Title
Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes

Journal
Psychological Science, 29(7)

ISSN
0956-7976

Authors
Watts, Tyler W
Duncan, Greg J
Quan, Haonan

Publication Date
2018-07-01

DOI
10.1177/0956797618761661

Peer reviewed

Research Article

Psychological Science
1–19

DOI: 10.1177/0956797618761661
www.psychologicalscience.org/PS

Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes

Tyler W. Watts[1], Greg J. Duncan[2], and Haonan Quan[2]
[1]Steinhardt School of Culture, Education, and Human Development, New York University, and
[2]School of Education, University of California, Irvine

Abstract

We replicated and extended Shoda, Mischel, and Peake's (1990) famous marshmallow study, which showed strong bivariate correlations between a child's ability to delay gratification just before entering school and both adolescent achievement and socioemotional behaviors. Concentrating on children whose mothers had not completed college, we found that an additional minute waited at age 4 predicted a gain of approximately one tenth of a standard deviation in achievement at age 15. But this bivariate correlation was only half the size of those reported in the original studies and was reduced by two thirds in the presence of controls for family background, early cognitive ability, and the home environment. Most of the variation in adolescent achievement came from being able to wait at least 20 s. Associations between delay time and measures of behavioral outcomes at age 15 were much smaller and rarely statistically significant.

Keywords

gratification delay, marshmallow test, achievement, behavioral problems, longitudinal analysis, early childhood, open data

Received 8/15/17; Revision accepted 1/21/18

In a series of studies based on children who attended a preschool on the Stanford University campus, Mischel, Shoda, and colleagues showed that under certain conditions, a child's success in delaying the gratification of eating marshmallows or a similar treat was related to later cognitive and social development, health, and even brain structure (Casey et al., 2011; Mischel et al., 2010; Shoda, Mischel, & Peake, 1990). Although only part of a larger research program investigating how children develop self-control, Mischel and Shoda's delay-time–later-outcome correlations and the preschooler videos accompanying them have become some of the most memorable findings from developmental research. Gratification delay is now viewed by many to be a fundamental "noncognitive" skill that, if developed early, can provide a lifetime of benefits (see Mischel et al., 2010, for a review).

Since the publication of Mischel and Shoda's seminal studies (e.g., Mischel, Shoda, & Peake, 1988; Mischel, Shoda, & Rodriguez, 1989; Shoda et al., 1990), other researchers have examined the processes underlying the ability to delay gratification. Some have modified the marshmallow test to illuminate the factors that affect a child's ability to delay gratification (e.g., Imuta, Hayne, & Scarf, 2014; Kidd, Palmeri, & Aslin, 2013; Michaelson & Munakata, 2016; Rodriguez, Mischel, & Shoda, 1989; Shimoni, Asbe, Eyal, & Berger, 2016); others have investigated the cognitive and socioemotional correlates of gratification delay (e.g., Bembenutty & Karabenick, 2004; Duckworth, Tsukayama, & Kirby, 2013; Romer, Duckworth, Sznitman, & Park, 2010). These studies have added to a growing body of literature on self-control suggesting that

Corresponding Author:
Tyler W. Watts, New York University, Steinhardt School of Culture, Education, and Human Development, 627 Broadway, 8th Floor, New York, NY, 10003
E-mail: tyler.watts@nyu.edu

gratification delay may constitute a critical early capacity. For example, Moffitt and Caspi demonstrated that self-control—typically understood to be an umbrella construct that includes gratification delay but also impulsivity, conscientiousness, self-regulation, and executive function—averaged across early and middle childhood, predicted outcomes across a host of adult domains (Moffitt et al., 2011). Duckworth and colleagues (2013) showed that the relation between early gratification delay and later outcomes was partially mediated by a composite measure of self-control, which has further fueled interventions designed to promote skills that fall under the "self-control" umbrella (e.g., Diamond & Lee, 2011). However, despite the proliferation of work on gratification delay, and the related construct of self-control, Mischel and Shoda's longitudinal studies still stand as the foundational examinations of the long-run correlates of the ability to delay gratification in early childhood.

Revisiting these studies reveals several limiting factors that warrant further investigation. First, Mischel and Shoda's reported longitudinal associations were based on very small and highly selective samples of children from the Stanford University community ($ns = 35–89$; Mischel et al., 1988; Mischel et al., 1989; Shoda et al., 1990). Although Mischel's original work included over 600 preschool-age children (Shoda et al., 1990), follow-up investigations focused on much smaller samples (e.g., for their investigation of SAT and behavioral outcomes, Shoda and colleagues were able to contact only 185 of the original 653 children). Moreover, these children originally underwent variations of the gratification-delay assessment; Mischel experimented with trials in which the treat was obscured from a child's vision, and some of the children were supplied with coping strategies to help them delay longer. They found positive associations between gratification delay and later outcomes only for children participating in trials in which no strategy was coached and the treat was clearly visible—a circumstance they called the "diagnostic condition."

For the 35 to 48 children who were tested in the diagnostic condition, and for whom adolescent follow-up data were available, Shoda and colleagues (1990) observed large correlations between delay time and SAT scores, $r(35) = .57$ for math, $r(35) = .42$ for verbal, and between delay time and parent-reported behaviors, for example, "[my child] is attentive and able to concentrate," $r(48) = .39$. These bivariate correlations were not adjusted for potential confounding factors that could affect both early delay ability and later outcomes. Because these findings have been cited as motivation both for interventions designed to boost gratification delay specifically (e.g., Kumst & Scarf, 2015; Murray, Theakston, & Wells, 2016; Rybanska, McKay, Jong, & Whitehouse, 2017) and for interventions seeking to promote self-control more generally (e.g., Diamond & Lee, 2011; Flook, Goldberg, Pinger, & Davidson, 2015; Rueda, Checa, & Cómbita, 2012), it is important to consider possible confounding factors that might lead bivariate correlations to be a poor projection of likely intervention effects.

In the current study, we pursued a conceptual replication of Mischel and Shoda's original longitudinal work. Specifically, we examined associations between performance on a modified version of the marshmallow test and later outcomes in a larger and more diverse sample of children, and we employed empirical methods that adjusted for confounding factors inherent in Mischel and Shoda's bivariate correlations. Several considerations motivated our effort. First, replication is a staple of sound science (Campbell, 1986; Duncan, Engel, Claessens, & Dowsett, 2014). Second, Mischel and Shoda's highly selective sample of children limits the generalizability of their results. Finally, if researchers are to extend Mischel and Shoda's work to develop interventions, a more sophisticated examination of the long-run correlates of early gratification delay is needed. Interventions that successfully boost early delay ability might have no effect on later life outcomes if associations between gratification delay and later outcomes are driven by factors unlikely to be altered by child-focused programs (e.g., socioeconomic status [SES], home parenting environment).

Current Study

We used data from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) to explore associations between preschoolers' ability to delay gratification and academic and behavioral outcomes at age 15. We focused most of our analysis on a sample of children born to mothers who had not completed college, for two reasons. First, it allowed us to investigate whether Mischel and Shoda's longitudinal findings extend to populations of greater interest to researchers and policymakers concerned with developing interventions (e.g., Mischel, 2014). Second, empirical concerns over the extent of truncation in our key gratification-delay measure in the college-educated sample limited our ability to reliably assess the correlation between gratification delay and later abilities. Because of these differences, we consider our study to be a conceptual, rather than traditional, replication of Mischel and Shoda's seminal work (Robins, 1978).

Method

More complete information regarding the study data and measures can be found in the Supplemental Material available online. Here, we provide a brief overview of key study components.

Data

Data for the current study were drawn from the NICHD SECCYD, a widely used data set in developmental psychology (NICHD Early Child Care Research Network, 2002). Participants were recruited at birth from 10 U.S. sites across the country, providing a geographically diverse, although not nationally representative, sample of children and mothers. Participants have been followed across childhood and adolescence, with the last full round of data collection occurring when children were 15 years old.

The current study relied on data collected when children were 54 months of age, and our outcome variables were measured during the assessments at Grade 1 and age 15. Our analysis sample was limited to children who had a valid measure of delay of gratification at age 54 months, as well as nonmissing achievement and behavioral data at age 15 ($n = 918$). For conceptual and analytic reasons (detailed below), we then split our sample on the basis of mother's education, and we focused much of our analyses on children whose mothers did not report having completed college when the child was 1 month old ($n = 552$, a sample that is 10 times larger than the sample size in the Shoda et al., 1990, study).

In Table 1, we present selected demographic characteristics for children included in our analytic sample, split by whether the child's mother did or did not receive a bachelor's degree. For purposes of comparison, we also present the same set of characteristics for a nationally representative sample of kindergarteners collected 2 to 3 years after our sample's 54-month wave of data

Table 1. Demographic Comparisons Between the Analytic Samples and a Nationally Representative Sample of Kindergarten Children (ECLS-K, 1998)

	NICHD SECCYD		ECLS-K, 1998
Variable	Children of nondegreed mothers	Children of degreed mothers	Nationally representative sample
Proportion male	.49	.46	.51
Proportion Black	.16	.02	.16
Proportion Hispanic	.07	.03	.19
Proportion White	.73	.91	.57
Mean age of mother (in years) at child's birth	26.84 (5.61)	31.67 (4.01)	27.28 (6.61)
Mother's education (proportions)			
Did not complete high school	.14	.00	.14
Graduated from high school	.32	.00	.29
Some college	.54	.00	.33
Bachelor's degree or higher	.00	1.00	.23
Income-to-needs ratio			
≤ 1	0.18	0	0.17
> 1 to ≤ 2	0.27	0.05	0.26
> 2 to ≤ 3	0.25	0.19	0.16
> 3 to ≤ 4	0.15	0.21	0.16
> 4	0.15	0.55	0.24
Proportion of mothers unemployed	.29	.23	.32
Mean number of children in home	2.32 (1.03)	2.16 (0.83)	2.49 (1.16)
Proportion of mothers married	.67	.93	.70
Number of observations	552	366	21,242

Note: Standard deviations are given in parentheses. The Early Childhood Longitudinal Survey—Kindergarten (ECLS-K) estimates were derived from data made publically available by the National Center for Education Statistics (https://nces.ed.gov/ecls/dataproducts.asp). All ECLS-K measures shown were collected during the fall of kindergarten (i.e., 1998), and National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) measures were collected during the 54-month interview (i.e., preschool; 1995–1996), except for mother's education and mother's age at child's birth, which were both collected at the 1-month interview. The ECLS-K variables were weighted using the C1CW0 weight to generate nationally representative estimates.

collection. These nationally representative data were drawn from the publically available Early Childhood Longitudinal Survey—Kindergarten Cohort, 1998–1999 (https://nces.ed.gov/ecls/dataproducts.asp; more information regarding this data set can be found in the Supplemental Material).

