Is It Really Self-Control? Examining the Predictive Power of the Delay of Gratification Task

Angela L. Duckworth[1], **Eli Tsukayama**[1], and **Teri A. Kirby**[2]
[1]University of Pennsylvania, Philadelphia USA

[2]University of Washington, Seattle, USA

Abstract

This investigation tests whether the predictive power of the delay of gratification task (colloquially known as the "marshmallow test") derives from its assessment of self-control or of theoretically unrelated traits. Among 56 school-age children in Study 1, delay time was associated with concurrent teacher ratings of self-control and Big Five conscientiousness—but not with other personality traits, intelligence, or reward-related impulses. Likewise, among 966 preschool children in Study 2, delay time was consistently associated with concurrent parent and caregiver ratings of self-control but not with reward-related impulses. While delay time in Study 2 was also related to concurrently measured intelligence, predictive relations with academic, health, and social outcomes in adolescence were more consistently explained by ratings of effortful control. Collectively, these findings suggest that delay task performance may be influenced by extraneous traits, but its predictive power derives primarily from its assessment of self-control.

Keywords

Impulses to seek pleasure and avoid displeasure are essential to survival, but impulses to pursue here-and-now rewards are often at odds with countervailing, goal-directed processes. Freud (1920) speculated that the ability to exercise self-control in such dilemmas is critical to healthy psychological development: "[U]nder the influence of the instructress Necessity," children must learn to "renounce immediate satisfaction, to postpone the obtaining of pleasure, to put up with a little unpleasure" (p. 444). Early attempts to measure individual differences in self-control used Rorschach and related projective tests (e.g., Singer, 1955), measures later found wanting in both face validity (Mischel, 2007) and predictive validity (Lilienfeld, Wood, & Garb, 2000). Subsequently, in a process entailing years of iterative prototyping and refinement, Mischel developed the delay of gratification task. Better known colloquially as the "marshmallow task," this paradigm quantifies self-control as the ability to wait for a preferred treat (e.g., two marshmallows later) while forgoing a less preferred reward (e.g., one marshmallow right now).

The delay of gratification task appears face-valid and predicts an array of positive academic, social, and health outcomes later in life (Ayduk et al., 2000; Mischel, Shoda, & Peake, 1988;

Corresponding Author: Angela Duckworth, Department of Psychology, University of Pennsylvania, 3701 Market Street, Suite 215, Philadelphia, PA 19104, USA. duckworth@psych.upenn.edu.

Shoda, Mischel, & Peake, 1990), but does it really assess self-control? Contrariwise, might delay time in this task reflect unrelated traits such as intelligence or attraction to rewards? If so, do such traits constitute third-variable confounds responsible for the delay task's predictive power? Surprisingly, given the prominence of the delay task in both scholarly research and public debate (Lehrer, 2009; Mischel & Brooks, 2011; Public Broadcasting Service, 2011), straightforward questions clarifying its interpretation have not been directly addressed in prior research. In the current investigation, school-age children (in Study 1) and preschool children (in Study 2) completed the delay of gratification task. Separately, children in both studies completed standard tests of intelligence, and adult informants provided ratings of their personality and motivation. We examined these data for evidence of convergent validity with concurrent informant ratings of self-control, discriminant validity vis-à-vis intelligence and reward-related impulses, and incremental predictive validity over and beyond possible confounding variables for longitudinally measured outcomes.

Delay of Gratification and Intelligence

Direct evidence on how intelligence relates to performance in the delay of gratification task is lacking, but there is sufficient indirect evidence to warrant speculation about intelligence as a confound. Like self-control, intelligence predicts academic, social, health, and economic well-being later in life (Borghans, Duckworth, Heckman, & ter Weel, 2008), and at least some studies have found that more intelligent children are rated as more self-controlled by parents and other informants (e.g., Moffitt et al., 2011; Olson, Sameroff, Kerr, Lopez, & Wellman, 2005). More intelligent children express preferences for larger, later rewards over smaller, sooner rewards (Lesure, 1977; Mischel & Metzner, 1962) as do more intelligent adolescents (Block & Funder, 1989; Olson, Hooper, Collins, & Luciana, 2007), and adults (Shamosh et al., 2008; Shamosh & Gray, 2008), though *preferring* to delay gratification and *sustaining* this commitment in the face of temptation are distinct psychological processes (Mischel, 2007; Reynolds & Schiffbauer, 2005). Of more direct relevance, in a sample of 95 girls and boys at the Bing Nursery School, performance in the delay task at age four strongly predicted parent impressions of intellectual competence a decade later (Mischel et al., 1988). Specifically, of 100 items in the parent-report California Child Q-Set (CCQ), the two that demonstrated the *strongest* positive associations in adolescence with preschool delay time were "is verbally fluent, can express ideas well in language" and "uses and responds to reason," rs = .48 and .47, respectively. Likewise, among 35 participants in the same sample whose parents reported their SAT scores, correlations between math and verbal SAT scores and delay time were large, rs = .42 and .57, respectively (Shoda et al., 1990). Separate research has established that associations between the SAT and tests of general intelligence are so high that some researchers use SAT scores as a proxy for IQ (Frey & Detterman, 2004).

There are at least two substantive reasons why intelligence might facilitate self-control in children and, hence, improve their performance in the delay of gratification task. First, more intelligent children may use more effective self-regulatory strategies (e.g., strategically distracting themselves) in the face of temptation (Rodriguez, Mischel, & Shoda, 1989). Put another way, it is possible that "learning to delay is intimately bound up with learning to think" (Mischel & Metzner, 1962, p. 425). Second, more intelligent individuals may be better at keeping necessarily abstract representations of distal goals in mind (Fujita & Carnevale, 2012; Shamosh et al., 2008). Consistent with this supposition, intelligence is strongly related to working memory capacity (Conway, Kane, & Engle, 2003), and deficits in working memory are related to impulsive behavior in children (Barkley, 1997) as well as preference for smaller, immediate rewards among adults (Shamosh & Gray, 2008).

Delay of Gratification and Reward-Related Impulses

Other than intelligence, reward-related impulses are the most obvious potential confound in the delay of gratification task. There is empirical evidence that processes supporting self-control capacity are distinct from those that give rise to involuntary reward-related impulses (Eisenberg, Spinrad, et al., 2004; Funder, Block, & Block, 1983; Heatherton & Wagner, 2011; Hofmann, Friese, & Strack, 2009). Individuals vary in their dispositional reactivity to rewards; some individuals are more easily excited by rewards than others (Blair, Peters, & Granger, 2004; Carver & White, 1994). But the distinction between impulses and their regulation is difficult if not impossible to ascertain from behavioral observation alone. For instance, if we see an individual resist temptation (e.g., pass on dessert), we cannot be sure whether they are exerting self-control over their impulses (e.g., to achieve a target weight) or, alternatively, are not very tempted in the first place (e.g., did not even want dessert). Likewise, the observation that a child waits longer than others in the delay paradigm is ambiguous as to whether they are exercising greater self-control or, contrariwise, are simply less tempted by the immediately available treat.

Eisenberg and colleagues have pointed out that while processes supporting voluntary self-control (e.g., strategic regulation of attention away from rewards) no doubt influence performance in the delay of gratification task,

> the reward *also* may activate impulsive reactive tendencies, such that children may be *pulled toward the reward* with little voluntary control. Therefore, children who cannot delay may be high in impulsive tendencies, whereas those who delay their gratification may be moderate or low in impulsive tendencies. (Eisenberg, Smith, Sadovsky, & Spinrad, 2004, p. 262, emphasis added)

This possibility—unexamined in prior research—muddies the interpretation of delay task performance and its predictive validity because the strength of the impulse to approach immediate reward (one marshmallow right away) is not measured separately.

