. In summary, theoretical and empirical evidence suggests that self-regulation exhibits heterotypic continuity. That is, behavioral manifestations of self-regulation change across development despite the persistence of the construct. However, whether self-regulation demonstrates heterotypic continuity is ultimately an empirical question, and limited empirical work has tested this possibility. Previous empirical work examining this question has been limited to specific components of self-regulation. It is thus important for empirical studies to investigate whether the broader self-regulation construct demonstrates heterotypic continuity.

If self-regulation shows heterotypic continuity, there are important measurement implications. Using the same measure across development may not reflect the same construct at different ages (Widaman et al., 2010). That is, a given measure may not be developmentally appropriate or construct-valid at all ages, such that scores on the same measure across time may reflect differences in the measure's meaning, rather than real change in an individual's level of self-regulation (Petersen et al., 2016). Consequently, accounting for heterotypic continuity of self-regulation may require using different measures across time (Widaman et al., 2010), because children are expected to display different behaviors at different ages for the same underlying construct of self-regulation (Bates & Novosad, 2005). Studies examining children's self-regulation development should account for these changes by using different measures across ages (Petersen et al., 2016, 2020). Using different, age-relevant measures over time provides more accurate growth estimates, at the group- and person-level than approaches that ignore heterotypic continuity (Chen & Jaffee, 2015; Petersen et al., 2018; Petersen, LeBeau, et al., 2021).

Although considerable research has examined different measures of self-regulation at different ages in recognition of its heterotypic continuity (e.g., Chang et al., 2015), no prior work has examined individuals’ self-regulation growth using different measures across development. Thus, we use developmental scaling to estimate children’s growth to better understand self-regulation development.

### 1.1 The present study

In the present study, we apply a domain-general hierarchical model of self-regulation that includes multiple related regulatory processes: inhibitory control, delayed gratification, sustained attention, and executive functions, consistent with Berger (2011), McClelland et al. (2010), Nigg (2017), and others. Given task impurity (McCoy, 2019), we use multiple measures and assessment methods (i.e., performance-based measures and questionnaires) for more robust estimates of children’s self-regulation (Gagne et al., 2021). Measures were chosen because they reflect a broad range of regulatory skills, including both lower-level and higher-level processes. Moreover, these measures are commonly used in studies examining development of self-regulation and have shown reliability and validity within the age range of the present study (Carlson, 2005; Petersen et al., 2016). To account for heterotypic continuity, we use some common measures across adjacent ages to capture the core self-regulation phenotype on the same scale, and some different measures across ages to capture the changing manifestation.

To date, no studies have developed a model to account for heterotypic continuity of self-regulation. In the present study, we use a Bayesian longitudinal mixed model for developmental scaling to link differing measures of self-regulation across ages onto the same scale. This model allows charting children’s growth across ages 3–7 years. Bayesian longitudinal mixed modeling is ideal for this developmental scaling scenario. Bayesian implementation relies upon conditional logic for the model structure which simplifies how the model is framed and allows information to be borrowed across multiple measurements of people’s standing on the latent self-regulation construct. The Bayesian model also leverages all available data to estimate latent growth curves. Furthermore, we have prior scientific knowledge of developmental scaling, and we apply this knowledge in the structure of the Bayesian model. By borrowing strength across the abundant data available per participant and by utilizing the informative prior structure, we can obtain reliable and interpretable estimates for all parameters in the model. Moreover, Bayesian item response theory (IRT) has benefits over frequentist approaches to IRT, including estimation for moderate and smaller sample sizes (Fox, 2010) and improved estimation of parameters (Natesan et al., 2016), van de Schoot et al. (2014) and Oldenhinkel (2016) provide accessible discussions of Bayesian approaches in developmental science.

As a criterion-related test of validity of our approach to developmental scaling, we examine children's trajectories in relation to adjustment outcomes, including school readiness (math and reading skills) and externalizing problems. We examine self-regulation development across ages 3–7 years because self-regulatory processes show rapid development in early childhood (Greene, 2017; Montroy et al., 2016). Moreover, the transition to formal education may represent a key developmental period when self-regulatory processes become especially important for school readiness as well as for future learning and achievement (Blair & Raver, 2015; Mazzocco & Kover, 2007). Consistent with prior studies of self-regulation, we hypothesized that children’s growth trajectories would show rapid development around age 3 and decelerate around age 7. We expected that measures would change in their strength of association with the self-regulation construct over time, consistent with heterotypic continuity (Petersen et al., 2016; Petersen, LeBeau, et al., 2021). Additionally, we hypothesized that boys would show poorer self-regulation, on average, than girls (Kochanska et al., 2001; Matthews et al., 2009, 2014; McClelland et al., 2016).
et al., 2007). We also hypothesized that children's developmentally scaled self-regulation would be associated with school readiness and externalizing problems. Specifically, we hypothesized that lower levels of self-regulation would be associated with poorer math and reading skills, as well as externalizing problems.