The children of college-completing mothers were largely White (91%), with 55% of them reporting family income that was at least 4 times above the poverty line (i.e., income-to-needs ratio over 4.0) and none of them reporting income at or below the poverty line (i.e., income-to-needs ratio at or below 1.0). The subsample of children with mothers without a college degree was more comparable with the nationally representative sample. In both samples, about 16% of children were Black, mother's age at birth was approximately 27 years, 14% of mothers did not complete high school, and between 17% and 18% of families were living at or below the poverty line. However, Hispanic children were still underrepresented in this sample, underscoring the fact that although diverse, our data were not nationally representative.

Measures

Delay of gratification. A variant of Mischel's (1974) self-imposed waiting task (i.e., the "marshmallow test") was administered to children when they were 54 months old. An interviewer would present children with an appealing edible treat based on the child's own stated preferences (e.g., marshmallows, M&M's, animal crackers). Children were then told that they would engage in a game in which the interviewer would leave the child alone in a room with the treat. If the child waited for 7 min, the interviewer would return, and the child could eat the treat and receive an additional portion as a reward for waiting. Children who chose not to wait could ring a bell to signal the experimenter to return early, and they would then receive only the amount of candy originally presented. The measure of delay of gratification was then recorded as the number of seconds the child waited, with 7 min being the ceiling.

The measure of gratification delay used here differed from the one employed by Mischel (1974) in several noteworthy ways. First, the 7-min cap was much shorter than Mischel's maximum assessment length; the children in Mischel's sample were asked to wait between 15 and 20 min, depending on the study, before the assessment ended. In our sample, approximately 55% of children hit the 7-min ceiling on the measure, presenting a potential analytic challenge to our models. However, we found that the ceiling was much more problematic for higher- than lower-SES children. Children whose mothers obtained college degrees hit the ceiling at a rate of 68%, compared with 45% for children whose mothers did not complete college ($p < .001$; see Table 2).

Table 2. Descriptive Characteristics of Key Analysis Variables

Variable	Children of nondegreed mothers (n = 552)	Children of degreed mothers (n = 366)	β	p value for difference
Delay of gratification (minutes waited)	3.99 (3.08)	5.38 (2.62)	0.45	.001
Delay of gratification (categories)				
7 min	.45	.68	0.21	.001
2–7 min	.16	.12	–0.02	.324
0.333–2 min	.16	.10	–0.06	.012
< 0.333 min	.23	.10	–0.13	.001
Outcome measures: Grade 1				
Achievement composite	108.42 (13.71)	117.29 (13.47)	0.63	.001
Behavior composite	49.15 (8.43)	47.40 (7.87)	–0.18	.008
Outcome measures: age 15				
Achievement composite	101.23 (11.63)	112.72 (13.19)	0.82	.001
Behavior composite	47.12 (9.37)	44.50 (8.66)	–0.27	.001

Note: In the columns for children with degreed and nondegreed mothers, the table reports the proportion of students falling within each delay-of-gratification category; all other values in these columns are means (with standard deviations in parentheses). The sample was split on the basis of mother's education, and p values were derived from a series of regressions in which each characteristic was regressed on a dummy for whether mother graduated from college and a series of site fixed effects. Beta values represent effect sizes measuring the standardized differences between the two groups.

We adopted several approaches to dealing with this truncation problem, principally exploring possible nonlinearities in the associations between time waited and outcome measures by dividing the distribution of waiting times into discrete intervals. We also focused much of our analyses on the children of mothers who did not complete college, as far fewer of the children in this sample hit the ceiling on the minutes-waited measure, and as explained above, this group of children complements the sample of children included in the Mischel and Shoda studies. But because the subsample of children with college-educated mothers allows for a more direct replication of Mischel and Shoda's famous work (e.g., Shoda et al., 1990), we also present results for them, bearing in mind the limitations imposed by the substantial delay truncation.

Finally, it should also be noted that children in the NICHD study were given only the version of the task that Shoda and colleagues (1990) called the diagnostic condition (i.e., the children were not offered strategies and were able to see the treat as they waited).

Academic achievement. Academic achievement was measured using the Woodcock-Johnson Psycho-Educational Battery Revised (WJ-R) test (Woodcock, McGrew, & Mather, 2001), a commonly used measure of cognitive ability and achievement (e.g., Watts, Duncan, Siegler, & Davis-Kean, 2014). For math achievement at Grade 1 and age 15, we used the Applied Problems subtest, which measured children's mathematical problem solving. At Grade 1, reading achievement was measured using the Letter-Word Identification task, a measure of word recognition and vocabulary, and at age 15, reading ability was measured using the Passage Comprehension test. The Passage Comprehension test asked students to read various pieces of text silently and then answer questions about their content.

For all the WJ-R tests, we used the standard scores, which were normed to have a mean of 100 and a standard deviation of 15 in each respective wave. We took the average of the Grade 1 math and reading measures and the age-15 math and reading measures, respectively, to create composite measures of academic achievement.

Behavioral problems. Following Shoda et al. (1990), we relied primarily on mothers' reports of child behavior. Mother-reported internalizing and externalizing behavioral problems were assessed using the Child Behavior Checklist (CBCL; Achenbach, 1991) at age 54 months, Grade 1, and age 15. The CBCL is a widely used measure of behavioral problems, and it includes approximately 100 items rated on 3-point scales that capture aspects of internalizing (i.e., depressive) and externalizing (i.e., antisocial) behavior. As with academic achievement, at Grade 1 and age 15, we averaged together the externalizing and internalizing measures to create a behavioral composite score that, before standardization, ranged from 32 to 83, with higher scores indicating higher levels of behavioral problems. We also tested models that used a host of alternative behavioral measures taken from youth reports and direct assessments at age 15; these measures and models are described in the Supplemental Material.

Additional covariates. All covariates included in our models are listed in Table 3, and we grouped the covariates into two distinct sets of control variables: *child background and Home Observation for Measurement of the Environment (HOME) controls* and *concurrent 54-month controls*.

Child background and HOME controls. Child demographic characteristics (i.e., gender and race), birth weight, mother's age at the child's birth, and mother's level of education were collected at the 1-month interview via interview with study mothers. Family income was collected from study mothers at the 1-, 6-, 15-, 24-, 36- and 54-month interviews. We took the average of all nonmissing income data over this span, and then log-transformed average family income to restrict the influence of outliers. Mother's Peabody Picture Vocabulary Test (PPVT) score was assessed in a lab visit when the focal child was 36 months old. The PPVT is a commonly used measure of intelligence (e.g., see meta-analysis by Protzko, 2015).

We also included early indicators of child cognitive functioning, as measured at age 24 months by the Bayley Mental Development Index (MDI; Bayley, 1991) and at age 36 months by the Bracken Basic Concept Scale (BBCS; Bracken, 1984). The MDI measured children's sensory-perceptual abilities, as well as their memory, problem solving, and verbal communication skills. The BBCS was an early measure of school readiness skills, and it required students to identify basic letters and numbers.

Child temperament was measured at age 6 months using the Early Infant Temperament Questionnaire (Medoff-Cooper, Carey, & McDevitt, 1993), a 38-item survey to which mothers responded. This questionnaire asked mothers to rate their child on a 6-point Likert-scale with items focused on the child's mood, adaptability, and intensity. We took the average score across these items as our measurement of temperament, with higher scores indicating more agreeable dispositions.

Finally, the set of controls measured prior to age 54 months also included indicators of the quality of the home environment, as measured by an observational assessment called the HOME inventory (Caldwell & Bradley, 1984). The HOME was assessed when the focal child was approximately 36 months old, and it was

Table 3. Descriptive Characteristics of All Control Variables

Variable	Children of nondegreed mothers				Children of degreed mothers			
	Waited 7 min (n = 251)	Did not wait 7 min (n = 301)	β	p value for difference	Waited 7 min (n = 250)	Did not wait 7 min (n = 116)	β	p value for difference
Child background and HOME controls								
Child background								
Proportion male	.47	.51	–0.04	.338	.45	.50	–0.05	.409
Proportion White	.82	.64	0.18	.001	.94	.85	0.10	.007
Proportion Black	.07	.24	–0.15	.001	.00	.05	–0.05	.024
Proportion Hispanic	.06	.07	–0.01	.545	.03	.03	–0.00	.962
Proportion other race/ethnicity	.04	.05	–0.01	.530	.03	.07	–0.05	.058
Child's age at delay measure (months)	56.11 (1.11)	56.01 (1.14)	0.13	.105	55.99 (1.13)	55.99 (1.15)	0.07	.519
Birth weight (g)	3490.23 (478.56)	3449.02 (540.26)	0.09	.320	3516.63 (520.52)	3572.53 (527.17)	–0.13	.268
BBCS standard score (36 months)	9.06 (2.56)	7.67 (2.86)	0.47	.001	10.67 (2.20)	10.14 (2.35)	0.19	.043
Bayley MDI (24 months)	93.89 (12.40)	85.91 (14.40)	0.53	.001	100.88 (11.78)	95.21 (14.10)	0.41	.001
Child temperament (6 months)	3.18 (0.42)	3.25 (0.38)	–0.17	.053	3.13 (0.37)	3.09 (0.43)	0.07	.531
Log of family income (1–54 months)	0.89 (0.61)	0.57 (0.73)	0.38	.001	1.54 (0.51)	1.42 (0.56)	0.14	.057
Mother's age at birth (years)	27.75 (5.66)	26.07 (5.46)	0.29	.001	31.58 (4.05)	31.87 (3.91)	–0.06	.438
Mother's education (years)	13.00 (1.41)	12.68 (1.50)	0.12	.017	17.02 (1.31)	16.82 (1.26)	0.07	.234
Mother's PPVT score	96.43 (13.38)	90.47 (17.03)	0.30	.001	114.10 (15.62)	105.63 (16.51)	0.44	.001
HOME score (36 months)								
Learning Materials	7.20 (2.36)	5.86 (2.51)	0.53	.001	8.64 (1.59)	8.41 (2.20)	0.12	.168
Language Stimulation	6.13 (1.04)	5.67 (1.24)	0.46	.001	6.38 (0.84)	6.17 (1.13)	0.21	.046
Physical Environment	6.16 (1.04)	5.64 (1.54)	0.40	.001	6.35 (0.83)	6.33 (0.91)	0.07	.372
Responsivity	5.67 (1.28)	5.17 (1.52)	0.31	.001	6.09 (0.99)	5.81 (1.30)	0.21	.033
Academic Stimulation	3.43 (1.21)	2.97 (1.29)	0.38	.001	3.74 (0.97)	3.57 (1.29)	0.17	.112
Modeling	3.13 (1.10)	2.82 (1.14)	0.29	.001	3.64 (0.93)	3.51 (1.04)	0.11	.285
Variety	6.80 (1.34)	6.14 (1.50)	0.45	.001	7.54 (1.17)	7.29 (1.36)	0.17	.088
Acceptance	3.39 (0.85)	3.22 (1.04)	0.18	.038	3.70 (0.59)	3.57 (0.82)	0.13	.162
Responsivity-Empirical Scale	5.54 (0.91)	5.14 (1.29)	0.37	.001	5.77 (0.52)	5.55 (0.91)	0.21	.026
Concurrent 54-month controls								
54-month WJ-R score								
Letter-Word Identification	99.03 (11.98)	93.22 (12.63)	0.42	.001	105.93 (12.19)	102.31 (11.94)	0.26	.011
Applied Problems	104.80 (12.88)	95.67 (15.72)	0.57	.001	112.36 (12.13)	106.06 (12.31)	0.40	.001
Picture Vocabulary	100.54 (13.07)	93.74 (13.80)	0.43	.001	109.11 (13.45)	103.47 (13.58)	0.36	.001