Current Investigation

The current investigation uses data from two longitudinal studies to clarify the theoretical interpretation of the delay of gratification paradigm. Our results extend previous research in several ways. First, we directly test convergent associations between delay task behavior and concurrent questionnaire measures of self-control completed by adult informants. Second, we examine evidence of discriminant validity, in particular by examining associations between delay task behavior and concurrent measures of general intelligence, reward-related impulses, and other traits in omnibus taxonomies of personality (in Study 1) and temperament (in Study 2). Finally, we use statistical techniques developed for mediational analyses to test whether the predictive power of the delay task derives from self-control or potentially spurious associations with other traits (e.g., intelligence) that also forecast positive developmental outcomes.

Study 1

In Study 1, 56 fifth-grade children at a socioeconomically and ethnically diverse public magnet school completed the delay of gratification task at the start of school and were followed through the end of the academic year. We examined evidence for convergent validity with concurrent informant-report questionnaire measures of self-control and, by contrast, discriminant validity vis-à-vis theoretically unrelated constructs, including general intelligence and reward-related impulses. To situate delay task performance within an omnibus framework of personality, we also examined associations between wait time in the delay task and concurrent teacher ratings of Big Five personality, a taxonomy originally

discovered to organize traits in adults but more recently found to be as relevant in school-age children (Shiner & De Young, 2013). We anticipated evidence of convergent validity between delay task performance and Big Five conscientiousness because of substantial conceptual overlap between self-control and this broad personality dimension, which encompasses "the propensity to follow socially prescribed norms for impulse control, to be goal directed, to plan, and to be able to delay gratification and follow norms and rules" (Roberts, Jackson, Fayard, Edmonds, & Meints, 2009, p. 369; also see Eisenberg, Duckworth, Spinrad, & Valiente, 2012; McCrae & Lockenhoff, 2010) and is used interchangeably by some authors with the term self-control (e.g., Moffitt et al., 2011). By contrast, we expected delay behavior to be relatively independent of the Big Five dimensions of agreeableness, extra-version, emotional stability, and openness to experience. Finally, we examined evidence for incremental predictive validity of the delay task for final report card grades over and beyond potential confounds.

Method

Participants—Participants were 56 fifth-grade children (mean age = 10.28 years, SD = . 40) at a magnet public middle school in the Northeast. About 39% of participants were White, 31% were Black, 14% were Asian, 7% were Hispanic, and 9% were of other ethnic backgrounds; 55% were female. Fourteen percent of participants were eligible for free or reduced-price lunch based on reported household incomes lower than 185% of the national poverty level. Participants did not differ significantly from nonparticipants on age, ethnicity, gender, or lunch status, ps > 05.

Procedure and Measures—During one-on-one testing sessions conducted at their school, children completed the delay of gratification task. Separately, children completed questionnaires and intelligence tests in small groups during nonacademic periods, and homeroom teachers completed questionnaires with the children as targets. All measures were completed by the end of October 2008, and data from school records were received in July 2009.

Delay of gratification—We made two minor changes to the preschool delay of gratification paradigm (Mischel, Ebbesen, & Zeiss, 1972) to be appropriate for school-age children. First, we extended the maximum wait time to 30 min. Second, to provide a plausible context for older children, we introduced the delay task as part of "a study of food preferences." Each child was excused individually from his or her classroom and escorted by a female experimenter to a nearby room cleared of distracting stimuli and containing a desk and a bell as well as a hidden camera. Once seated at the desk, the child was left alone to complete a brief survey presenting hypothetical choices between pairs of food items (e.g., "Would you rather have a bowl of Honey Nut Cheerios or a bowl of Lucky Charms?" and "Would you rather have a 12 oz. can of Coke or a 20 oz. bottle of Coke?") and to indicate using a 7-point scale "How hungry are you right now?" Before leaving the room, the experimenter instructed the child to ring the bell to indicate that he or she had completed the survey.

Next, the experimenter showed the child a variety of snacks (e.g., cookies, chocolate candies, pretzels, grapes, chips) and asked which he or she liked best. The experimenter then asked, "Would you rather have [small amount of chosen snack] or [large amount of chosen snack]?" All of the participants preferred the larger amount. Explaining that she had to set up a task for another student, the experimenter said,

> If you wait without eating [the snack] and without getting out of your seat until I
> come back by myself, then you can have [large amount of snack]. If you don't want

> to wait, you can ring the bell at any time, and I will come in right away. But then you can only have [small amount of snack].

Once the child understood the task contingency and clearly indicated a preference for waiting, the experimenter left the room, returning and ending the task if the child rang the bell or was observed through the hidden camera to leave his or her seat or begin to eat the snack. Otherwise, the experimenter returned after 30 min and gave the child the larger snack.

Teacher ratings of self-control—With their students as targets, homeroom teachers reported on the frequency of self-control lapses in the domains of schoolwork (e.g., "This student's mind wandered when he or she should have been listening") and interpersonal relationships (e.g., "This student lost his or her temper"). Specifically, teachers rated the frequency of eight different behaviors identified in a separate sample of middle school students and teachers as failures of self-control (Tsukayama, Duckworth, & Kim, 2012) using a 6-point frequency scale ranging from *0 of the last 5 school days* to *5 of the last 5 school days*. The observed internal reliability was .89. Items were coded and averaged such that higher scores indicated higher self-control.

Reward-related impulses—Children completed three subscales from the Behavioral Inhibition System (BIS) and Behavioral Activation System (BAS) Questionnaire (Carver & White, 1994), identified by Eisenberg and Morris (2002) as appropriate for assessing reward-related impulses. These included the BAS Reward Responsiveness (e.g., "When I get something I want, I feel excited and energized") and BAS Drive (e.g., "When I want something, I usually go all-out to get it") subscales. We omitted the BAS Fun Seeking subscale because, unlike the other two BAS subscales, it does not reliably predict "positive affective responses to the signals of impending reward" (Carver & White, 1994, p. 330). While individual differences in sensitivity to punishment have a less obvious relationship with delay of gratification, we also included the Behavioral Inhibition subscale (e.g., "Criticism or scolding hurts me quite a bit"). All BIS/ BAS items were endorsed using a 5-point Likert-type scale where 5 = *agree strongly* and 1 = *disagree strongly*. The internal reliability coefficients were .65, .75, and .54 for the Reward Responsiveness, Drive, and Behavioral Inhibition subscales, respectively.

Intelligence—Children completed the Raven's Progressive Matrices and the Junior version of the Mill Hill Vocabulary Scale (Raven, Raven, & Court, 1988), widely used, untimed tests of nonverbal and verbal intelligence, respectively. Because standardized scores by age group are not published for either test, we included age as a covariate in all analyses involving nonverbal and verbal intelligence raw scores.

Big Five personality—With their students as targets, teachers completed the Big Five Inventory (BFI; John & Srivastava, 1999), which measures the personality dimensions of *conscientiousness* (e.g., "Does a thorough job"), *openness to experience* (e.g., "Is curious about many different things"), *emotional stability* (e.g., "Is relaxed, handles stress well"), *agreeableness* (e.g., "Is considerate and kind to almost everyone"), and *extraversion* (e.g., "Is outgoing, sociable") using a 5-point Likert-type response scale ranging from 5 = *agree strongly* to 1 = *disagree strongly*. Internal reliability coefficients ranged from .87 to .95 (avg. = .91).

Results and Discussion

The 10-year-old children in Study 1 waited an average of 24.50 min (*SD* = 8.52). About 41 % of children ended the task early in exchange for the smaller reward. Because data for the

remaining 59% of participants were censored (i.e., the task was ended by the experimenter at 30 min before the child voluntarily terminated), we used the Cox proportional hazards regression models. To facilitate interpretation and comparison of hazard ratios, we standardized continuous variables prior to entry as predictors in Cox models. The effect size estimates produced in these Cox models are hazard ratios, interpreted as the proportional change in the hazard (i.e., probability of ending the delay task early) associated with a one-unit change in the predictor. Consequently, hazard ratios less than one indicate a greater ability to delay, whereas hazard ratios greater than one indicate less ability to delay. As shown in Table 1, in separate Cox models, delay time was unrelated to age, gender, or free lunch status. To preserve degrees of freedom given the modest sample size, we therefore excluded these variables from subsequent analyses. However, results were virtually identical when these covariates were included (results available upon request). Because teacher ratings of students were not always independent (i.e., one teacher might rate several students), we controlled for rater in analyses with teacher ratings.