2 | METHOD

2.1 | Participants

Participants consisted of a community sample of young children (N = 108) and their families, who took part in an ongoing accelerated longitudinal study. Children were recruited from 2018 to 2022 at one of the following ages (cohorts): 36 (n = 29), 45 (n = 29), 54 (n = 21), or 63 (n = 29) months and were assessed every 9 months over four time points (see Figure 1). The full sample of children spanned 3–7.5 years of age. Participants were recruited through a biomedical registry of children who had well-child checkups at University of Iowa Hospital and Clinics, university email listservs, and from advertisements and in-person recruitment activities at their school or preschool, Women, Infants, and Children (WIC) programs, pediatricians’ offices, and community events. Exclusion criteria were: the child's primary caregiver did not speak English, or the child did not have a permanent guardian, did not have normal or corrected-to-normal vision and hearing, or was not capable of following basic instructions in English.

Figure 2 depicts the flow of participants from screening to consent. The final sample consisted of 108 children (M = 4.82 years, SD = 1.22 years; 51 girls), their primary caregiver, the primary caregiver’s parenting partner (as applicable), and a teacher/secondary caregiver (e.g., nanny, babysitter, or someone else who knew the child well). Participant demographics are detailed in Appendix S1. The ethnic composition of children in the sample was: 67.6% Non-Hispanic White, 7.4% Black or African American, 6.5% Asian, 7.4% Hispanic or Latino, 5.6% Multiracial, and 5.6% other. Participants received money and small gift bags as compensation for participation.

Extent of missingness for each model variable is in Table S1. Reasons for missingness and tests of systematic missingness are in Appendix S2.

The number of children with self-regulation scores by wave is depicted in Figure S1.

2.2 | Procedure

At each time point (i.e., every 9 months for four time points), the child and their primary caregiver completed two lab visits, approximately 1 week apart. The primary caregiver completed electronic questionnaires during both lab visits or from home. Additionally, the primary caregiver’s parenting partner and the child’s teacher/secondary caregiver were emailed or mailed the questionnaires to complete.

2.2.1 | Lab visit 1

The first lab visit lasted approximately 120–180 min (M = 150.78, SD = 20.31). During this visit, the child and their primary caregiver came to the lab. The child completed a series of tasks with an experimenter, including self-regulation tasks, parent-child interaction tasks, standardized assessments of academic achievement, and other assessments, while the primary caregiver completed questionnaires about their child.

2.2.2 | Lab visit 2

The second lab visit lasted approximately 70–120 min (M = 86.96, SD = 18.97). During this visit, the child completed computerized tasks, including a go/no-go (Fish/Sharks) and stop-signal (Food Finder) task, while wearing an electroencephalography cap and brainwaves were recorded. The primary caregiver completed additional questionnaires.

2.3 | Measures

The present study is part of a larger study, the School Readiness Study. Measures and hypotheses for the School Readiness Study were pre-registered: https://osf.io/jzxb8. Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are
208 contacted to screen

5 refused the screen:
- 1 was "too busy"
- 1 was "not interested"
- 3 refused for "other" reasons

203 screened

14 ineligible:
- 10 children were age 5 years 4 months or older
- 1 child did not speak and understand English
- 3 children did not have a permanent guardian

189 eligible

17 not interested:
- 9 said they were "not interested"
- 2 were "too busy"
- 2 moved or relocated
- 1 said that doing EEG would be difficult for child
- 3 were not interested for "other" reasons

172 interested

22 not (yet) contacted to enroll:
- 18 have not yet aged into scheduling window
- 4 due to COVID-19 suspending lab operations

150 contacted to enroll

22 not (yet) enrolled:
- 5 are currently negotiating scheduling
- 13 were unable to be contacted
- 4 for unknown reasons

128 enrolled

20 not (yet) consented:
- 7 are scheduled
- 2 are to be scheduled
- 3 were unable to be contacted
- 8 for unknown reasons