(*continued*)

Table 3. (Continued)

Variable	Children of nondegreed mothers				Children of degreed mothers			
	Waited 7 min (n = 251)	Did not wait 7 min (n = 301)	β	p value for difference	Waited 7 min (n = 250)	Did not wait 7 min (n = 116)	β	p value for difference
Memory for Sentences	93.21 (15.59)	85.43 (17.67)	0.43	.001	100.99 (18.73)	92.34 (17.45)	0.49	.001
Incomplete Words	98.08 (12.91)	92.72 (13.52)	0.41	.001	102.18 (11.69)	98.05 (11.98)	0.35	.001
54-month Child Behavior Checklist								
Internalizing	47.36 (9.11)	47.94 (8.51)	–0.06	.477	46.55 (8.84)	46.81 (8.17)	–0.01	.988
Externalizing	51.14 (9.34)	53.09 (9.84)	–0.21	.020	50.44 (9.11)	50.99 (8.53)	–0.06	.604

Note: In the columns for children who did and did not wait 7 min, the table reports proportions for race/ethnicity; all other values in these columns are means (with standard deviations in parentheses). The p value column compares children who successfully completed the task and waited 7 min with children who did not, and the betas represent effect sizes measuring the standardized differences between the two groups. A series of regressions in which each variable was regressed on a dummy indicating whether the child completed the marshmallow test was used to generated p values, and a series of site dummy variables was also included to adjust for site differences (ps below .001 have been rounded to .001). BBCS = Bracken Basic Concept Scale; HOME = Home Observation for Measurement of the Environment; MDI = Mental Development Index; PPVT = Peabody Picture Vocabulary Test; WJ-R = Woodcock-Johnson Psycho-Educational Battery Revised.

designed to capture aspects of the home environment known to support positive cognitive, emotional, and behavioral functioning. We used nine subscales of the HOME in our models: The first eight subscales are commonly used with the HOME measure (Learning Materials, Language Stimulation, Physical Environment, Responsivity, Academic Stimulation, Modeling, Variety, and Acceptance), and the ninth subscale, called the Responsivity-Empirical Scale, was derived by the NICHD SECCYD study from factor analyses of the HOME items. This final scale was distinct from the traditional Responsivity scale, as it included items from the Language Stimulation scale that also measured mother responsivity and sensitivity to the child.

Concurrent 54-month controls. For models that included controls for concurrent cognitive and behavioral skills, we also included subscales taken from the age 54-month WJ-R test. As our measure of early reading, we included the Letter-Word Identification task, which tested children's ability to sound out simple words, and the Applied Problems test at age 54 months was our measure of early math skills. For preschool children, the Applied Problems test requires them to count and solve simple addition problems. We also used the Memory for Sentences and Incomplete Words subtests as measures of cognitive ability. The Incomplete Words test measured auditory closure and processing, and children listened to an audio recording where words missing a phoneme were listed. They were then asked to name the complete word. Finally, the Picture Vocabulary test was a measure of verbal comprehension and crystallized intelligence. In this task, children were asked to name pictured objects. All of these tasks have been widely used as measures of children's early cognitive skills and their measurement properties have been widely reported (e.g., Watts et al., 2014).

Finally, we also included the mother's report of children's externalizing and internalizing problems from the Child Behavior Checklist at age 54 months. Much like the measure used for age-15 behavioral problems, the 54-month survey included a battery of items designed to assess children's antisocial and disruptive behavior (i.e., externalizing) and depressive symptoms (i.e., internalizing).

Analysis

Our primary goal was to estimate the association between early gratification delay and long-run measures of academic achievement and behavioral functioning. Like the work of Shoda and colleagues (1990), our study did not include a measure of gratification delay in which between-child differences were generated from some exogenous intervention, so we do not claim that the associations we estimated reflect causal impacts. Instead, our goal was to assess how much bias might be contained in longitudinal bivariate correlations between gratification delay and later outcomes as a result of failure to control for characteristics of children and their environments. Regression-adjusted correlations should provide better guidance regarding whether interventions boosting gratification delay might also improve later achievement and behavior.

To accomplish our analytic goals, we modeled later academic achievement and behavior (measured at both Grade 1 and age 15) as a function of a measure of

gratification delay at age 54 months. We then tested models that added controls for background characteristics and measures of the home environment before moving to models that also included measures of cognitive and behavioral skills assessed at age 54 months (see Table 3).

These two approaches reflect different assumptions regarding how variation in gratification-delay ability might arise. Models with controls measured between birth and age 36 months still allow for variation in age 54-months gratification delay caused by the differential development of general cognitive or behavioral skills (e.g., executive function, self-control) between 36 and 54 months. Put another way, these models contain controls only for factors that even ambitious preschool-child-focused interventions are unlikely to alter (e.g., birth weight, temperament at 6 months of age, early home environment).

In contrast, the models with concurrent-54-months covariates controlled for variation in a range of cognitive capacities and behavioral problems developed by age 54 months. They helped to isolate the possible effects of an intervention that targets only the narrow set of skills involved with gratification delay (e.g., a program that merely provided children with strategies to help them delay longer; see Mischel, 2014, p. 40) but not concurrent general cognitive ability or socioemotional behaviors.

Although it is impossible to know exactly how individual differences in gratification delay emerge (e.g., changes in parenting, development of cognitive skills), by controlling for factors unlikely to be altered by interventions (e.g., ethnicity, parental background), we can purge our estimates of bias due to observable characteristics that are correlated with gratification delay and later outcomes. If remaining unobserved factors also contribute to gratification delay and later outcomes (e.g., changes in parenting), and if these unobserved factors are unlikely to be altered by a particular intervention, then bias in our estimates may still remain. Yet our estimates should serve as an improvement over the unadjusted correlations reported previously (e.g., Shoda et al., 1990).

In all models shown, continuous variables were standardized so that coefficients could be read as effect sizes, and all models with control variables included a set of dummy variables for each site to adjust for any between-site differences. In order to account for missing data on control variables, we used structural equation modeling with full information maximum likelihood in Stata Version 15.0 (StataCorp, 2017) to estimate all analytic models. Finally, we report all estimated p values to the thousandth decimal place (with p values below .001 displayed as $<.001$), and we describe any estimate corresponding to a p value less than .05 as statistically significant. Though we recognize the arbitrariness of focusing only on results with a p value less than .05, we selected this alpha level because it was the minimum threshold for statistical significance used in the studies we attempted to replicate and extend (i.e., Mischel et al., 1988; Mischel et al., 1989; Shoda et al., 1990). Consequently, any differences in conclusions reached between our studies and those reported in the previous literature should be attributed to design and sample differences rather than alpha-level choices.

Results

Descriptive findings

Table 2 provides descriptive results for key analysis variables, including the 54-months delay-of-gratification measure, split by mother's education level. In the sample of children with nondegreed mothers, children waited an average of 3.99 min ($SD = 3.08$) before ending the task. We also present the proportion of children falling within certain ranges on the measure, with the 7-min category representing children who successfully completed the trial. In the lower-SES sample, 45% of children waited the maximum of 7 min, and 23% waited less than 20 s (i.e., 0.33 min). In the higher-SES sample, only 10% of children waited less than 20 s, and the average time waited was 5.38 min (statistically significantly longer than the lower-SES group, $p < .001$).

Because the 7-min ceiling presented a potential analytic challenge for both samples, we estimated models that substituted the four dummy categories shown in Table 2 for the continuous minutes-waited variable as a way to assess nonlinearities in the relationship between delay time and academic and socioemotional outcomes. Importantly, these models also provide information on how much our analysis might be compromised by the 7-min truncation.

Table 3 presents descriptive information for the various control measures used in the analysis, and means are presented separately for children who did and did not complete the delay task. In both the higher- and lower-SES samples, performance on the delay-of-gratification task was highly correlated with differences on most observable characteristics considered. For example, for children from nondegreed mothers, those who completed the delay-of-gratification task were from higher income families ($p < .001$) than noncompleters, had mothers with higher PPVT scores ($p < .001$), and had higher scores on dimensions of the HOME observational assessment (ps = .04 to $<.001$). Null or smaller differences were generally observed for the children of

degreed mothers, perhaps owing to the lack of heterogeneity in this subsample.

Regression results

Results for children of nondegreed mothers. Table 4 presents coefficients and standard errors from models that estimated the association between delay of gratification at 54 months and our Grade 1 and age-15 achievement and behavioral composites for the sample of children from nondegreed mothers. Table 4 displays the results for a standardized continuous measure of gratification delay (i.e., the number of minutes waited during the marshmallow test). As Column 1 reflects, the bivariate association between minutes waited and academic achievement was 0.28 (SE = 0.04, $p < .001$), considerably less than the .57 correlation Shoda and colleagues found for SAT math scores and the .42 correlation they found for verbal scores. These linear results suggest that children's Grade 1 achievement would improve by approximately one tenth of a standard deviation for every additional minute waited at age 4. When the controls measured prior to age 54 months (second column of Table 3) were added to the model, the standardized association fell to 0.10 (SE = 0.03, p = .002), and when concurrent 54-months controls were added (third column of Table 1), the association fell to a statistically nonsignificant 0.05 (SE = 0.03, p = .114).