As shown in Table 1, children who were more self-controlled according to teacher ratings waited longer in the delay of gratification task, $rh = 0.63$, 95% confidence interval (CI) = [0.40, 0.98], $p = .043$. Specifically, children one standard deviation higher than average in self-control were about a third less likely to terminate the delay task as a function of time. In contrast, delay time was not associated with self-reported reward-related impulses, including reward responsiveness ($rh = 0.86$, 95% CI = [0.58, 1.28], $p = .45$) and drive, ($rh = 0.95$, 95% CI = [0.65, 1.41], $p = .81$), or hunger at the start of the task ($rh = 1.01$, 95% CI = [0.67, 1.53], $p = .97$) nor to behavioral inhibition, $rh = 1.07$, 95% CI = [0.70, 1.64], $p = .75$. Likewise, when controlling for age, delay time was unrelated to either nonverbal ($rh = 0.98$, 95% CI = [0.62, 1.54], $p = .91$) or verbal intelligence ($rh = 1.17$, 95% CI = [0.77, 1.76], $p = .46$). However, comparing means and standard deviations for the nonverbal and verbal intelligence test with published percentile norms for U.S. children taking the same tests in the mid-1980s (see Table 8 in Raven, 2000; see Table SPM9 in Raven, Raven, & Court, 2000) clearly suggested a restriction on range in our convenience sample, even when considering secular trends toward increasing intelligence scores (Raven, 2000). Thus, we did not interpret the absence of evidence for relations between intelligence and delay performance as evidence of absence in the general population.

Among Big Five personality factors, only Big Five conscientiousness was associated significantly with delay time ($rh = 0.59$, 95% CI = [0.37, 0.93], $p = .023$). Because Big Five factors were related (e.g., conscientiousness and agreeableness ratings were associated, $r = .51$, $p < .001$), we entered all Big Five factors into a simultaneous Cox regression model. Consistent with bivariate analyses, only conscientiousness predicted significant variance over and beyond other Big Five factors, $rh = 0.55$, 95% CI = [0.30, 0.99], $p = .047$.

Delay time measured at the start of the school year was positively associated with final grade point average (GPA) measured at the end of the school year in bivariate analyses, but this relationship failed to reach significance, $rh = 0.79$, 95% CI = [0.55, 1.14], $p = .21$. However, because nonverbal and verbal intelligence each predicted GPA ($r = .36$ and .38, $ps < .01$, respectively), we included these covariates in the same model to reduce error and found that the relationship between delay time and GPA was marginally significant, $rh = 0.65$, 95% CI = [0.41, 1.03], $p = .067$. Thus, although not quite significant, the inclusion of measures of intelligence increased (in magnitude) rather than diminished the predictive validity of the delay task for academic achievement.

Study 2

In Study 1, school-age children who waited longer in the delay of gratification paradigm were considered more self-controlled and conscientious by their teachers but no different in other dimensions of personality nor in their attraction to rewards. However, in this convenience sample, we documented restriction on range in intelligence. Moreover, the small sample size constrained statistical power; it is possible that with more participants, weaker associations between delay behavior and other variables would have reached statistical significance. Finally, while we could confirm that the delay task marginally predicted report card grades when controlling for intelligence, health and social outcomes were not available in Study 1, nor were any outcomes assessed later than a year after the delay task was administered.

In Study 2, we addressed these limitations by conducting secondary analysis of data from a national sample of 966 children who completed the delay of gratification task at age 4. In addition to concurrent ratings of temperament by teachers and caregivers and IQ scores, follow-up data collected a decade later were available, making possible prospective, longitudinal analyses with early adolescent outcomes, including objectively measured report card grades, standardized achievement test scores, body mass index (BMI), and self-reported risky behavior. To our knowledge, none of the analyses reported here have been conducted previously.

Method

Participants—The participants were 966 children from the National Institute of Child Health and Development (NICHD) Study of Early Child Care and Youth Development (SECCYD; https://secc.rti.org/) who completed the preschool delay of gratification task. Approximately 80% of participants were White, 11% were Black, 5% were Hispanic, 1% were Asian, and 3% were other ethnicities; 52% were female. The median household income-to-needs ratio (assessed in terms of income compared with the U.S. Census Bureau–defined poverty line) for this sample was 2.9, and on average, mothers in this sample had completed 14 years of education.

Procedure and Measures

Delay of gratification—When they were 4 years old, children participated in a laboratory task in which they first selected their favorite among several snacks (e.g., chocolate candies, cookies, pretzels). Next, the experimenter placed a plate with a small amount of snack and a plate with a larger amount of snack in front of the child and asked which the child preferred. Once it was established that the child preferred the larger amount, the child was told that she or he would be allowed to eat the larger amount if she or he waited until the experimenter returned, but if the child could not wait, then she or he could ring a bell, the experimenter would return, and the child would be given the smaller amount of snack. The child was also instructed to remain seated and not to eat the snack until the experimenter returned. Once the child understood the instructions, the experimenter left the room and watched the child from an observation booth. The experimenter returned and wait time was measured when the child rang the bell, left the seat, ate the snack, became distressed, or called for the experimenter or a parent. Otherwise, the experimenter returned after 7 min and gave the child the larger amount of snack.

Self-control, reward-related impulses, and other dimensions of temperament —When participants were 4 years old, their mothers and caregivers (e.g., preschool teachers) completed selected subscales of the Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994). Because effortful control— "the ability to inhibit a

dominant response to perform a subdominant response" (Rothbart & Bates, 1998, p. 137)—corresponds to our definition of self-control, we used the Attention Focusing and Inhibitory Control subscales as measures of self-control. The Attention Focusing subscale reflects the capacity to maintain attentional focus (e.g., "Hard time concentrating on activity"), and Inhibitory Control reflects the capacity to plan and to suppress inappropriate responses (e.g., "Able to resist temptation"). Correlations between mother and caregiver ratings of Attention Focusing and Inhibitory Control were $r = .32$ and $r = .34$, respectively. These associations compare favorably to the meta-analytically derived average correlation of $r = .28$ between two different types of informant (e.g., parent/teacher) by Achenbach, McConaughy, and Howell (1987). Correlations among all indicators of self-control ranged from .24 to .68 (avg. = .40), and internal reliability coefficients ranged from .74 to .84 (avg. = .78).

We used mother ratings of Approach/Anticipation and Activity Level as measures of reward-related impulses. Caregivers did not complete these subscales. The Approach/Anticipation subscale reflects excitement and positive anticipation for expected pleasurable activities (e.g., "When she or he sees a toy she or he wants, gets very excited"), and Activity Level reflects gross motor activity (e.g., "Always in a bit of a hurry to get from one place to another"). The correlation between Approach/Anticipation and Activity Level was .46, and the internal reliability coefficients were .69 and .67, respectively.

To further assess discriminant validity, we also analyzed all other CBQ subscales completed by mothers and/or caregivers. The Anger/Frustration subscale reflects negative affect related to interruption of ongoing tasks or goal blocking (e.g., "Has temper tantrums when she or he does not get her or his way"); Fear reflects unease, worry, or nervousness related to anticipated pain or distress and/or potentially threatening situations (e.g., "Is afraid of the dark?"); Sadness reflects negative affect and lowered mood and energy related to exposure to suffering, disappointment, and object loss (e.g., "Sometimes appears downcast for no reason"); and Shyness reflects an inhibited approach in social situations involving novelty or uncertainty (e.g., "Acts shy around new people"). Internal reliability coefficients for these subscales ranged from .59 to .90 (avg. = .76). Intercorrelations between mother and caregiver ratings were .16, .12, and .43 for anger, sadness, and shyness, respectively.