108 consented

FIGURE 2 Participant flowchart. Note. EEG = "electroencephalography"

2.3.1 | Self-regulation

Measures of self-regulation included 15 laboratory tasks and two questionnaires. We assessed four facets of self-regulation: inhibitory control, delayed gratification, sustained attention, and executive functions. We assessed inhibitory control with Bear/Dragon, Day/Night, Fish/Sharks, Food Finder Stop-Signal Task, Grass/Snow, Hand Game, Knock/Tap, Less is More, Peg Tapping, Shape Stroop, Simon Says, and the Children's Behavior Questionnaire. We assessed delayed gratification with Gift Delay, a self-imposed waiting task, and Snack Delay. We assessed sustained attention with Token/Bead Sort. We assessed various executive functions, such as inhibition, shifting, and working memory, by parents’ reports on the Behavior Rating Inventory of Executive Function (BRIEF). The BRIEF is a widely used questionnaire that was designed to assess executive functions within the context of children’s everyday environment. Except for computer-scored tasks (Fish/Sharks and Stop-Signal Task) and Token/Bead Sort, children’s
performance on tasks was scored after the lab visit from video recording. All scored cases were double-coded to evaluate inter-rater reliability via intraclass correlation. Raters met to resolve any large discrepancies between raters’ codes. Estimates of reliability (inter-rater, internal consistency, cross-time stability) are in Table S5. Scores were averaged across raters. Estimates of time to administer each task are in Table S6.

For developmental scaling, scores of each self-regulation measure were converted to proportion of maximum (POM) scores to have the same possible range (0–1), with higher scores reflecting greater self-regulation. Proportion scores are widely recommended by longitudinal researchers for studying growth with different measures (Little, 2013; Moeller, 2015). For measures that had a minimum and maximum possible score, the POM score reflected the proportion of the maximum possible score. For measures that did not have a minimum or maximum possible score (Stop-Signal Task and Token/Bead Sort), the POM score reflected the proportion of the maximum observed score. POM scores were calculated as: 

\[ \text{POM score} = \frac{\text{score} - \text{minimum}}{\text{maximum} - \text{minimum}} \]

where minimum and maximum were the minimum and maximum possible or observed score. Tasks (Token/Bead Sort; Stop-Signal Task) and questionnaires (BRIEF) were adapted to accommodate the developmental capacity of the child and the changing expression of self-regulation with age.

### 2.3.2 School Readiness

**Woodcock Johnson IV–Tests of Achievement**

The Woodcock Johnson IV–Tests of Achievement (Schrank et al., 2014, 2018) assess academic achievement. Children completed two subtests to assess their early (pre-)reading and math skills: Letter-Word Identification and Applied Problems, respectively. Letter-Word Identification assesses word identification skills and reading-writing ability. The child was asked to identify letters and eventually asked to read aloud individual words. Applied Problems assesses quantitative ability. The child was asked to analyze and solve math problems. Items were scored on accuracy (1 = correct, 0 = incorrect). Raw scores (i.e., number of correct responses) were used.

### 2.3.3 Externalizing behavior

**Achenbach System of Empirically Based Assessment**

The Achenbach System of Empirically Based Assessment (ASEBA) assesses children's emotional and behavioral problems. Items were rated on a 3-point Likert scale according to how well the item described the child (0 = not true, 1 = somewhat or sometimes true, 2 = very true). Multiple versions were used based on the child's age and rater type. Parents completed the Child Behavior Checklist 1.5–5 (Achenbach & Rescorla, 2000) if the child was 3–5 years old or the Child Behavior Checklist 6–18 (Achenbach & Rescorla, 2000) if the child was 6–7 years old. Secondary caregivers completed the Caregiver–Teacher Report Form (Achenbach & Rescorla, 2001) if the child was 3–5 years old or the Teacher’s Report Form (Achenbach & Rescorla, 2001) if the child was 6–7 years old. Scores on the Externalizing scale were used. Externalizing problem scores were then converted to POM scores to put scores from different ASEBA measures onto a metric with the same possible range.

### 2.4 Statistical analysis

We used different measures of self-regulation across ages to account for heterotypic continuity.

#### 2.4.1 Exploratory factor analysis

We first examined whether measures’ scores were able to be modeled with item response modeling by examining their scores in exploratory factor analysis (EFA). We conducted EFA with maximum likelihood estimation using the psych 2.1.9 package (Revelle, 2020) in R 4.1.2 (R Core Team, 2021).