Columns 4 through 6 show analogous models for the measure of achievement at age 15. The magnitudes of the age-15 correlations were remarkably similar to the Grade 1 correlations. The age-15 achievement correlation in the absence of other controls was of moderate size and statistically significant, β = 0.24, SE = 0.04, $p < .001$; but fell substantially when controls for earlier characteristics were added, β = 0.08, SE = 0.03, p = .016; and became nonsignificant when 54-months controls were added, β = 0.05, SE = 0.03, p = .140. Given that Shoda and colleagues found almost as strong correlations with later behavior as with later achievement, we were surprised to find virtually no relationship—even in the absence of controls—between delay of gratification and the composite score of mother-reported internalizing and externalizing at either Grade 1 or age 15 (right half of Table 4).

Children who waited less than 20 s (i.e., the lowest category) served as the comparison group for our models that represented delay times in a set of dummy variables (see Table 2 for the proportion of students in each category). As shown in Table 4, models of outcomes at both Grade 1 and age 15 that lack control variables show a strong gradient between gratification delay and later achievement. Relative to children who waited less than 20 s, children who waited between 20 s and 2 min scored about one third of a standard deviation higher on the achievement measure at Grade 1 and age 15, and this difference grew to nearly three fourth of a standard deviation for the group that waited the entire 7 min. The entry for Model 1 in the row labeled "p value for test of equality of second, third, and fourth categories" shows that the coefficients produced by the three groups of children who waited longer than 20 s differed significantly from one another ($p < .001$), as did coefficient differences across all four categorical variables (the p value that is shown in the row labeled "p value for test of equality of all categories").

At both Grade 1 and age 15, when controls for early child and family characteristics were added to the model (Column 2 for Grade 1; Column 5 for age 15), the coefficients estimated for all three delay-time groups fell by roughly 50%. Surprisingly, the addition of the background controls also flattened out the gradient of the prediction across the gratification-delay distribution. Relative to the less-than-20-s reference group, achievement differences for children who waited more than 20 s but not the full 7 min were strikingly similar to the difference for children who waited the full 7 min. At age 15, the threshold nature of the relationship was most apparent; the coefficients produced by the three groups that waited longer than 20 s all fell between 0.23 and 0.30, and were not close to being statistically significantly different from one another (p = .752).

When concurrent 54-months controls were added, coefficients fell even further. At age 15, only the coefficient produced by the group describing children who waited 2 to 7 min retained statistical significance (β = 0.24, SE = 0.10, p = .018), though once again the set of coefficients on the included categories of delay time did not differ from one another (p = .630). As with the models shown for delay minutes in the achievement-composite columns in Table 4, we found no statistically significant relationships between gratification delay and the first-grade and age-15 behavioral composites.

In our focal case of age-15 achievement, the return for delaying gratification appeared to be driven by differences between children who managed to wait at least 20 s and those who did not. Figure 1 illustrates this threshold effect with three lines showing the coefficients produced by our delay-of-gratification categories in the age-15 achievement models (i.e., the "Delay minutes (categorical)" section of Table 4). The solid line shows coefficients drawn from the no-control model (i.e., Column 4 of Table 4), the dashed line shows coefficients from the model with early controls (i.e., Column 5 of Table 4), and the dotted line shows coefficients produced by models with the 54-months controls (i.e., Column 6 of Table 4).

The uncontrolled line has a steep initial jump, followed by a more gradual increase for wait times longer than 20 s. Both lines for the models with controls

Table 4. Associations Between Delay of Gratification at Age 54 Months and Later Measures of Academic Achievement and Behavior for Children of Mothers Without College Degrees

	Achievement composite						Behavior composite					
	Grade 1			Age 15			Grade 1			Age 15		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Delay minutes (continuous)	0.279*	0.102*	0.047	0.236*	0.081*	0.050	–0.060	–0.015	0.023	–0.062	–0.026	0.003
	(0.038)	(0.033)	(0.030)	(0.037)	(0.034)	(0.032)	(0.043)	(0.044)	(0.044)	(0.046)	(0.047)	(0.042)
Delay minutes (categorical)												
<0.333 min	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333–2 min	0.298*	0.189	0.127	0.353*	0.230*	0.178	0.055	0.090	0.079	–0.140	–0.071	–0.106
	(0.126)	(0.105)	(0.093)	(0.122)	(0.103)	(0.098)	(0.144)	(0.138)	(0.105)	(0.152)	(0.148)	(0.132)
2–7 min	0.424*	0.206	0.041	0.457*	0.300*	0.235*	–0.088	–0.020	0.039	–0.182	–0.109	–0.053
	(0.126)	(0.104)	(0.093)	(0.123)	(0.103)	(0.099)	(0.144)	(0.137)	(0.106)	(0.151)	(0.145)	(0.131)
7 min	0.720*	0.284*	0.141	0.646*	0.234*	0.150	–0.121	–0.007	0.072	–0.193	–0.095	–0.048
	(0.098)	(0.086)	(0.078)	(0.098)	(0.088)	(0.084)	(0.112)	(0.114)	(0.087)	(0.120)	(0.123)	(0.111)
p value for test of equality of all categories	.001	.012	.247	.001	.015	.093	.477	.866	.837	.428	.861	.885
p value for test of equality of second, third, and fourth categories	.001	.563	.475	.015	.752	.630	.382	.700	.923	.927	.969	.882
Control variables included												
Child background and HOME	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Concurrent 54 month	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Note: $n = 552$. For the continuous and categorical measures of delay minutes, the table gives standardized coefficients (with standard errors in parentheses). For the categorical measure, <0.333 min was the reference category. Because outcome variables were standardized, coefficients can be interpreted as effect sizes. Estimates shown in the first column of each set (i.e., Columns 1, 4, 7, and 10) contained only the measure of delay of gratification and a given outcome measure. Estimates shown in the second column of each set (i.e., Columns 2, 5, 8, and 11) added child background characteristics, Home Observation for Measurement of the Environment (HOME) scores, and site dummy variables. Estimates shown in the third column of each set (i.e., Columns 3, 6, 9, and 12) added other behavioral and cognitive measures also measured at age 54 months. Post hoc chi-square tests were used to generate *p* values in order to assess whether respective sets of variables were different from one another (*p*s below .001 have been rounded to .001).

*$p < .05$.

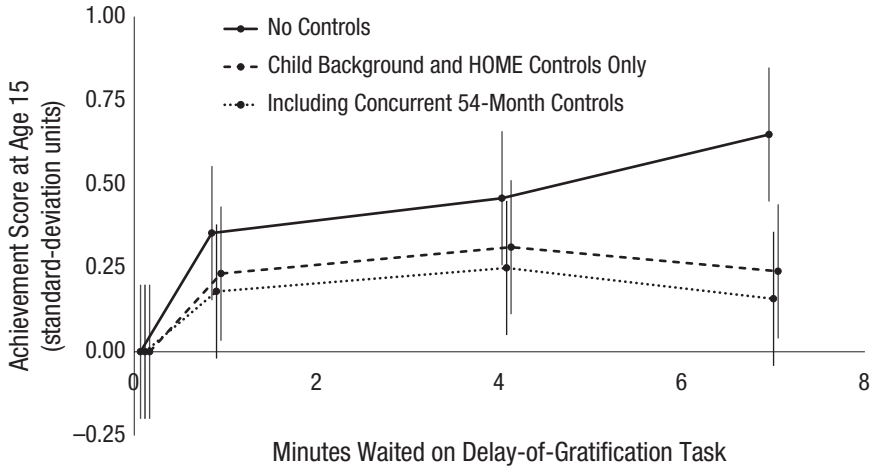

Fig. 1. Predicted achievement score by minutes of delay for children of mothers with no college degree. Error bars represent 95% confidence intervals. Values are shown separately for each of the four delay-of-gratification groups (< 0.333 min, 0.333–2 min, 2–7 min, 7 min); the *x*-axis shows the deviation in achievement composite scores from the reference group (delay < 0.333 min) against the within-group average amount of time waited. The average wait times for the models with no controls and with child background and Home Observation for Measurement of the Environment (HOME) controls only are displaced by ±.025 to distinguish the sets of error bars. The high-delay group's coefficients are plotted at 7 min, although the 7-min truncation prevents us from knowing what the mean value of minutes waited would have been for this group in the absence of this limit.

decrease after 4 min. Using 7 min to anchor the more-than-7-min group is probably an underestimate, but it is clear from the downward trajectory that no assumptions about the distribution of wait times above 7 min would produce a strong positive slope for the last segment of the line. Thus, in the case of children with mothers who lack college degrees, the truncation of delay time at 7 min does not affect the conclusion that children with the highest delay times show similar achievement levels at age 15 as other children who are able to delay for at least 20 s.

Results for children from mothers with college degrees. In Table 5, we present key results for children of mothers with college degrees. As in Table 4, we again present results for the continuous measure of delay of gratification and the categorical measures split along parts of the delay-of-gratification distribution. For the continuous measure, we again found evidence of positive unadjusted associations between delay of gratification and later achievement at both Grade 1 ($\beta = 0.18$, $SE = 0.06$, $p = .001$) and age 15 ($\beta = 0.17$, $SE = 0.06$, $p = .007$), and the categorical results suggested that much of this association was somewhat linear through the distribution. For the age-15 models, these relations became statistically indistinguishable from zero once controls were added, and the point estimate for the more-than-7-min category was surprisingly small and negative ($\beta = -0.04$, $SE = 0.15$, $p = .816$). As with the models shown in Table 4, we again found no evidence of associations between delay of gratification and the behavioral measures at first grade or age 15 in the high-SES sample.

Despite statistically nonsignificant results, point estimates were sometimes positive and substantial (e.g., the 2–7 min group coefficient shown in Column 1; $\beta = 0.40$, $SE = 0.21$, $p = .054$), but the standard errors were nearly double those estimated for children of nondegreed mothers (Table 4). This is due in part to the somewhat smaller sample size for the higher-SES sample but also to the lack of variation in the delay-of-gratification measure for this sample. Thus, although we found even less evidence of associations between delay of gratification and measures of later achievement when considering only the children of mothers with college degrees, it is difficult to draw strong conclusions from these models given the imprecise nature of their coefficient estimates.