Intelligence—At age 4, participants completed the Memory for Sentences, Incomplete Words, and Picture Vocabulary subscales of the Woodcock–Johnson Psycho-Educational Battery–Revised (WJ-R) Tests of Cognitive Abilities (WJ-R COG; Woodcock & Johnson, 1989). In a validity study (McGrew, Werder, & Woodcock, 1991), the WJ-R COG correlated highly (rs > .70) with similar tests of intelligence (e.g., Stanford-Binet, McCarthy, and Kaufman Assessment Battery for Children [ABC]). Correlations among these indicators of intelligence ranged from .38 to .49 (avg. = .44), and internal reliability coefficients ranged from .75 to .85 (avg. = .81).

Academic performance—Principals or their designated staff members reported final grades for math, English, science, and social studies for participants at the end of the eighth grade. Schools provided official student transcripts at the end of the ninth grade. Final grades for math, science, English, and social studies were converted to a numeric scale where $A+ = 4.33$ to $F = 0.00$. Eighth- and ninth-grade GPA were highly correlated ($r = .72$), so we averaged them to create a composite GPA.

In ninth grade, the children completed the Passage Comprehension and Applied Problems Achievement subscales of the Woodcock–Johnson Psycho-Educational Battery–Revised Tests of Achievement (WJ-R ACH). The WJ-R includes separate tests of cognitive ability and achievement, the latter of which are designed to assess academic skills and knowledge

(Mather, 1991). Passage Comprehension and Applied Problems were highly correlated, $r = .65$, so we averaged these scales to create composite standardized achievement test scores.

BMI—BMI has been demonstrated as a reliable marker of overall physical health in adolescence (Swallen, Reither, Haas, & Meier, 2005). Height and weight were measured using standardized protocols at ninth grade, and these data were used to calculate an age- and sex-specific BMI z score for each participant. The average BMI z score was .56, indicating that the average adolescent in our sample was slightly overweight.

Risky behavior—In ninth grade, participants completed a questionnaire asking how many times in the past year they engaged in different risky behaviors, including substance use (e.g., "Used or smoked marijuana"), endangerment to their safety (e.g., "Ridden a motorcycle without wearing a helmet"), and social risks (e.g., "Stolen something"). This scale was adapted for the NICHD-SECCYD from work by Conger and Elder (1994) and Halpern-Felsher, Biehl, Kropp, and Rubinstein (2004). We coded items such that 0 = "never" and 1 = "once or twice and more than twice" before creating a summed score. On average, participants endorsed about 6 out of 53 risky behaviors. The observed internal reliability was .89.

Analytic Strategy—After assessing descriptive statistics and bivariate associations, we fit a series of structural equation models (SEMs) to estimate the extent to which the predictive validity of the preschool delay of gratification task can be explained by synchronous latent measures of temperament and intelligence. Our intent was not to conduct mediation analyses for the purpose of making causal inferences about the mechanisms by which delay ability translates into life outcomes. Rather, using SEM techniques originally developed to assess multiple mediators (MacKinnon, 2008), we aimed to disaggregate the longitudinal associations between delay time and later life outcomes in terms of variance shared with self-control versus theoretically unrelated traits. Because SEM procedures have not yet been developed for censored predictor variables, we treated delay of gratification behavior as a binary variable where 1 = "delayed until task conclusion" (7 min) and 0 = "ended task early." We followed the recommendations of MacKinnon, Lockwood, Hoffman, West, and Sheets (2002) and, rather than conduct Sobel tests for significance, examined for each proposed mediator "the joint significance of the two effects comprising the intervening variable effect" (p. 83). Because children from higher-income households or with more educated mothers were more likely to delay gratification, as were White children, we included these and all other demographic covariates in all SEMs. Because income-to-needs and risky behavior were positively skewed, we log-transformed both to normalize their distributions (adding one to the latter before log-transforming to remove zeros). For bivariate associations with delay time in Table 2, we report hazard ratios from Cox regression models using the censored variable.

To correct for measurement error, we used latent variables for self-control, reward-related impulses, and intelligence, with their respective subscales as observed indicators. All other variables were treated as observed variables. GPA, standardized academic achievement, BMI z score, and risky behavior at age 15 were the outcome variables, and their disturbances were allowed to covary. For the self-control latent variable, the error variances for the same reporter (e.g., mother-report attention focusing and mother-report inhibitory control) and the same subscale (e.g., mother-report attention focusing and caregiver-report attention focusing) were allowed to covary. We used full information maximum likelihood (FIML) to handle missing data (about 7% of the data were missing; see Table 2). FIML is less biased and more efficient than traditional missing data techniques (Enders & Bandalos, 2001; Peters & Enders, 2002).

Results and Discussion

After declaring their intention to wait for a preferred treat in the delay of gratification paradigm, 4-year-olds in Study 2 waited an average of 4.5 min ($SD = 3.0$). About 47% of children terminated the task early (i.e., before 7 min had elapsed and the experimenter returned to the room). As summarized in Table 2, we fit separate Cox regression models to estimate bivariate associations between delay of gratification and other variables. Consistent with prior longitudinal studies, delay time at age 4 was related to each of the outcomes assessed in adolescence, $ps < .001$: Children who delayed longer later earned higher GPAs ($rh = 0.72$, 95% CI = [0.65, 0.80], $p < .001$) and standardized achievement test scores ($rh = 0.59$, 95% CI = [0.52, 0.67], $p < .001$) and had healthier (i.e., lower) BMI scores ($rh = 1.29$, 95% CI = [1.15, 1.44], $p < .001$) and engaged in fewer risky behaviors ($rh = 1.47$, 95% CI = [1.28, 1.69], $p < .001$).

Delay time was associated with concurrent ratings of self-control by mothers and caregivers. That is, children who were rated one standard deviation higher than average in attention focusing or inhibitory control by their mother or caregiver were 19% to 25% less likely to terminate the delay task as a function of time, rhs from 0.75 to 0.81, $ps < .001$. In contrast, delay time was less reliably related to reward-related impulses: Mother ratings of motor activity level predicted delay time ($rh = 1.21$, 95% CI = [1.10, 1.33], $p < .001$), but mother ratings of approach/anticipation tendencies did not, $rh = 1.07$, 95% CI = [0.98, 1.18], $p = .15$. Likewise, for other measured dimensions of temperament, including anger/frustration, fear, and sadness, associations with delay performance failed to reach significance for one or both raters.

Other than self-control, the only aspect of temperament that demonstrated reliable associations with delay performance was shyness. Children rated one standard deviation higher in shyness by their mother or caregiver were 9% and 12% less likely to terminate the delay task as a function of time, $rh = 0.91$, 95% CI = [0.83, 1.00] and 0.88, 95% CI = [0.78, 0.99], $ps < .05$. One post hoc explanation for this somewhat unexpected finding was that interacting with a novel adult (i.e., the female experimenter who conducted the delay task) precipitated some degree of fearfulness in shyer children, causing them to freeze up and, by default, to wait longer. Before fitting SEMs, we confirmed that mother ratings of shyness did not predict any adolescent outcomes, and caregiver ratings of shyness predicted only two of four outcomes: BMI ($r = -.09$, $p = .04$) and risk taking ($r = -.11$, $p = .009$). Given that only two of eight possible associations between shyness and outcomes reached significance, we did not include shyness in subsequent analyses.[1]

In our first SEM, we confirmed that the direct effects of preschool delay of gratification on adolescent outcomes held when controlling for demographic covariates. When controlling for family income, maternal education, ethnicity, and age, delay of gratification at age 4 continued to predict higher standardized achievement test scores (= .12, $SE = .03$, $p < .001$) and GPAs (= .08, $SE = .03$, $p = .016$), as well as lower BMI scores (= −.10, $SE = .04$, $p = .01$), and fewer risky behaviors (= −.07, $SE = .04$, $p = .037$). Because this model was just-identified and only included observed variables, model fit statistics were not available.