#### 2.4.2 Developmental scaling

We used developmental scaling to link scores from the different measures across ages onto the same scale. In this way, we could make meaningful comparisons of scores from different measures across ages and estimate accurate trajectories of children’s self-regulation growth. A detailed description of the developmental scaling approach is in Appendix S4. To perform developmental scaling, we used a two-parameter Bayesian longitudinal item response model in a mixed modeling item response theory (IRT) framework. Such a model allows us to simultaneously account for heterotypic continuity of self-regulation using different measures across time and to model children’s self-regulation trajectories. Given the numerous measures assessed, the many items, and the varying number of items per measure, we used measure-level (POM) scores (rather than item- and trial-level scores) as the “items” in the item-response model. The model linked scores from measures across all ages in the same model, known as concurrent calibration. Concurrent calibration accounts for within-person dependence of scores across time and results in more precise and stable estimates than two-stage calibration in which separate models are fit (Kolen & Brennan, 2014; McArdle et al., 2009). The two-parameter item response model estimates two parameters: easiness (\(\xi\); the inverse of difficulty) and discrimination (\(\alpha\)). The item’s easiness parameter is the expected score on an item at a given level of the construct (Bürkner, 2020). The item’s discrimination parameter is how strongly the item is associated with the construct. In our study, easiness and discrimination provide information about the functioning and usefulness of each measure—and the whole measurement scheme—at a given age.

In the present study, the self-regulation scores were continuous proportion scores that ranged from 0 to 1. Because some scores were zero or one (especially one; see Figure S2), we used a zero-one-inflated beta distribution for the outcome variable (Ospina & Ferrari, 2012). A tra-
dential beta distribution is a continuous probability distribution that does not allow zeros or ones. A zero-one-inflated beta distribution is a mixed continuous-discrete probability distribution, which includes a continuous beta distribution (to capture the continuous distribution of proportion scores) and a Bernoulli distribution (to capture zeros and ones).

We performed the developmental scaling, estimation of growth curves, and tests of differential item (measure) functioning (DIF) in the same model. A given child had up to four time points. Thus, a quadratic was the most complex polynomial of nonlinear growth we could estimate for children’s trajectories that still allow measurement error. Because of prior work demonstrating that growth in self-regulation is non-linear, such that children showed faster growth in preschool than elementary school (Montroy et al., 2016), we modeled children’s growth in self-regulation with a quadratic term. We modeled random intercepts and random linear and quadratic slopes to allow each child to differ in their starting point, form of growth, and curvature. Age in years was centered to set the intercepts at age 3. We included the child’s sex (female = 1, male = 0) as a predictor of the intercepts and slopes.

To examine DIF across ages, we estimated a random intercept and slope for measure (in addition to the terms described above) to allow the item parameters for each measure to differ by age. This allowed us to examine the extent to which the measures changed across development in easiness and discrimination.

Our model had no missing data in the predictors (age and sex); missingness was only in the outcome (scores on self-regulation measures). Mixed models handle missing data in the outcomes. Mixed models provide valid inferences if the data are missing at random or completely at random. Because much of our missingness was due to COVID-19, and we observed limited patterns of systematic missingness as a function of demographics, predictors, or outcomes with small effect sizes, we felt this modeling approach was appropriate. Moreover, researchers have argued against using multiple imputation in longitudinal designs that use mixed models because multiple imputation can lead to unstable estimates (Twisk et al., 2013).

Developmentally scaled self-regulation factor scores were estimated for each child at each of their measurement occasions. This allowed each child to have a different factor score at each of their measurement occasions.

We fit the Bayesian longitudinal mixed model using the brms package 2.16.3 (Bürkner, 2017) in R, which uses the RStan 2.21.1.3 (Stan Development Team, 2020a) interface to Stan 2.21.0 (Stan Development Team, 2020b) for Bayesian modeling. The model included eight chains and 10,000 iterations.

2.4.3 Sensitivity analyses

We conducted sensitivity analyses using two models, as described in Appendix S5. First, we examined a model that imposed approximate longitudinal measurement invariance. Second, we fit a model that excluded scores for a given measure at ages when the proportion of maximum score on that measure could reflect ceiling effects (i.e., mean proportion score > 0.90).