Additional results and sensitivity checks

Heterogeneity. Because we found little evidence supporting associations between early delay ability and later outcomes for the higher-SES sample, we next tested whether the different pattern of results observed between the higher- and lower-SES samples constituted a statistically significant difference. In Table 6, we present models that included interaction terms between the various

Table 5. Associations Between Delay of Gratification at Age 54 Months and Later Measures of Academic Achievement and Behavior for Children of Mothers With College Degrees

	Achievement composite						Behavior composite					
	Grade 1			Age 15			Grade 1			Age 15		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Delay minutes (continuous)	0.178*	0.120*	0.048	0.167*	0.062	0.007	–0.049	–0.059	–0.050	0.031	0.038	0.043
	(0.056)	(0.053)	(0.045)	(0.062)	(0.059)	(0.054)	(0.057)	(0.061)	(0.046)	(0.059)	(0.063)	(0.055)
Delay minutes (categorical)												
<0.333 min	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333–2 min	0.327	0.039	0.148	0.079	–0.131	–0.085	–0.069	–0.088	–0.184	–0.065	0.027	–0.083
	(0.220)	(0.198)	(0.168)	(0.245)	(0.216)	(0.197)	(0.227)	(0.228)	(0.173)	(0.231)	(0.232)	(0.200)
2–7 min	0.397	0.147	0.134	0.216	0.028	–0.032	–0.277	–0.240	–0.265	–0.318	–0.217	–0.227
	(0.206)	(0.184)	(0.155)	(0.227)	(0.199)	(0.182)	(0.210)	(0.209)	(0.157)	(0.218)	(0.216)	(0.185)
7 min	0.562*	0.301	0.193	0.404*	0.077	–0.036	–0.194	–0.208	–0.214	–0.007	0.068	0.052
	(0.166)	(0.154)	(0.131)	(0.183)	(0.166)	(0.152)	(0.168)	(0.174)	(0.131)	(0.174)	(0.180)	(0.155)
p value for test of equality of all categories	.005	.100	.521	.059	.674	.979	.515	.584	.350	.267	.367	.227
p value for test of equality of second, third, and fourth categories	.238	.153	.843	.149	.477	.948	.629	.753	.867	.147	.206	.115
Control variables included												
Child background and HOME	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Concurrent 54 month	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Note: $n = 366$. For the continuous and categorical measures of delay minutes, the table gives standardized coefficients (with standard errors in parentheses). For the categorical measure, <0.333 min was the reference category. Because outcome variables were standardized, coefficients can be interpreted as effect sizes. Estimates shown in the first column of each set (i.e., Columns 1, 4, 7, and 10) contained only the measure of delay of gratification and a given outcome measure. Estimates shown in the second column of each set (i.e., Columns 2, 5, 8, and 11) added child background characteristics, Home Observation for Measurement of the Environment (HOME) scores, and site dummy variables. Estimates shown in the third column of each set (i.e., Columns 3, 6, 9, and 12) added other behavioral and cognitive measures also measured at age 54 months. Post hoc chi-square tests were used to generate *p* values in order to assess whether respective sets of variables were different from one another (*p*s below .001 have been rounded to .001). Estimates in this table can be directly compared with estimates from Table 4. The sample was limited to children whose mothers had completed at least 16 years of education (i.e., completed college).

*$p < .05$.

Table 6. Associations Between Delay of Gratification at Age 54 Months and Later Measures of Academic Achievement With Interactions Between Delay of Gratification and Socioeconomic Status

	Achievement composite						Behavior composite					
	Grade 1			Age 15			Grade 1			Age 15		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Delay minutes (continuous)	0.279*	0.115*	0.050	0.236*	0.083*	0.040	–0.059	–0.019	0.012	–0.062	–0.023	0.009
	(0.038)	(0.035)	(0.030)	(0.040)	(0.037)	(0.034)	(0.042)	(0.043)	(0.033)	(0.044)	(0.046)	(0.040)
High-SES indicator	0.509*	0.050	0.032	0.747*	0.270*	0.266*	–0.187*	0.026	0.031	–0.286*	–0.119	–0.127
	(0.064)	(0.068)	(0.059)	(0.067)	(0.071)	(0.066)	(0.070)	(0.084)	(0.064)	(0.074)	(0.088)	(0.077)
Interaction	–0.101	–0.043	–0.035	–0.069	–0.007	–0.018	0.010	–0.038	–0.058	0.094	0.040	0.017
	(0.067)	(0.058)	(0.050)	(0.069)	(0.061)	(0.057)	(0.073)	(0.071)	(0.054)	(0.076)	(0.075)	(0.066)
Delay minutes (categorical)												
<0.333 min	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333–2 min	0.298*	0.182	0.109	0.353*	0.202	0.151	0.055	0.060	0.050	–0.140	–0.082	–0.097
	(0.127)	(0.110)	(0.096)	(0.131)	(0.115)	(0.107)	(0.140)	(0.137)	(0.104)	(0.148)	(0.145)	(0.127)
2–7 min	0.424*	0.215	0.053	0.457*	0.288*	0.199	–0.088	–0.046	0.006	–0.182	–0.103	–0.024
	(0.127)	(0.110)	(0.097)	(0.132)	(0.115)	(0.108)	(0.140)	(0.137)	(0.105)	(0.146)	(0.143)	(0.126)
7 min	0.721*	0.308*	0.147	0.646*	0.222*	0.121	–0.121	–0.025	0.034	–0.193	–0.087	–0.028
	(0.099)	(0.090)	(0.079)	(0.105)	(0.097)	(0.091)	(0.109)	(0.112)	(0.086)	(0.116)	(0.120)	(0.106)
High-SES indicator	0.585*	0.154	0.041	0.951*	0.428*	0.417*	–0.097	0.163	0.191	–0.375	–0.185	–0.138
	(0.174)	(0.156)	(0.136)	(0.178)	(0.163)	(0.151)	(0.187)	(0.190)	(0.144)	(0.195)	(0.199)	(0.174)
Interactions												
High SES × < 0.333 min	0.029	–0.164	0.032	–0.274	–0.337	–0.266	–0.124	–0.127	–0.160	0.075	0.119	0.035
	(0.252)	(0.218)	(0.190)	(0.259)	(0.226)	(0.210)	(0.275)	(0.269)	(0.205)	(0.284)	(0.276)	(0.243)
High SES × 2–7 min	–0.027	–0.138	0.010	–0.241	–0.293	–0.258	–0.188	–0.185	–0.199	–0.136	–0.090	–0.156
	(0.240)	(0.206)	(0.179)	(0.246)	(0.213)	(0.198)	(0.260)	(0.252)	(0.192)	(0.272)	(0.261)	(0.229)
High SES × 7 min	–0.159	–0.119	–0.033	–0.242	–0.119	–0.134	–0.073	–0.167	–0.203	0.186	0.115	0.049
	(0.192)	(0.165)	(0.144)	(0.197)	(0.173)	(0.161)	(0.207)	(0.201)	(0.153)	(0.217)	(0.212)	(0.186)
p value from interaction-term joint *F* test	.668	.870	.968	.640	.342	.507	.899	.859	.610	.450	.753	.720
Control variables included												
Child background and HOME	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Concurrent 54 month	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Note: $n = 918$. For the continuous and categorical measures of delay minutes, the table gives standardized coefficients (with standard errors in parentheses). For the categorical measure, <0.333 min was the reference category. Because continuous variables were standardized, coefficients can be interpreted as effect sizes. Estimates shown in the first column of each set (i.e., Columns 1, 4, 7, and 10) contained only the measure of delay of gratification and a given outcome measure. Estimates shown in the second column of each set (i.e., Columns 2, 5, 8, and 11) added child background characteristics, Home Observation for Measurement of the Environment (HOME) scores, and site dummy variables. Estimates shown in the third column of each set (i.e., Columns 3, 6, 9, and 12) added other behavioral and cognitive measures also measured at age 54 months. Post hoc chi-square tests were used to generate *p* values in order to assess whether respective sets of variables were different from one another (*p*s below .001 have been rounded to .001). The joint *F* test evaluated whether the set of interaction terms jointly contributed to the model. In other words, it tested whether the set of interactions were statistically significantly different from zero.
*$p < .05$.

measures of delay of gratification (i.e., the continuous and categorical measures) and the indicator for whether the participant's mother completed college. None of the interactions tested were statistically significant, and our series of joint *F* tests indicated that the set of interactions for the categorical measures of delay of gratification did not statistically significantly contribute to any of the models (*p*s = .342–.968). However, as with the models that were run solely on the sample of children with college-educated mothers, standard errors were quite large for the interaction terms, indicating a substantial level of statistical imprecision. Unfortunately, the wide confidence intervals on many of the interaction terms render it impossible to provide a definitive answer to whether the relation between early delay ability and later achievement differs by SES.

Measurement considerations. In Table 7, we present correlations between the marshmallow test and all analysis variables for the full sample of children considered in our analyses ($n = 918$; see the Supplemental Material for correlation matrices for both the lower-SES and higher-SES samples, respectively). In Table 7, we also included the 54-month measure of the Continuous Performance Task (CPT; Barkley, 1994), which is a commonly used indicator of attention and impulsivity, and we included the Duckworth et al. (2013) parent- and teacher-report index of 54-month self-control (see the Supplemental Material for measurement details). We included these additional measures to further investigate how the marshmallow test might relate to theoretically relevant constructs (see Diamond & Lee, 2011). Surprisingly, the marshmallow test had the strongest correlation with the Applied Problems subtest of the WJ-R, $r(916) = .37$, $p < .001$; and correlations with measures of attention, impulsivity, and self-control were lower in magnitude (*r*s = .22–.30, $p < .001$). Although these correlational results were far from conclusive, they suggest that the marshmallow test should not be thought of as a mere behavioral proxy for self-control, as the measure clearly relates strongly to basic measures of cognitive capacity.

In the Supplemental Material, we report further assessments of the extent to which self-control and attention could account for the associations between delay of gratification and later achievement. In Table S3, we included the 54-months measures of attention and impulse control taken from the CPT in the Table 4 models and found that inclusion of the CPT measures accounted for only 21% to 27% of the effect for the less-than-7-min group. In Table S4, we present results from a parallel analysis using the Duckworth et al. (2013) index of self-control, and again we found that coefficients were hardly reduced when the self-control index was included. The small change in the coefficient for the delay-of-gratification measure between models that did and did not include indicators of attention, impulsivity, and self-control raises further questions regarding what constructs are measured by the marshmallow test.

Alternative outcome measures. Returning to our focal sample of children with mothers who had not completed college, we were surprised to see the lack of significant associations between our delay-of-gratification measure and the behavioral measures at Grade 1 and age 15. We also tested models that used alternative indicators of behavior assessed at age 15, including measures of risky behavior from youth self-reports and assessments of impulse control. Surprisingly, we still found virtually no associations between delay of gratification and behavior across any of these alternative measures (Tables S5–S7 in the Supplemental Material). Furthermore, because we relied on aggregated measures of achievement and behavior, we also tested separate models for math, reading, externalizing behaviors, and internalizing behaviors (Table S8 in the Supplemental Material). Results indicated that the achievement associations were similar for both the math and reading measures, and we still found no statistically significant effects on either measure of problem behaviors.