In a second SEM, we added separate latent factors for self-control, reward-related impulses, and intelligence. Factor loadings for self-control ranged from .40 to .68 (avg. = .53), loadings for reward-related impulses ranged from .47 to .99 (avg. = .73), and loadings for

---

[1]In a structural equation model (SEM; not reported), with shyness as a latent variable, shyness did not mediate the effect of delay on risk taking or body mass index. This model was based on the final model reported below, except that it included shyness as a latent variable. The results (i.e., other paths in the model) were virtually identical to the final model.

intelligence ranged from .58 to .73 (avg. = .66), $ps < .001$. Model 2 fit the data well: $^2(98) = 249.38$, $p < .001$; comparative fit index (CFI) = .96; root mean square error of approximation (RMSEA) = .04 (90% CI = [.03 to .05]). Children who delayed gratification were higher in self-control (= .20, $SE = .04$, $p < .001$) and were more intelligent (= .25, $SE = .03$, $p < .001$). In contrast, a weaker relationship was observed between delay behavior and reward-related impulses (= −.09, $SE = .03$, $p = .006$). Moreover, reward-related impulses predicted none of the four adolescent outcomes, all $ps > .15$. Because reward-related impulses did not mediate any of the effects of delay behavior and were highly correlated with self-control ($r = −.72$), we reduced multicollinearity by excluding this construct in our final model. Path coefficients in this final model, described below, for self-control and intelligence were nearly identical but, as expected, standard errors were reduced.

Our final SEM, illustrated in Figure 1, fit the data well: $^2(74) = 162.08$, $p < .001$; CFI = .97; RMSEA = .04 (90% CI = [.03 to .04]). Demographic covariates of gender, age, ethnicity, family income, and maternal education were included in this model but are not shown in the figure. Preschool delay performance was associated with concurrently measured self-control (= .21, $SE = .05$, $p < .001$) and intelligence (= .25, $SE = .03$, $p < .001$) in this model.

As shown in Figure 1, for report card grades in eighth and ninth grade, the predictive power of the delay task was explained by self-control = .31, $SE = .07$, $p < .001$) but not intelligence, = .10, $SE = .06$, $p = .10$. This finding adds to a growing literature demonstrating that self-control (more typically assessed using informant or self-report ratings) predicts report card grades better than does any other aspect of temperament or personality (Duckworth & Allred, 2012).

The prediction of higher standardized achievement test scores by delay performance was explained in part by self-control (= .21, $SE = .07$, $p = .001$) as well as verbal intelligence (= .48, $SE = .06$, $p < .001$). This finding comports with separate research suggesting that intelligence is a better predictor of standardized achievement test scores than self-control, whereas self-control is a better predictor of report card grades than intelligence (Duckworth, Quinn, & Tsukayama, 2012). One possible explanation for divergent associations with different measures of academic achievement is that report card grades differentially reward positive classroom behavior, studying, and homework, whereas achievement tests differentially tap the ability to solve novel problems without formal instruction.

For self-reported risky behavior in adolescence, self-control (= −.13, $SE = .08$, $p = .074$), but not intelligence (= .06, $SE = .07$, $p = .34$), was a marginal predictor. When accounting for these two factors, delay performance was no longer a significant predictor of risky behavior, = −.06, $SE = .04$, $p = .11$. The association between self-control and risky behavior in adolescence has been well-documented in other studies (e.g., Romer, Duckworth, Sznitman, & Park, 2010). Current theory suggests that risky behaviors peak in adolescence because self-control processes are still maturing while reward-related impulses dramatically increase in strength during this developmental epoch (Steinberg, 2008). The current findings support the view that variance in risky behavior during adolescence is predicted by early emerging differences in self-control but not reward-related impulses.

As for physical health, preschool children rated higher in self-control maintained a healthier BMI in adolescence (= −.26, $SE = .08$, $p = .002$), whereas more intelligent preschool children ended up slightly heavier (= . 15, $SE = .07$, $p = .045$). This finding corroborates separate longitudinal research showing that more self-controlled children maintain healthier bodyweights, particularly as they enter adolescence and more independently make choices about what and how much to eat (Duckworth, Tsukayama, & Geier, 2010; Tsukayama, Toomey, Faith, & Duckworth, 2010). Delay performance was a marginal predictor of BMI

when controlling for self-control and intelligence ($ = -.08$, $SE = .04$, $p = .054$). The finding that more intelligent children ended up heavier was surprising given longitudinal research identifying intelligence as a protective factor against weight gain (Chandola, Deary, Blane, & Batty, 2006). Given the borderline p value of .045, this result may be due to chance and/or due to suppression effects from the other variables in the model. Nonetheless, this finding does not change our conclusion that self-control, rather than intelligence, is responsible for the protective relationship between delay of gratification and BMI.

General Discussion

Overall, our findings suggest the delay of gratification task predicts life outcomes because it measures self-control, rather than intelligence or reward-related impulses. Among school-age children in Study 1 and preschool children in Study 2, self-imposed wait time in this task converged with concurrent ratings of self-control by adult informants. These associations were small to medium in terms of effect size (Bedard, Krzyzanowska, Pintilie, & Tannock, 2007), comparing favorably to meta-analytic estimates of correlations between task and questionnaire measures in general (Meyer et al., 2001) and for self-control in particular (Duckworth & Kern, 2011). Moreover, wait time was less reliably related to reward-related impulses (in both studies) or to conceptually distinct traits in taxonomies of personality (in Study 1) or temperament (in Study 2). Finally, we confirmed that performance in the delay task provided incremental predictive validity over and beyond intelligence for GPA (in Studies 1 and 2), as well as standardized achievement test scores, BMI, and risky behavior (in Study 2). As expected, informant ratings of preschool self-control consistently explained the predictive validity of the delay task for adolescent outcomes, whereas informant ratings of preschool reward-related impulses did not.

Kelvin (1883) famously observed,

> when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be. (p. 73)

We can think of no better exemplar of Kelvin's dictum than the delay of gratification paradigm. Long assumed central to successful development, self-control has only within the last half century become the object of productive scientific inquiry. While not the only valid measure of self-control available to researchers, the delay of gratification task has crucial advantages. Most notably, the delay task obviates the well-known limitations of questionnaire measures (e.g., faking, social desirability bias, acquiescence bias, and reference bias).

What makes the delay of gratification task so exquisitely sensitive to individual differences in self-control? We can only speculate, but several features of the paradigm seem worth highlighting. First, the child is presented with a range of treats from which they choose their favorite. Temptation is thus maximized by using a treat the child really likes, but the very trivial amount of snack likely precludes hunger impulses to swamp self-regulatory processes, as evidenced by a near-zero correlation between self-reported hunger ratings at the start of the task and delay time in Study 1. Second, the task is administered in a quiet, empty room in which the child is left alone to ponder, continuously, his or her choice—shall I continue to wait or shall I gobble up this smaller treat right now? In the absence of external distractions, with temptation lying within easy reach and in plain sight, children rely on self-regulatory strategies of varying effectiveness (Carlson & Beck, 2009). Third, before leaving, the experimenter emphasizes to the child that she doesn't care much what the child

ultimately decides to do. This minimizes the possibility that children wait to comply with authority, as seems to be the case in other tasks (e.g., the gift delay task in Funder et al., 1983). Finally, unlike more easily administered measures in which individuals make discreet (and irrevocable) choices between smaller, sooner and larger, later rewards, the delay task begins with the (universal) election for larger, later treats and then tests the ability to *sustain* the decision to wait.

Limitations

We see three important limitations of the present investigation. The first concerns the lack of adult outcomes for the children who completed the delay task in both Studies 1 and 2. While research using alternative measures suggests that self-control contributes to a wide range of outcomes in adulthood (e.g., Moffitt et al., 2011), it will be years before comparable outcome data is available for the participants in this investigation.

Second, our analyses were restricted to data that had been collected, particularly in Study 2, which relied upon a large, public data set. Thus, while in both studies we were able to situate the delay task within nomological networks established by omnibus measures of personality and temperament, there is no way to know for certain whether some unmeasured trait would have demonstrated the same pattern of results as self-control. Likewise, it is possible that better measures of reward-related impulses would have produced stronger associations with delay performance and outcomes. For school-age children, for instance, it is possible that reward-related impulses might be more accurately elicited using an implicit association task (Hofmann, Deutsch, Lancaster, & Banaji, 2010).