2.4.4 Self-regulation predicting outcomes

To examine whether children’s developmentally scaled self-regulation factor scores were associated with externalizing problems and school readiness, we used multiple regression with a cluster variable specifying the participant (i.e., clustered regression). Clustered regression accounts for the longitudinal dependency in the data. Clustered regression models were fit using the rms package 6.2 (Harrell, 2015) in R that calculates robust standard errors using a Huber-White sandwich estimator of the covariance matrix (Huber, 1967; White, 1980). Power analyses of our ability to detect associations predicting outcomes are in Appendix S6.

3 RESULTS

3.1 Descriptive statistics and correlations

Average proportion self-regulation scores by measure and age are shown in Figure 3. Bivariate correlations and descriptive statistics of model variables are in Tables S2–S4. Partial correlations controlling for age are in Tables S7–S9. Although there were exceptions, self-regulation scores were largely inter-correlated across measures. Moreover, a combination of self-regulation scores across measures showed strong internal consistency (ω = 0.94; see Table S5).

3.2 Exploratory factor analysis

We examined scores from the self-regulation measures in EFA. Results of the EFA are in Table S10. A one-factor model accounted for 35% of the variance. All but three measures’ scores (Food Finder Stop-Signal Task, mothers’ and fathers’ ratings on the BRIEF, and mothers’ and secondary caregivers’ ratings on the CBQ) had a standardized factor loading above 0.40. In a two-factor model, the second factor accounted for 8% of the variance. Moreover, all measures that had loadings above 0.40 on the second factor were questionnaire measures, suggesting that the factor that accounted for the most variance after the primary factor was a method factor. Findings remained consistent when controlling for the child’s age. Thus, although the self-regulation measures clearly assessed multiple dimensions, a single factor accounted for considerable variance, and accounted for considerably more variance than the second factor. Based on this evidence, the primary factor appeared to reflect a meaningful operationalization of self-regulation. Given our goals to examine children’s self-regulation development by aggregating scores from multiple methods, we conducted item response modeling with a single factor.
FIGURE 3  Average proportion self-regulation scores by measure and age. Note. The bars correspond to waves 36, 45, 54, 63, 72, 81, and 90 months. Scores on some performance-based measures were largely missing at 90 months due to COVID-19 (see Appendix S1). “BRIEF” = Behavior Rating Inventory of Executive Function; “CBQ” = Children’s Behavior Questionnaire; “Secondary” = secondary caregiver.

3.3 Bayesian longitudinal item response model

We fit a Bayesian longitudinal item response model in a mixed modeling framework to perform the developmental scaling, estimation of growth curves, and tests of DIF. All Gelman-Rubin diagnostic criteria for convergence (R) were 1.00, and visual examination of trace plots showed that all chains adequately mixed, indicating that the model converged. The R² from leave-one-out (LOO) cross validation was 0.47, indicating that the model explained nearly half of the variance in children’s scores on the self-regulation measures across time. Measures’ easiness and discrimination are in Figure 4. All measures showed significant associations with the self-regulation construct; the 95% credible interval of the discrimination estimates did not include zero. Measures’ empirical characteristic curves are in Figure S4. Model results are in Table S11.

3.3.1 Differential item (measure) functioning

Tests of differential item functioning are described in detail in Appendix S5. Changes in item easiness and discrimination are depicted in Figure S3. Four measures became easier—relative to the same ability—with age: Fish/Sharks, Gift Delay, Snack Delay, and mothers’ ratings on the BRIEF–P. All measures except Fish/Sharks and Simon Says showed decreases in discrimination with age, consistent with heterotypic continuity. Effect sizes of non-invariance were small, and measures remained strongly discriminating across ages, so we proceeded to interpret the growth curves and predictors of the trajectories.

3.3.2 Form of growth

There was a positive mean of the quadratic slope. As depicted in Figure 5, children showed rapid growth in self-regulation from ages 3 to 6, after which growth slowed and leveled off.

3.3.3 Sex-related differences

We examined whether the child’s sex predicted differences in intercepts and slopes. As depicted in Figure 5, girls showed higher intercepts...
3.4 Validation of developmentally scaled self-regulation scores

As a validation of the developmentally scaled self-regulation scores, we examined whether the developmentally scaled self-regulation scores were associated with theoretically relevant outcomes, including externalizing problems and school readiness.