Discussion

We attempted to extend the famous findings of Mischel and Shoda (Mischel et al., 1988; Mischel et al., 1989; Shoda et al., 1990) by examining associations between early delay of gratification and adolescent outcomes in a more diverse sample of children and with more sophisticated statistical models. As with the earlier studies, we found statistically significant, although smaller, bivariate associations between early delay ability and later achievement. But we also found that these associations were highly sensitive to the inclusion of controls. Moreover, we failed to find even bivariate associations between delay of gratification at age 54 months and a host of behavioral outcomes at age 15, which was remarkable given the stability in self-control measures found in other studies (e.g., Moffitt et al., 2011).

It surprised us that for the children of nondegreed mothers, most of the achievement boost for early delay ability was gained by waiting a mere 20 s. Shoda et al. (1990) argued that the relationship between delay of gratification and academic achievement might be driven by the ability to generate useful metacognitive strategies that will influence self-regulation throughout one's life. Such strategies are unlikely to have played much of a role in a child's ability to wait for only 20 s. Instead, our findings suggest that impulse control may be a key

Table 7. Correlations Between All Analysis Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Gratification delay (54 months)																									
1. Continuous	—																								
2. <0.333 min	–.69	—																							
3. 0.333–2 min	–.47	–.18	—																						
4. 2–7 min	–.07	–.19	–.16	—																					
5. 7 min	.90	–.51	–.43	–.45	—																				
Related measures																									
6. Self–control (54 months)	.24	–.15	–.15	–.03	.24	—																			
7. Attention (54 months)	.22	–.18	–.07	–.08	.24	.15	—																		
8. Impulsivity (54 months)	–.30	.26	.06	.05	–.28	–.28	–.26	—																	
Outcome measures																									
9. Achievement (Grade 1)	.31	–.26	–.08	–.03	.28	.33	.30	–.27	—																
10. Achievement (age 15)	.30	–.25	–.09	–.02	.27	.32	.20	–.23	.64	—															
11. Behavior (Grade 1)	–.08	.06	.05	–.02	–.07	–.30	–.08	.05	–.09	–.11	—														
12. Behavior (age 15)	–.06	.08	.01	–.04	–.04	–.23	–.06	.06	–.11	–.13	.55	—													
Background controls																									
13. Male	–.05	.06	–.02	.02	–.05	–.20	–.01	.23	–.01	.05	–.00	–.04	—												
14. Black	–.25	.21	.07	.05	–.24	–.16	–.12	.20	–.29	–.33	.06	.00	–.00	—											
15. Hispanic	–.03	–.00	.06	–.02	–.02	–.04	–.02	.03	–.05	–.03	.01	.04	.03	–.08	—										
16. Other	–.04	.00	.03	.04	–.05	–.00	.02	–.02	.02	.01	–.01	–.01	–.03	–.07	–.05	—									
17. Age	.03	–.04	.03	–.02	.02	.03	.06	–.02	.04	–.05	.03	.04	–.00	.03	.01	–.04	—								
18. Log of income	.30	–.26	–.08	–.03	.27	.26	.19	–.19	.37	.40	–.16	–.17	–.02	–.36	–.08	–.01	–.01	—							
19. Mother's age	.20	–.18	–.05	–.00	.18	.18	.12	–.14	.22	.32	–.18	–.21	–.04	–.28	–.10	–.04	–.04	.54	—						
20. Mother's education (years)	.25	–.19	–.09	–.04	.24	.27	.16	–.20	.35	.42	–.13	–.17	–.04	–.22	–.11	–.03	–.00	.61	.52	—					
21. Mother's PPVT score	.28	–.22	–.09	–.08	.28	.29	.12	–.18	.35	.48	–.10	–.07	–.01	–.37	–.11	–.09	–.03	.49	.46	.57	—				
22. Site 1	–.04	.02	.00	.06	–.06	–.06	.06	–.02	.03	–.14	.09	.07	–.00	.11	–.07	–.07	–.03	–.09	–.09	–.06	–.10	—			
23. Site 2	.00	–.06	.05	.01	.00	.04	.03	–.03	.06	.10	–.06	–.07	.00	–.11	.23	–.01	–.18	.16	.06	.02	.04	–.10	—		
24. Site 3	.07	–.05	–.03	–.02	.07	–.04	.02	–.09	–.04	–.08	.04	.02	–.02	–.02	.08	–.03	.10	–.07	–.05	.02	–.02	–.10	–.11	—	
25. Site 4	–.00	.02	–.01	–.01	–.00	.02	.04	.09	–.02	.05	–.07	–.04	.02	–.05	–.04	.04	.00	.05	.07	–.04	–.02	–.10	–.11	–.11	—
26. Site 5	–.06	.02	.06	–.00	–.06	.02	.03	.01	–.02	–.05	–.02	–.02	.02	.13	–.06	–.02	.22	–.05	.03	.02	–.01	–.11	–.11	–.11	–.11
27. Site 6	.03	–.01	–.04	–.01	.04	.06	.04	–.03	.04	.09	–.07	–.08	–.02	.11	–.06	–.01	–.10	.12	.09	.13	.10	–.10	–.10	–.11	–.10
28. Site 7	–.05	.04	.00	.01	–.04	–.02	–.10	.12	–.05	.02	–.02	–.03	.02	.01	.00	.02	.14	–.09	–.07	–.08	–.01	–.11	–.11	–.11	–.11
29. Site 8	.06	.00	–.05	–.09	.09	.10	–.00	–.08	–.01	.05	–.02	.05	–.00	–.05	.00	.11	–.19	.08	.10	.10	.14	–.11	–.11	–.11	–.11
30. Site 9	–.04	–.00	.04	.04	–.06	–.07	–.01	.00	.05	.02	.03	.03	.01	–.05	–.06	–.05	.04	–.06	–.08	–.12	–.11	–.11	–.11	–.11	–.11
31. Birth weight (g)	–.01	.02	.01	–.06	.02	–.02	.05	–.01	.11	.10	.02	.09	.12	–.14	.04	–.07	.01	.04	.05	.07	.13	–.02	.03	–.06	.04
32. BBCS	.28	–.22	–.10	–.04	.26	.32	.26	–.29	.54	.50	–.09	–.10	–.15	–.32	–.07	–.02	.01	.45	.32	.42	.40	–.05	.09	.00	.04
33. Bayley MDI	.34	–.27	–.08	–.06	.31	.29	.24	–.24	.42	.39	–.08	–.13	–.17	–.32	–.08	–.01	–.02	.40	.23	.36	.34	–.05	.10	–.02	.11
34. Temperament	–.08	.11	.00	–.02	–.06	–.14	–.04	.08	–.11	–.12	.12	.15	–.04	.17	–.01	.05	–.01	–.19	–.19	–.13	–.19	.06	–.08	–.01	–.01
HOME controls																									
35. Learning Materials	.29	–.23	–.11	–.02	.27	.31	.15	–.23	.38	.40	–.10	–.11	–.05	–.39	–.12	–.08	.03	.49	.35	.48	.47	–.06	–.09	–.01	–.04
36. Language Stimulation	.21	–.18	–.05	–.04	.20	.17	.08	–.14	.25	.21	–.01	–.06	–.03	–.11	–.12	–.11	.01	.27	.12	.24	.24	.06	–.16	–.15	–.20
37. Physical Environment	.20	–.13	–.13	.02	.17	.15	.13	–.12	.23	.21	–.09	–.08	.01	–.24	.00	.01	–.03	.28	.18	.23	.19	–.07	–.06	–.05	.03
38. Responsivity	.19	–.13	–.08	–.05	.20	.18	.14	–.12	.19	.17	–.09	–.07	–.02	–.22	–.06	–.07	–.09	.32	.26	.28	.27	–.11	–.01	–.06	–.02
39. Academic Stimulation	.21	–.17	–.06	–.01	.18	.15	.05	–.15	.23	.20	.00	–.01	–.03	–.17	–.09	–.04	.01	.24	.12	.25	.24	–.01	–.18	–.06	–.07
40. Modeling	.17	–.11	–.06	–.05	.16	.17	.10	–.07	.23	.25	–.07	–.04	–.05	–.15	–.06	–.06	–.05	.31	.23	.33	.29	–.00	.01	–.09	–.08
41. Variety	.25	–.15	–.14	–.04	.24	.22	.12	–.21	.28	.29	–.10	–.07	–.03	–.27	–.09	–.05	.01	.41	.27	.39	.37	.02	–.14	–.06	–.04
42. Acceptance	.12	–.07	–.07	–.04	.13	.21	.13	–.17	.16	.19	–.16	–.14	–.05	–.10	–.01	.00	–.04	.23	.21	.24	.20	–.10	–.00	–.06	.01
43. Responsivity-Empirical Scale	.20	–.14	–.08	–.05	.20	.16	.12	–.10	.20	.16	–.06	–.05	–.02	–.18	–.06	–.04	–.06	.31	.20	.26	.25	.04	–.03	–.07	–.03
54-month controls																									
44. Letter-Word ID (WJ-R)	.28	–.22	–.09	–.03	.25	.29	.25	–.24	.60	.49	–.07	–.07	–.10	–.20	–.08	.05	–.01	.38	.19	.38	.34	–.02	.01	–.01	.03
45. Applied Problems (WJ-R)	.37	–.28	–.16	–.01	.33	.35	.32	–.32	.62	.56	–.04	–.12	–.10	–.32	–.07	.01	–.02	.42	.28	.40	.43	–.08	.03	.01	.07
46. Picture Vocabulary (WJ-R)	.28	–.21	–.08	–.09	.28	.25	.22	–.18	.42	.50	–.09	–.04	.10	–.33	–.10	–.01	.02	.42	.32	.40	.48	–.12	.01	.01	.07
47. Memory for Sentences (WJ-R)	.29	–.25	–.09	–.02	.26	.28	.22	–.21	.42	.43	–.11	–.09	–.04	–.18	–.11	.04	.02	.29	.21	.28	.30	–.08	–.06	–.06	.07
48. Incomplete Words (WJ-R)	.23	–.17	–.08	–.06	.22	.15	.19	–.17	.39	.34	–.05	–.12	–.00	–.18	–.10	.01	.03	.24	.18	.24	.27	–.07	–.06	–.04	.02
49. Internalizing (CBCL)	–.04	.04	.02	–.00	–.04	–.17	–.05	.08	–.07	–.08	.53	.38	.03	.04	–.01	.03	.09	–.06	–.09	–.10	–.10	.01	–.04	–.02	.03
50. Externalizing (CBCL)	–.10	.07	.07	–.02	–.09	–.39	–.07	.09	–.10	–.12	.63	.47	–.08	.05	.01	–.00	.01	–.12	–.13	–.13	–.11	.04	.01	.03	–.03