Finally, neither of our samples were nationally representative. Methodologists (e.g., Grace & Bollen, 2005) have argued that interpreting standardized coefficients in convenience samples is problematic because standardized coefficients are based on both unstandardized effects as well as (possibly truncated) sample standard deviations. The implication is that the relative strength of standardized coefficients in a sample may not reflect the population if there is restricted range on some variables but not others. While Study 1 had a small sample from a single school, this issue is less of a problem for Study 2, which included a socioeconomically, ethnically, and geographically diverse sample of children from across the United States. Furthermore, a significant unstandardized coefficient suggests a significant standardized coefficient (because the standardized coefficient is zero if the unstandardized coefficient is zero). Therefore, regardless of the relative strength of the predictors, the pattern of significant results supports our hypothesis that self-control, rather than intelligence or reward-related impulses, is responsible for the predictive power of the delay task.

Conclusion

Performance task measures of competencies other than mental ability are regrettably few in modern psychology research. Despite heroic attempts in this direction earlier in psychology's history (e.g., Hartshorne & May, 1929), psychological research these days is dominated by "introspective self-reports, hypothetical scenarios, and questionnaire ratings" (Baumeister, Vohs, & Funder, 2007, p. 396). The current investigation affirms the value of directly measuring human behavior under standardized conditions explicitly designed to elicit theoretically interpretable responses and verifying, through systematic investigation of its correlates and consequences, that the task indeed assesses what it was intended to assess.

Acknowledgments

References

Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin. 1987; 101:213–232. [PubMed: 3562706]

Ayduk O, Mendoza-Denton R, Mischel W, Downey G, Peake PK, Rodriguez M. Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. Journal of Personality and Social Psychology. 2000; 79:776–792. [PubMed: 11079241]

Barkley RA. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. Psychological Bulletin. 1997; 121:65–94. [PubMed: 9000892]

Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on Psychological Science. 2007; 2:396–403.

Bedard PL, Krzyzanowska MK, Pintilie M, Tannock IF. Statistical power of negative randomized controlled trials presented at American society for clinical oncology annual meetings. Journal of Clinical Oncology. 2007; 25:3482–3487. [PubMed: 17687153]

Blair C, Peters R, Granger D. Physiological and neuropsychological correlates of approach/withdrawal tendencies in preschool: Further examination of the behavioral inhibition system/behavioral activation system scales for young children. Developmental Psychobiology. 2004; 45:113–124. [PubMed: 15505800]

Block J, Funder DC. The role of ego-control, ego-resiliency, and IQ in delay of gratification in adolescence. Journal of Personality and Social Psychology. 1989; 57:1041–1050. [PubMed: 2614657]

Borghans L, Duckworth AL, Heckman JJ, ter Weel B. The economics and psychology of personality traits. Journal of Human Resources. 2008; 43:972–1059.

Carlson, SM.; Beck, DM. Symbols as tools in the development of executive function. In: Winsler, A.; Fernyhough, C.; Montero, I., editors. Private speech, executive functioning, and the development of verbal self-regulation. New York, NY: Cambridge University Press; 2009. p. 163-175.

Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology. 1994; 67:319–333.

Chandola T, Deary IJ, Blane D, Batty GD. Childhood IQ in relation to obesity and weight gain in adult life: The National Child Development (1958) Study. International Journal of Obesity. 2006; 30:1422–1432. [PubMed: 16520812]

Conger, RD.; Elder, GH. Families in troubled times: Adapting to change in rural America. New York, NY: Aldine de Gruyter; 1994.

Conway ARA, Kane MJ, Engle RW. Working memory capacity and its relation to general intelligence. Trends in Cognitive Sciences. 2003; 7:547–552. [PubMed: 14643371]

Duckworth, AL.; Allred, KM. Temperament in the classroom handbook of temperament. New York, NY: Guilford; 2012.

Duckworth AL, Kern ML. A meta-analysis of the convergent validity of self-control measures. Journal of Research in Personality. 2011; 45:259–268. [PubMed: 21643479]

Duckworth AL, Quinn PD, Tsukayama E. What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. Journal of Educational Psychology. 2012; 104:439–451. [PubMed: 24072936]

Duckworth AL, Tsukayama E, Geier AB. Self-controlled children stay leaner in the transition to adolescence. Appetite. 2010; 54:304–308. [PubMed: 20004223]

Eisenberg N, Duckworth AL, Spinrad TL, Valiente C. Conscientiousness: Origins in childhood?. 2012 Manuscript submitted for publication.

Eisenberg, N.; Morris, AS. Children's emotion-related regulation. In: Kail, RV., editor. Advances in child development and behavior. Vol. Vol. 30. San Diego, CA: Academic Press; 2002. p. 189-229.

Eisenberg, N.; Smith, CL.; Sadovsky, A.; Spinrad, T., editors. Effortful control: Relations with emotion regulation, adjustment, and socialization in childhood. New York, NY: Guilford; 2004.

Eisenberg N, Spinrad TL, Fabes RA, Reiser M, Cumberland A, Shepard SA, Thompson M. The relations of effortful control and impulsivity to children's resiliency and adjustment. Child Development. 2004; 75:25–46. [PubMed: 15015673]

Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling. 2001; 8:430–457.

Freud, S. Introductory lectures on psychoanalysis. New York, NY: Norton; 1920.

Frey MC, Determan DK. Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. Psychological Science. 2004; 15:373–378. [PubMed: 15147489]

Fujita K, Carnevale JJ. Transcending temptation through abstraction. Current Directions in Psychological Science. 2012; 21:248–252.

Funder DC, Block JH, Block J. Delay of gratification: Some longitudinal personality correlates. Journal of Personality and Social Psychology. 1983; 44:1198–1213.

Grace JB, Bollen KA. Interpreting the results from multiple regression and structural equation models. Bulletin of the Ecological Society of America. 2005; 86:283–295.

Halpern-Felsher BL, Biehl M, Kropp RY, Rubinstein ML. Perceived risks and benefits of smoking: Differences among adolescents with different smoking experiences and intentions. Preventive Medicine. 2004; 39:559–567. [PubMed: 15313096]

Hartshorne, H.; May, MA. Studies in the nature of character, Volume II: Studies in self-control. Vol. Vol. 2. New York, NY: Macmillan; 1929.

Heatherton TF, Wagner DD. Cognitive neuroscience of self-regulation failure. Trends in Cognitive Science. 2011; 15:132–139.

Hofmann W, Deutsch R, Lancaster K, Banaji MR. Cooling the heat of temptation: Mental self-control and the automatic evaluation of tempting stimuli. European Journal of Social Psychology. 2010; 40:17–25.

Hofmann W, Friese M, Stack F. Impulse and self-control from dual-systems perspective. Perspectives on Psychological Science. 2009; 4:162–176.

John, OP.; Srivastava, S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, LA.; John, OP., editors. Handbook of personality: Theory and research. 2nd ed.. New York, NY: Guilford; 1999. p. 102-138.

Kelvin, WT. Popular lectures and addresses. Vol. Vol. 1. London, England: Macmillan; 1883.

Lehrer, J. Don't! The secret of self-control. The New Yorker. 2009 May 18. Retrieved from http://www.newyorker.com/reporting/2009/05/18/090518fa_fact_lehrer

Lesure GE. Relationship between intelligence and preferences to work for delayed rewards. Psychological Reports. 1977; 40:493–494.

Lilienfeld SO, Wood JM, Garb HN. The scientific status of projective techniques. Psychological Science in the Public Interest. 2000; 1:27–66.

MacKinnon, DP. Introduction to statistical mediation analysis. New York, NY: Lawrence Erlbaum; 2008.

MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychological Methods. 2002; 7:83–104. [PubMed: 11928892]

Mather, N. An instructional guide to the Woodcock-Johnson Psycho-Educational Battery—Revised. New York, NY: John Wiley; 1991.

McCrae, RR.; Lockenhoff, CE. Self-regulation and the five-factor model of personality traits. In: Hoyle, RH., editor. Handbook of personality and self-regulation. Chichester, West Sussex: Blackwell; 2010. p. 145-168.