3.4.1 Predicting externalizing problems

Model results from the regression models of developmentally scaled self-regulation scores predicting externalizing problems are in Table S12–S13. Self-regulation was moderately negatively associated with externalizing problems in a model without covariates ($\beta = -0.28$). However, the association became marginally significant when controlling for the child’s age ($\beta = -0.13$).

3.4.2 Predicting school readiness

Model results from the regression models of developmentally scaled self-regulation scores predicting school readiness are in Table S14–S16. Self-regulation was moderately to strongly positively associated with reading ($\beta = 0.27$) and math ($\beta = 0.51$) skills, controlling for age, grade, and SES. Moreover, self-regulation remained associated with math skills ($\beta = 0.35$) but was marginally associated with reading skills ($\beta = 0.19$), when controlling for intelligence.

3.4.3 Sensitivity analyses

The sensitivity analyses are described in detail in Appendix S5. Findings were substantially similar when examining the model with approximate
longitudinal invariance imposed, and when examining the model with potential mean-level ceiling effects. Both models yielded similar trajectories of self-regulation. In addition, criterion-related tests yielded similar results. Self-regulation was negatively associated with externalizing problems, controlling for age, and was positively associated with math (but not reading) skills when controlling for age, grade, SES, and intelligence.

4 | DISCUSSION

Self-regulation is thought to demonstrate changes in its behavioral manifestation across development. Researchers have argued that children develop lower-level processes in early childhood (e.g., early forms of inhibitory control and delayed gratification) and higher-level processes (e.g., executive functions) in later childhood, which are then integrated with lower-level processes to form a hierarchically organized regulatory system, in later childhood (Greene, 2017; Montroy et al., 2016). However, limited empirical work has examined whether self-regulation shows heterotypic continuity, despite evidence for heterotypic continuity of specific components of self-regulation (Chang et al., 2015; Geeraerts et al., 2021; Petersen, Bates, et al., 2021; Petersen et al., 2016; Putnam et al., 2008; Zimmermann & Iwanski, 2014). We found evidence of heterotypic continuity of self-regulation in the present study, such that measures changed in their strength of association with the latent construct across ages. Prior work has not accounted for heterotypic continuity of self-regulation when studying children’s growth curves. In the present study, we accounted for heterotypic continuity of self-regulation based on theoretical, methodological, and analytical considerations. We followed theoretical conceptualizations of self-regulation as encompassing multiple processes, including inhibitory control, delayed gratification, sustained attention, and executive functions (Baumeister & Vohs, 2004; Berger, 2011; Blair & Raver, 2015; Gagne et al., 2021; McClelland et al., 2010, 2015; Nigg, 2017). Methodologically, we used different measures across ages to account for the changing nature of the construct, while using some common measures across adjacent ages to ensure scores could be linked across ages. Analytically, we used developmental scaling to link scores from different measures across ages onto the same scale so we could examine children’s self-regulation growth. We describe our approach to developmental scaling below.

4.1 Developmental scaling

We used a Bayesian longitudinal item response modeling approach to developmental scaling, in which children’s proportion scores from each measure were used as items in the model. The model simultaneously estimated item response model parameters and longitudinal growth curves using a concurrent calibration approach, in which scores from measures across all ages were linked in the same model (Kolen & Brennan, 2014). Our approach to developmental scaling is consistent with prior work that has linked different measures of cognitive ability across the lifespan (McArdle et al., 2009). When concurrent calibration is used, people’s estimated construct levels are on the same scale across ages if IRT assumptions are met (Kolen & Brennan, 2014). Although the measures likely assessed multiple dimensions, EFA demonstrated that the data were likely uni-dimensional enough for IRT, thus providing greater confidence in children’s estimated growth curves. Moreover, the developmentally scaled scores showed strong internal consistency (ω = 0.94) and cross-time stability (r = 0.68), providing evidence that they reflected meaningful operationalization of children’s self-regulation.

4.2 Form of growth

Based on model-implied trajectories from the longitudinal item response model, children showed rapid growth in self-regulation from ages 3 to 6, after which growth slowed and leveled off from ages 6 to 7. This pattern of growth is consistent with concomitant changes in brain development. Brain size increases four-fold during the preschool period, reaching approximately 90% of the adult volume by age 6 (Brown & Jernigan, 2012; Stiles & Jernigan, 2010). Moreover, children between ages 3 and 6 showed marked improvement in working memory and inhibitory control abilities, which are thought to depend on the dorsolateral prefrontal cortex (DL-PFC; Berger, 2011; Diamond, 2001, 2002). During this time, the DL-PFC undergoes important changes, including rapid decreases in neuronal density between 2 and 7 years of age and expansion of dendritic trees in layer III pyramidal cells between 2 and 5 years of age (Diamond, 2001).