(*continued*)

Table 7. (Continued)

Variable	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
Gratification delay (54 months)																								
1. Continuous																								
2. <0.333 min																								
3. 0.333–2 min																								
4. 2–7 min																								
5. 7 min																								
Related measures																								
6. Self–control (54 months)																								
7. Attention (54 months)																								
8. Impulsivity (54 months)																								
Outcome measures																								
9. Achievement (Grade 1)																								
10. Achievement (age 15)																								
11. Behavior (Grade 1)																								
12. Behavior (age 15)																								
Demographic controls																								
13. Male																								
14. Black																								
15. Hispanic																								
16. Other																								
17. Age																								
18. Log of income																								
19. Mother's age																								
20. Mother's education (years)																								
21. Mother's PPVT score																								
22. Site 1																								
23. Site 2																								
24. Site 3																								
25. Site 4																								
26. Site 5	—																							
27. Site 6	−.11	—																						
28. Site 7	−.12	−.11	—																					
29. Site 8	−.12	−.11	−.12	—																				
30. Site 9	−.12	−.11	−.12	−.12	—																			
31. Birth weight (g)	−.02	−.03	−.01	.03	.02	—																		
32. BBCS	.00	.05	−.14	.05	−.03	.08	—																	
33. Bayley MDI	.02	.02	−.18	−.06	−.01	.06	.52	—																
34. Temperament	−.04	.02	.02	.02	.03	−.04	−.15	−.12	—															
HOME controls																								
35. Learning Materials	−.01	.08	−.04	.00	.05	.07	.47	.43	−.12	—														
36. Language Stimulation	.02	.18	.01	.05	.09	.10	.28	.23	−.04	.51	—													
37. Physical Environment	.09	.02	−.25	−.05	.19	.01	.25	.24	−.08	.41	.28	—												
38. Responsivity	−.11	.30	−.30	.12	.06	.08	.31	.25	−.11	.38	.38	.26	—											
39. Academic Stimulation	.04	.12	−.11	.05	.09	.08	.33	.26	−.02	.55	.55	.31	.33	—										
40. Modeling	−.00	.15	−.08	.03	−.06	.07	.24	.24	−.10	.37	.31	.26	.28	.28	—									
41. Variety	−.03	.16	−.09	.07	.08	.04	.36	.37	−.09	.56	.41	.33	.33	.43	.35	—								
42. Acceptance	−.04	.14	.04	.04	−.14	.05	.22	.19	−.05	.28	.20	.17	.19	.14	.32	.23	—							
43. Responsivity-Empirical Scale	−.15	.17	−.12	.08	.02	.04	.24	.19	−.12	.35	.48	.26	.77	.29	.27	.28	.23	—						
54-month controls																								
44. Letter-Word ID (WJ-R)	−.01	.10	−.08	.03	−.05	.07	.61	.40	−.08	.40	.29	.23	.26	.34	.23	.32	.19	.21	—					
45. Applied Problems (WJ-R)	−.02	.09	−.08	.04	−.03	.09	.57	.56	−.13	.43	.24	.27	.25	.25	.18	.31	.22	.21	.58	—				
46. Picture Vocabulary (WJ-R)	−.01	.05	−.04	.10	−.03	.12	.46	.44	−.12	.43	.25	.23	.27	.28	.21	.38	.16	.22	.46	.52	—			
47. Memory for Sentences (WJ-R)	.08	.06	−.05	.04	.03	.08	.39	.43	−.08	.31	.20	.19	.17	.22	.14	.30	.15	.13	.42	.47	.46	—		
48. Incomplete Words (WJ-R)	.05	.04	.00	−.05	.13	.09	.30	.36	−.10	.30	.24	.23	.16	.22	.14	.27	.11	.16	.36	.45	.37	.49	—	
49. Internalizing (CBCL)	.05	−.05	.01	−.02	−.03	.04	−.03	−.03	.14	−.07	−.04	−.02	−.04	.02	−.04	−.07	−.07	−.06	−.01	−.04	−.08	−.06	−.04	—
50. Externalizing (CBCL)	.01	−.03	−.01	−.06	.01	.04	−.11	−.05	.12	−.12	−.06	−.09	−.10	−.07	−.11	−.11	−.14	−.11	−.04	−.05	−.09	−.10	−.05	.58

Note: n=918. All nonmissing cases for each pairwise correlation were included. The Supplemental Material presents correlations for all variables shown separately by mother's education. BBCS = Bracken Basic Concept Scale; CBCL = Child Behavior Checklist; HOME = Home Observation for Measurement of the Environment; MDI = Mental Development Index; PPVT = Peabody Picture Vocabulary Test; WJ-R = Woodcock-Johnson Psycho-Educational Battery Revised.

mechanism, although post hoc inclusion of an explicit measure of impulse control explained some but certainly not most of the delay-of-gratification effect.

These results create further questions regarding what the marshmallow test might measure and how it relates to the umbrella construct of self-control. We observed that delay of gratification was strongly correlated with concurrent measures of cognitive ability, and controlling for a composite measure of self-control explained only about 25% of our reported effects on achievement. These results suggest that the marshmallow test may capture something rather distinct from self-control. Indeed, Duckworth and colleagues (2013) also investigated the relations among delay of gratification, self-control, and intelligence using the data employed here, and they found that both self-control and intelligence mediated the relation between early delay ability and later outcomes. Our results further suggest that simply viewing delay of gratification as a component of self-control may oversimplify how it operates in young children.

When considering how our results might inform intervention development, recall that models with controls for concurrent measures of cognitive skills and behavior reduced the association between delay of gratification and age-15 achievement to nearly zero. This implies that an intervention that altered a child's ability to delay but failed to change more general cognitive and behavioral capacities would likely have limited effects on later outcomes. If intervention developers hope to generate program impacts that replicate the long-term marshmallow test findings, targeting the broader cognitive and behavioral abilities related to delay of gratification might prove more fruitful.

Indeed, Mischel and Shoda's original results (Shoda et al., 1990) supported similar conclusions. Recall that they reported long-run correlations between delay of gratification and later outcomes only for children who were not provided with strategies for delaying longer. That the prediction was strong only in trials that relied on natural variation in children's ability to delay suggests that unobserved factors underlying children's delay ability may have driven the long-run correlations. Our results support this interpretation.

Our study is not without weaknesses. The 7-min ceiling was limiting, although our nonlinear models indicated that it was unlikely to affect conclusions drawn for the lower-SES sample. For the higher-SES sample, the 7-min ceiling prevented a direct replication of Mischel and Shoda's original work (e.g., Shoda et al., 1990), as a substantial majority of higher-SES children hit the ceiling. The lack of precision in our higher-SES results was unfortunate, though it should be noted that point estimates in fully controlled models were often very small. At the very least, these results further suggest that bivariate associations between delay of gratification and later outcomes probably contain substantial bias, even for more privileged children.

It should also be noted that variation in our delay-of-gratification measure at age 54 months was not exogenous, so our models could not truly capture the effects that would be produced by exogenously spurred gains in early delay-of-gratification ability. However, our models included an extensive set of control variables that go well beyond the bivariate specifications employed in previous studies (e.g., Shoda et al., 1990). Finally, data not drawn to be nationally representative provide a shaky foundation for generalization.

In sum, our findings suggest that although early delay of gratification did indeed correlate with later achievement for children whose mothers had not completed college, the magnitude of this association was highly sensitive to the inclusion of control variables and did not appear to be linear across the delay-of-gratification distribution. Future work on delay of gratification should continue to examine the processes captured by the marshmallow test and whether early delay-of-gratification interventions would be worthwhile investments for promoting children's long-run success.

Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

T. W. Watts and G. J. Duncan developed the study concept and design and wrote the manuscript. T. W. Watts and H. Quan analyzed the data. All authors approved the final manuscript.

Acknowledgments

We are grateful to Ana Auger, Drew Bailey, Daniel Belsky, Jay Belsky, Clancy Blair, Peg Burchinal, Angela Duckworth, Dorothy Duncan, Jade Jenkins, Terrie Moffitt, Cybele Raver, and Deborah Vandell for helpful comments on drafts of this manuscript.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number P01-HD065704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618761661

Open Practices

The primary data set used in this study was from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development. Our data-use agreement prevented us from posting these data online, but the data set is available on request from the ICPSR website (https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00233). The secondary data set was from the Early Childhood Longitudinal Program and can be found at the National Center for Education Statistics website (https://nces.ed.gov/ecls/dataproducts.asp). The three Stata files necessary to replicate the results given here in the tables, along with the complete Open Practices Disclosure for this article, can be found in the Supplemental Material (http://journals.sagepub.com/doi/suppl/10.1177/0956797618761661).

The design and analysis plans for the study were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 Profile*. Burlington: Department of Psychiatry, University of Vermont.

Barkley, R. A. (1994). The assessment of attention in children. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 69–102). Baltimore, MD: Brookes.

Bayley, N. (1991). *Bayley scales of infant development* (2nd ed.). New York, NY: Psychological Corp.

Bembenutty, H., & Karabenick, S. A. (2004). Inherent association between academic delay of gratification, future time perspective, and self-regulated learning. *Educational Psychology Review*, *16*, 35–57.

Bracken, B. A. (1984). *Bracken Basic Concept Scale*. Chicago, IL: Psychological Corp.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas at Little Rock.

Campbell, D. (1986). Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In D. Fiske & R. Shweder (Eds.), *Metatheory in social science* (pp. 108–135). Chicago, IL: University of Chicago Press.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., . . . & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences, USA*, *108*, 14998–15003. doi:10.1073/pnas.1108561108

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

Duckworth, A. L., Tsukayama, E., & Kirby, T. A. (2013). Is it really self-control? Examining the predictive power of the delay of gratification task. *Personality and Social Psychology Bulletin*, *39*, 843–855.

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*, 2417.

Flook, L., Goldberg, S. B., Pinger, L., & Davidson, R. J. (2015). Promoting prosocial behavior and self-regulatory skills in preschool children through a mindfulness-based kindness curriculum. *Developmental Psychology*, *51*, 44–51.

Imuta, K., Hayne, H., & Scarf, D. (2014). I want it all and I want it now: Delay of gratification in preschool children. *Developmental Psychobiology*, *56*, 1541–1552.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, *126*, 109–114.