McGrew, KS.; Werder, JK.; Woodcock, RW. WJ-R technical manual. Allen, TX: DLM; 1991.

Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, Read GM. Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist. 2001; 56:128–165. [PubMed: 11279806]

Mischel, W. Walter Mischel. In: Lindzey, G.; Runyan, WM., editors. A history of psychology in autobiography. Vol. Vol. IX. Washington, DC: American Psychological Association; 2007. p. 229-267.

Mischel W, Brooks D. The news from psychological science: A conversation between David Brooks and Walter Mischel. Perspectives on Psychological Science. 2011; 6:515–520.

Mischel W, Ebbesen EB, Zeiss AR. Cognitive and attentional mechanisms in delay of gratification. Journal of Personality and Social Psychology. 1972; 21:204–218. [PubMed: 5010404]

Mischel W, Metzner R. Preference for delayed reward as a function of age, intelligence, and length of delay interval. Journal of Abnormal and Social Psychology. 1962; 64:425–431. [PubMed: 14474526]

Mischel W, Shoda Y, Peake PK. The nature of adolescent competencies predicted by preschool delay of gratification. Journal of Personality and Social Psychology. 1988; 54:687–696. [PubMed: 3367285]

Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington HL, Caspi A. A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:2693–2698. [PubMed: 21262822]

Olson EA, Hooper CJ, Collins P, Luciana M. Adolescents' performance on delay and probability discounting tasks: Contributions of age, intelligence, executive functioning, and self-reported externalizing behavior. Personality and Individual Differences. 2007; 43:1886–1897. [PubMed: 18978926]

Olson SL, Sameroff AJ, Kerr DCR, Lopez NL, Wellman HM. Developmental foundations of externalizing problems in young children: The role of effortful control. Development and Psychopathology. 2005; 17:25–45. [PubMed: 15971758]

Peters CLO, Enders C. A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. Journal of Targeting, Measurement and Analysis for Marketing. 2002; 11:81–95.

Public Broadcasting Service. "Sesame street" tells you how to get to sunnier days financially. 2011. Retrieved from http://www.pbs.org/newshour/bb/business/jan-june11/making-sense_06-03.html

Raven J. The Raven's progressive matrices: Change and stability over culture and time. Cognitive Psychology. 2000; 41:1–48. [PubMed: 10945921]

Raven, J.; Raven, JC.; Court, JH. Manual for Raven's progressive matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment; 1988.

Raven, J.; Raven, JC.; Court, JH. Manual for Raven's progressive matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment; 2000.

Reynolds B, Schiffbauer R. Delay of gratification and delay discounting: A unifying feedback model of delay-related impulsive behavior. Psychological Record. 2005; 55:439–460.

Roberts, BW.; Jackson, JJ.; Fayard, JV.; Edmonds, G.; Meints, J. Conscientiousness. In: Leary, M.; Hoyle, R., editors. Handbook of individual differences in social behavior. New York, NY: Guilford; 2009. p. 369-381.

Rodriguez ML, Mischel W, Shoda Y. Cognitive person variables in the delay of gratification of older children at risk. Journal of Personality and Social Psychology. 1989; 57:358–367. [PubMed: 2760808]

Romer D, Duckworth AL, Sznitman S, Park S. Can adolescents learn self-control? Delay of gratification in the development of control over risk taking. Prevention Science. 2010; 11:319–330. [PubMed: 20306298]

Rothbart MK, Ahadi SA, &Hershey KL. Temperament and social behavior in childhood. Merrill-Palmer Quarterly: Journal of Developmental Psychology. 1994; 40:21–39.

Rothbart, MK.; Bates, JE. Temperament. In: Damon, W.; Eisenberg, N., editors. Handbook of child psychology: Social, emotional, and personality development. 5th ed. Vol. Vol. 3. New York, NY: John Wiley; 1998. p. 105-176.

Shamosh NA, DeYoung CG, Green AE, Reis DL, Johnson MR, Conway ARA, Gray JR. Individual differences in delay discounting: Relation to intelligence, working memory, and anterior prefrontal cortex. Psychological Science. 2008; 19:904–911. [PubMed: 18947356]

Shamosh NA, Gray JR. Delay discounting and intelligence: A meta-analysis. Intelligence. 2008; 36:289–305.

Shiner, RL.; DeYoung, CG. The structure of temperament and personality traits: A developmental perspective. In: Zelazo, PD., editor. Oxford handbook of developmental psychology. New York, NY: Oxford University Press; 2013. p. 113-141.

Shoda Y, Mischel W, Peake PK. Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. Developmental Psychology. 1990; 26:978–986.

Singer JL. Delayed gratification and ego development: Implications for clinical and experimental research. Journal of Consulting Psychology. 1955; 19:259–266. [PubMed: 13242702]

Steinberg L. A social neuroscience perspective on adolescent risk-taking. Development Review. 2008; 28:78–106.

Swallen KC, Reither EN, Haas SA, Meier AM. Overweight, obesity, and health-related quality of life among adolescents: The National Longitudinal Study of Adolescent Health. Pediatrics. 2005; 115:340–347. [PubMed: 15687442]

Tsukayama E, Duckworth AL, Kim BE. Domain-specific impulsivity in school-age children. Developmental Science. in press.

Tsukayama E, Toomey SL, Faith MS, Duckworth AL. Self-control as a protective factor against overweight status in the transition from childhood to adolescence. Archives of Pediatrics & Adolescent Medicine. 2010; 164:631–635. [PubMed: 20603463]

Woodcock, RW.; Johnson, MB. Woodcock-Johnson Psycho-Educational Battery—Revised. Allen, TX: DLM; 1989.
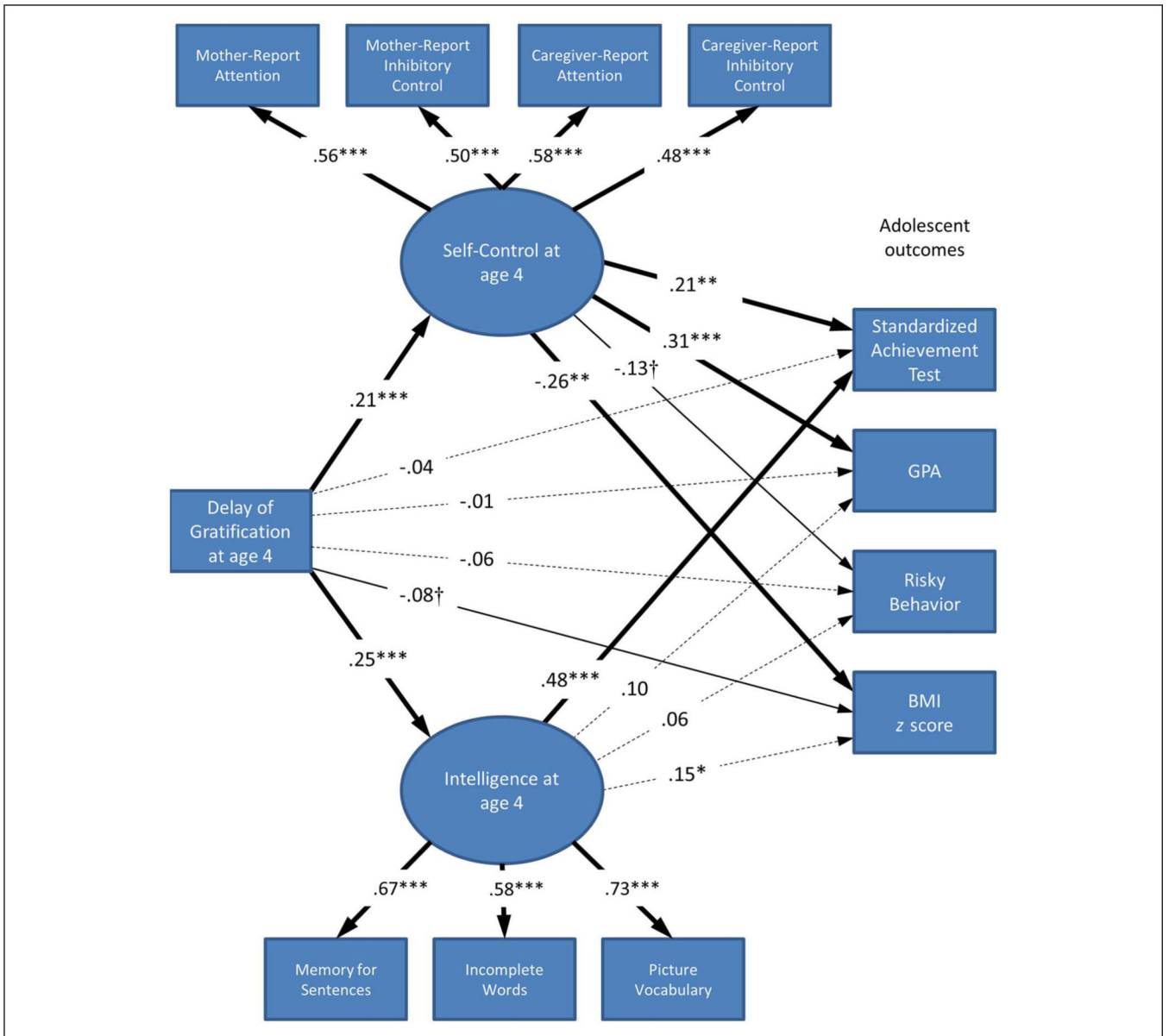
Figure 1.