Consistent with prior work, girls had modestly higher mean self-regulation at age 3 compared to boys (Kochanska et al., 2001; Matthews et al., 2009, 2014; McClelland et al., 2007). Boys and girls did not significantly differ in their slopes, but boys appeared to nearly catch up to girls by age 7.
4.3 | Validation of developmentally scaled self-regulation scores

As a criterion-related test of the validity of our approach to developmental scaling, we examined children's developmentally scaled self-regulation scores in relation to adjustment outcomes, including externalizing problems and school readiness (math and reading skills). We hypothesized that lower levels of self-regulation would be associated with externalizing problems and poorer math and reading skills. We found that lower levels of self-regulation were moderately associated with externalizing problems. However, the association was only at trend level when controlling for the child’s age, which was inconsistent with hypotheses. It is possible that developmental improvements in self-regulation could account for normative age-related reductions in externalizing problems. Alternatively, developmental changes in self-regulation may reflect other processes, such as language development (Petersen & LeBeau, 2021), that lead to age-related reductions in externalizing problems. Or, perhaps our study was under-powered to detect the association, given meta-analytic evidence that the effect size of self-regulation on externalizing problems is small (Berger & Buttelmann, 2021). Future work will be important to examine the role of self-regulation in the development of externalizing problems.

Consistent with hypotheses, lower levels of self-regulation were associated with poorer (pre-)reading and math skills. The effect size was large, and the association held when controlling for covariates (age, grade, and SES). Moreover, the association held with math (but not reading) skills when controlling for the child’s intelligence. Thus, performance on the self-regulation measures does not appear to merely reflect better comprehension of task rules (likely influenced by language ability, i.e., a dimension of intelligence). The finding that self-regulation was strongly associated with math skills above and beyond intelligence supports the possibility raised by prior research that self-regulation plays an important role in development of school readiness (e.g., Blair & Raver, 2015; Eisenberg et al., 2010; Ursache et al., 2012). The finding that self-regulation was more strongly associated with academic skills than with externalizing problems is consistent with prior research showing that preschoolers' executive function predicted mathematic skills but not aggression (Sasser et al., 2015). In sum, the criterion-related association between children's developmentally scaled self-regulation scores and their school readiness provides further support for the validity and utility of our approach to developmental scaling.

4.4 | Implications for understanding development of self-regulation

Theory (Berger, 2011; Kopp, 1982; McClelland et al., 2010), empirical work (Chang et al., 2015; Geeraerts et al., 2021; Petersen, Bates, et al., 2021; Putnam et al., 2008; Zimmermann & Iwanski, 2014), meta-analysis (Petersen et al., 2016), and findings in the present study collectively provide considerable evidence that self-regulation changes in its behavioral manifestation across development. Our developmental scaling approach of self-regulation replicates and extends prior literature examining development of self-regulation. Consistent with previous studies, we observed rapid growth in self-regulation between ages 3 and 6, which slowed and leveled off between ages 6 and 7 (Greene, 2017; Montroy et al., 2016). Moreover, we observed robust associations with school readiness outcomes. Crucially, developmental scaling simultaneously accounted for heterotypic continuity and charted children's growth over time. Accounting for changing behavioral manifestations at different ages provides a more accurate understanding of self-regulation development across early childhood than previous models, and our approach can be extended to adolescence and adulthood. Indeed, research has shown that adolescents show a marked increase in cognitive flexibility and improvements in planning, organizing, and strategic thinking skills, which carry into adulthood (Anderson, 2002; Greene, 2017). Moreover, Zimmermann and Iwanski (2014) found differences in emotion-regulation strategies from early adolescence to middle adulthood, consistent with heterotypic continuity. Nevertheless, self-regulation or other constructs do not need to show heterotypic continuity for our modeling approach to be useful for charting children’s growth.

4.5 | Strengths

The study had several strengths. First, the study was longitudinal, which allowed examining children's self-regulation development. Second, we assessed multiple facets of self-regulation to be consistent with theory and prior research on the structure of self-regulation. Third, our assessment of self-regulation included multiple measurement methods including performance-based assessment and questionnaires to reduce common method variance. Fourth, we included multiple informants, including mothers, fathers, and teachers or other caregivers to gain a more accurate estimate of children’s real-world functioning. Fifth, we used developmental scaling to link differing measures across ages onto the same scale, which allowed examining children’s absolute growth in self-regulation. Our multi-wave, multifacet, multi-method, multi-measure, multi-rater, developmental scaling approach is the most comprehensive to date for assessing development of self-regulation. Prior research using developmental scaling has used primarily dichotomous or polytomous items. The Bayesian approach we used successfully handled a moderate sample size when fitting longitudinal item response models with continuous data, which potentially increases its practicality for use in developmental research. We also make our data and analysis scripts freely available to promote dissemination.

4.6 | Limitations

The study also had limitations. First, the sample size may limit our ability to detect smaller effects. Second, the study was observational, so we cannot make causal inferences. Third, there was considerable missing data at later ages, including limited performance-based assessments at participants’ fourth time points, largely due to COVID-19. In addition,
a modest number of children had self-regulation scores at later ages due to the accelerated nature of the longitudinal design. Moreover, many measures showed increases in easiness and/or decreases in discrimination across ages, and several performance-based tasks showed ceiling effects at later ages, which may have contributed to their somewhat weaker discrimination. Thus, we have less confidence about children's level of self-regulation at later ages in our study (6–7 years of age). Nevertheless, researchers have argued that establishing longitudinal measurement invariance is unnecessary when the construct shows heterotypic continuity (Edwards & Wirth, 2012; Knight & Zerr, 2010; Petersen et al., 2020). Our model accounted for changes in measures' easiness and discrimination. Moreover, the form of growth we observed aligned with prior findings, and it was consistent even when we imposed approximate longitudinal measurement invariance and removed scores at ages with potential mean-level ceiling effects, which increases confidence in our findings.

Another limitation relates to assumptions regarding self-regulation. We modeled self-regulation using item response modeling, which assumes there is a latent factor (i.e., reflective construct) that influences scores on all self-regulation measures. In support of a reflective model of self-regulation, we found that the measures assessing various components of self-regulation (i.e., inhibitory control, delayed gratification, sustained attention, and executive functions) were robustly correlated, and a one-factor model captured a large portion of the variance in scores across measures. We acknowledge, however, that we may not have assessed all relevant components, for instance emotion regulation, which could limit the interpretation of the latent factor and generalizability of findings. Nevertheless, our assessment included many measures of multiple facets, providing a more comprehensive assessment than prior research examining self-regulation growth, which has mainly examined one or a few measures and one or a few facets (Montroy et al., 2016; Sulik et al., 2010). Alternatively, emerging research suggests that self-regulatory processes may be operationalyzed using formative constructs (Camerota et al., 2020; Willoughby et al., 2017), whereby self-regulation is defined as the summation of relevant measures, rather than as their shared variance. Future research should examine how to operationalize self-regulation, including its structure, whether it is a reflective or formative construct, and how its structure changes with development. Better developmental models of self-regulation that account for changes in its structure will lead to better understanding of how self-regulation develops across the lifespan.

5 | CONCLUSION

Self-regulation is thought to change in its behavioral manifestation across development. We accounted for heterotypic continuity of self-regulation by using different, theoretically relevant measures across ages to account for the changing manifestation of the construct. We used developmental scaling to link scores from differing measures across ages onto the same scale so we could examine children's self-regulation growth. Children's developmentally scaled self-regulation scores were validated against their externalizing problems and school readiness, including math and reading skills. Findings suggest that developmental scaling permits studying the development of self-regulation across lengthy spans and key developmental transitions. Future research should adapt measurement schemes to be developmentally appropriate and valid across ages. Developmental scaling may enable studying development of self-regulation and other constructs across the lifespan.

ACKNOWLEDGMENTS

The present study was approved by the University of Iowa Institutional Review Board (Study #: 201708761). The project was funded by Grants HD098235 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), T32GM108540 from the National Institute of General Medical Sciences (NIGMS), and UL1TR002537 from the National Center for Advancing Translational Sciences (NCATS).

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/5xnrh

ORCID

Alexis Hosch https://orcid.org/0000-0003-2874-2340
Jacob J. Oleson https://orcid.org/0000-0001-6343-3274
Jordan L. Harris https://orcid.org/0000-0002-1335-9448
Isaac T. Petersen https://orcid.org/0000-0003-3072-6673

REFERENCES