Structural equation model in Study 2

Note: The covariates of gender, age, ethnicity, log-transformed income-to-needs, and maternal education, as well as error variances, and disturbances, are not displayed. Also not shown is the correlation between self-control and verbal intelligence, $r = .38$. Bold lines indicate significant relationships, normal lines indicate marginal relationships at $p < .10$, and dotted lines indicate nonsignificant paths or paths significant in the nonpredicted direction. †p< .10. *p<.05. ***p< .001.

Table 1

Summary Statistics and Bivariate Associations With Delay Time in Study 1.

Measure	M (SD)	Range	Hazard ratio	95% CI
Delay time in minutes	24.50 (8.52)	0.75–30.00		
Teacher-rated self-control	4.24 (0.88)	1.63–5.00	0.63[*]	[0.40, 0.99]
Reward-related impulses				
BAS reward responsiveness	4.08 (0.53)	2.60–5.00	0.86	[0.58, 1.28]
BAS drive	3.36 (0.86)	1.00–4.75	0.95	[0.65, 1.41]
Self-reported hunger	3.04(1.64)	1–7	1.01	[0.67, 1.53]
BIS inhibition	3.34 (0.56)	2.14–4.71	1.07	[0.70, 1.64]
Intelligence				
Raven's progressive matrices	45.20 (5.54)	29–56	0.98	[0.62, 1.54]
Mill hill Junior Vocabulary Scale	36.77 (4.70)	25–46	1.17	[0.77, 1.76]
Big Five personality				
Conscientiousness	3.57 (0.89)	1.78–5.00	0.59[*]	[0.37, 0.93]
Extraversion	3.12(0.74)	1.75–5.00	0.87	[0.58, 1.30]
Openness to experience	3.66 (0.52)	2.70–4.80	0.64	[0.38, 1.08]
Agreeableness	3.97 (0.63)	2.22–5.00	0.85	[0.52, 1.39]
Emotional stability	3.55 (0.62)	2.00–5.00	0.86	[0.5 1, 1.45]
Final GPA	89.30 (4.89)	76–97	0.79	[0.55, 1.14]
Demographics				
Age in years	10.28(0.40)	9.25–10.83	0.90	[0.63, 1.31]
Female	55%		1.00	[0.66, 1.51]
Free/reduced lunch status	14%		0.71	[0.34, 1.46]

Note: BAS = behavioral activation system; BIS = behavioral inhibition system; CI = confidence interval; GPA = grade point average, n = 56 for all variables except GPA, where *n* = 54. Cox regression models for verbal and nonverbal intelligence included age as a covariate.

[*]
p < .05.

Table 2

Summary Statistics and Bivariate Associations With Delay Time in Study 2.

	M	SD	Observed range	n	Hazard ratio	95% CI
Delay time in minutes	4.48	3.01	0.00–7.00	961	—	
Ratings of self-control at age 4						
Mother-report CBQ attention focusing	4.72	0.84	1.25–6.88	922	0.75 ***	[0.69, 0.82]
Caregiver-report CBQ attention focusing	4.88	0.99	1.25–7.00	697	0.80 ***	[0.72, 0.90]
Mother-report CBQ inhibitory control	4.69	0.77	2.00–6.70	958	0.76 ***	[0.69, 0.83]
Caregiver-report CBQ inhibitory control	5.11	1.02	1.80–7.00	705	0.81 ***	[0.72, 0.90]
Ratings of reward-related impulses at age 4						
Mother-report CBQ activity level	4.78	0.76	1.60–6.90	924	1.21 ***	[1.10, 1.33]
Mother-report CBQ approach/anticipation	5.19	0.63	2.90–7.00	898	1.07	[0.98, 1.18]
Ratings of other dimensions of temperament						
Mother-report CBQ anger/frustration	4.73	0.83	1.60–6.90	958	1.05	[0.96, 1.16]
Caregiver-report CBQ anger/frustration	3.44	1.10	1.00–6.50	694	1.18 **	[1.05, 1.32]
Mother-report CBQ fear	4.09	0.85	1.40–6.20	734	1.02	[0.92, 1.14]
Mother-report CBQ sadness	3.96	0.70	1.60–5.90	867	1.01	[0.91, 1.11]
Caregiver-report CBQ sadness	3.45	0.96	1.00–5.63	435	1.08	[0.93, 1.24]
Mother-report CBQ shyness	3.54	1.10	1.00–6.60	952	0.91 *	[0.83, 1.00]
Caregiver-report CBQ shyness	3.46	0.29	1.00–6.89	670	0.88 *	[0.78, 0.99]
Intelligence at age 4						
WJ-R memory for sentences	92.39	18.36	17–142	962	0.65 ***	[0.59, 0.71]
WJ-R incomplete words	97.14	13.29	57–132	962	0.72 ***	[0.67, 0.78]
WJ-R picture vocabulary	100.55	14.90	10–143	965	0.68 ***	[0.63, 0.74]
Outcomes assessed in adolescence						
GPA	2.99	0.78	0.00–4.17	750	0.72 ***	[0.65, 0.80]
Standardized achievement test	105.86	13.53	60.50–160.00	770	0.59 ***	[0.52, 0.67]
BMI z score	0.56	0.98	−3.08–2.85	729	1.29 ***	[1.15, 1.44]
Risky behavior[a]	6.07	5.74	0–53	807	1.47 ***	[1.28, 1.69]

	M	SD	Observed range	n	Hazard ratio	95% CI
Demographics						
Income to needs[a]	3.56	2.73	0.10–20.20	948	0.66 ***	[0.60,0.71]
Maternal education (years)	14.42	2.47	7–21	966	0.69 ***	[0.63, 0.76]
Age	4.63	0.23	4.17–5.00	966	0.91	[0.83, 1.00]
Female	52%			966	0.91	[0.83, 1.00]
White	80%			966	0.72 ***	[0.66, 0.78]
Black	1 1%			966	1 41 ***	[1.31, 1.51]
Hispanic	5%			966	1.04	[0.95, 1.14]
Asian	1%			966	1.01	[0.93, I.I 1]
Other	3%			966	1.06	[0.98, 1.15]

Note: CI = confidence interval; CBQ = Child Behavior Questionnaire; WJ-R = Woodcock–Johnson Psycho-Educational Battery-revised; GPA = grade point average; BMI = body mass index.

[a] Mean, standard deviation, and range are based on raw scores; hazard ratio is based on log-transformed scores.

*
$p < .05$.

**
$p < .01$.

***
$p < .001$