Kumst, S., & Scarf, D. (2015). Your wish is my command! The influence of symbolic modelling on preschool children's delay of gratification. *PeerJ*, *3*, Article e774. doi:10.7717/peerj.774

Medoff-Cooper, B., Carey, W. B., & McDevitt, S. C. (1993). The Early Infancy Temperament Questionnaire. *Journal of Developmental & Behavioral Pediatrics*, *14*, 230–235.

Michaelson, L. E., & Munakata, Y. (2016). Trust matters: Seeing how an adult treats another person influences preschoolers' willingness to delay gratification. *Developmental Science*, *19*, 1011–1019.

Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 249–292). New York, NY: Academic Press.

Mischel, W. (2014). *The marshmallow test: Why self-control is the engine of success*. New York, NY: Little, Brown.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., & Shoda, Y. (2010). 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*, 252–256.

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, *54*, 687–696.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*, 933–938.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, USA*, *108*, 2693–2698.

Murray, J., Theakston, A., & Wells, A. (2016). Can the attention training technique turn one marshmallow into two? Improving children's ability to delay gratification. *Behaviour Research and Therapy*, *77*, 34–39.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*, 133–164.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, *53*, 202–210. doi:10.1016/j.intell.2015.10.006

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behaviour: Replications from longitudinal studies. *Psychological Medicine*, *8*, 611–622.

Rodriguez, M. L., Mischel, W., & Shoda, Y. (1989). Cognitive person variables in the delay of gratification of older children at risk. *Journal of Personality and Social Psychology*, *57*, 358–367.

Romer, D., Duckworth, A. L., Sznitman, S., & Park, S. (2010). Can adolescents learn self-control? Delay of gratification in the development of control over risk taking. *Prevention Science*, *11*, 319–330.

Rueda, M. R., Checa, P., & Cómbita, L. M. (2012). Enhanced efficiency of the executive attention network after training in preschool children: Immediate changes and effects after two months. *Developmental Cognitive Neuroscience*, *2*, S192–S204. doi:10.1016/j.dcn.2011.09.004

Rybanska, V., McKay, R., Jong, J., & Whitehouse, H. (2017). Rituals improve children's ability to delay gratification. *Child Development*, *89*, 349–359. doi:10.1111/cdev.12762

Shimoni, E., Asbe, M., Eyal, T., & Berger, A. (2016). Too proud to regulate: The differential effect of pride versus joy on children's ability to delay gratification. *Journal of Experimental Child Psychology*, *141*, 275–282.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*, 978–986.

StataCorp. (2017). Stata Statistical Software: Version 15.0 [Computer software]. College Station, TX: StataCorp LLC.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*, 352–360.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. Itasca, IL: Riverside.

Online supplementary materials for

"Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Gratification Delay and Later Outcomes"

Method

Data

Data for the current study were drawn from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD). Participants were recruited from ten U.S. sites in both urban and rural settings. Mothers who had recently given birth were recruited from nearby hospitals, and mothers were excluded if they could not speak English or if they planned to move within the next 18 months. The first wave of major child-level data collection occurred when the focus child was 1 month of age, at which point 1,364 children remained in the study (roughly 50% of the originally recruited sample). Most of the attrition between initial recruitment and the 1-month interview was concentrated among low-SES mothers and children (see Duncan & Gibson, 2000; NICHD Early Child Care Research Network, 2002). Data collection continued at various times throughout children's lives, with the last full round of measurement occurring when children were 15 years old. The current study relies primarily on data collected when children were 54 month old, in grade 1, and 15 years old. However, for models with covariates, we also rely on data collected between the 1 month child interview and the age 36 month interview.

ECLS-K. For purposes of comparison, we also show demographic characteristics for children and families recruited for participation in the Early Childhood Longitudinal Study-Kindergarten Cohort of 1998-1999 (ECLS-K; Tourangeau, Nord, Le, Sorongon, & Najarian, 2009). The data were collected by the National Center for Educational Statistics (NCES), and the sample was designed to be nationally representative of children in kindergarten during the fall of 1998. This publically-available dataset has been broadly used to study child development and education, and information regarding data collection procedures and study measures has been widely reported.

In Table 1 of the main text, we present information taken from study families during the fall 1998 parent survey. Mothers reported their age at the time of the survey, and we subtracted the focus child's age from the reported mother's age to calculate "Mother's Age at Child Birth." NCES presented family annual income as a categorical variable, with respondents falling into income ranges (e.g., $20,000-$25,000). We then gave each participant the middle income value for each category, and used the number of children and adults in the home to calculate the "income to needs ratio" following the guidelines given by the Census Bureau (https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html). Finally, for all descriptive statistics presented, we weighted estimates using the C1C1W0 weight to recover nationally representative estimates.

Measures

Academic achievement. The Woodcock Johnson- Revised (WJ-R; Woodcock, McGrew & Mather, 2001) is a commonly used assessment of cognitive and academic ability, and it contains a number of subtests each designed to measure mathematics and verbal achievement, as well as more general cognitive abilities such as executive function. Each subtest was age-normed and administered by a trained examiner in a one-on-one interview with the child. We focus on WJ-R subtests designed to measure mathematics and reading achievement, which were administered at grade 1 and age 15 years. In some models, we also use WJ-R subtests given at age 54-months as control variables. For all WJ-R subtests, we used the WJ-R standard scores, which were standardized to have a mean of 100 and standard deviation of 15 at each age.

Mathematics achievement. At age 54-months, grade 1, and age 15-years, mathematics

achievement was measured using the WJ-R *Applied Problems* subtest, a commonly used measure of math ability (e.g., Siegler et al., 2012; Watts et al., 2014) that tested mathematical reasoning and problem solving. The examiner administering the test established a basal and ceiling for each child, and items were ordered hierarchically. At age 54 months, the *Applied Problems* items focused on counting and simple arithmetic; by age 15, items focused on algebraic concepts and fractions.

Reading achievement. At age 54 months and grade 1, reading achievement was measured using the *Letter-Word Identification* task, which tested children's ability to sound out simple words. Like the *Applied Problems* test, items on the *Letter-Word Identification* test were also ordered hierarchically, and the examiner established a basal and a ceiling for each child. Early items on the test asked children to match a pictographic representation of a word with an actual picture of the object, and the remaining items asked subjects to read aloud isolated letters and words. At age 15, reading achievement was measured using the *Passage Comprehension* subtest. Early items on this test involved reading a phrase then identifying a picture that depicted the phrase. Later items asked children to read a passage and fill in a missing key word.

Additional 54-month cognitive measures. In models with 54-month covariates, we also used the *Memory for Sentences* and *Incomplete Words* subtests as measures of cognitive ability. The *Incomplete Words* test measured auditory closure and processing, and children listened to an audio recording where words missing a phenome were listed off. They were then asked to name the complete word. Finally, the *Picture Vocabulary* test was a measure of verbal comprehension and crystallized intelligence. In this task, children were asked to name pictured objects.

Additional problem behaviors. In addition to using the mother-reported Child Behavior Checklist, we investigated the relationship between delay of gratification and other measures of behavioral functioning assessed at age 15. These measures included the Stoplight Task (Gardner & Steinberg, 2005), a measure of risk-taking in which children played a computer game that asked them to control a car attempting to reach a destination in a limited amount of time. While driving the car, children encountered stoplights that forced them to face the choice of either slowing down and losing time or running the light and possibly crashing the vehicle. From this task, we used two measures of risk taking: 1) the amount of time (in milliseconds) between the appearance of a yellow light and the application of the brake; 2) number of brake applications during the entire task. The stoplight task has been used in other developmental studies as a measure of adolescent risk taking (see Chein et al., 2011; Kim-Spoon et al., 2016)

As an alternative measure of internalizing and externalizing, we included scales taken from the age-15 Youth Self Report (YSR). The YSR was adapted from the CBCL, and it included 119 items that allowed youth to assess their own behavior. We also included a self-reported measure of impulse control, taken from the Winberger Adjustment Inventory (Weinberger & Schwarts, 1990), in which youth responded to 8 items designed to help them assess their own ability to control counterproductive impulses. Finally, we used a youth-reported measure of risky behavior called the Risky Behavior Questionnaire, which was adapted from several different measures used in large studies of child development (Conger & Elder, 1994; Halpern-Felsher, 2005). With this measure, youth responded to 61 items asking them how many times in the past year they had engaged in 55 different risk behaviors (e.g., alcohol use and sexual risk taking).

Continuous Performance Task. In supplemental models described below, we controlled for a 54-month measure of attention and impulsivity called the Continuous Performance Task (CPT; Barkley, 1994). With the CPT, children interacted with a computer

game in which they were asked to click a key every time a certain object appeared on the screen. Attention was measured by the proportion of trials in which the child correctly clicked the key in response to a target object. Impulsivity was measured by the proportion of trials in which a child incorrectly clicked a key in response to a non-target object. These controls were introduced to gauge the extent to which the age-54 month gratification delay "effect" is reduced when differences in concurrent impulsivity are taken into account.

Self-control. We also tested whether controlling for a measure of self-control would reduce the effect of gratification delay on later achievement and behavior. Following the example of Duckworth and colleagues (2013), we created a composite score of self-control from mother and teacher reports at age 54 months. Both mothers and caregivers completed the Child Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994), and both surveys contained sub-scales that measured "attention focusing" and "inhibitory control." As with the Duckworth et al. (2013) study, we averaged together the teacher and mother "attention focusing" and "inhibitory" control scales to create a self-control composite (see Duckworth et al. (2013) for a description of the measure's psychometric properties).

Analysis

Our primary goal is to estimate the association between early gratification delay and long-run measures of academic achievement and behavioral functioning. Like the work of Mischel and Shoda (e.g., Mischel, Shoda, Peake, 1988; Mischel, Shoda, Rodriguez, 1989; Shoda, Mischel, Peake, 1990) our data did not include a measure of gratification delay in which cross-child differences were generated from some exogenous intervention, so we do not pretend that the associations we estimate reflect causal impacts. Instead, the goal of our investigation is to estimate how much bias might be contained in longitudinal correlations between measures of delay of gratification and measures of child cognitive and behavioral functioning as a result of failure to control for characteristics of children and their environments, all measured before age 54 months, that may be causing both differences in 54-month gratification delay times and adolescent outcomes of interest.

To accomplish our analytic goals, we modeled later academic achievement and behavior as a function of an age-54-month measure of gratification delay and pre-54-month controls:

$$(1)\quad Y_{iLATE} = \alpha_0 + \beta_1 DOG_{i54} + \emptyset Back\&Fam_{iEARLY} + \delta Child_{iEARLY} + \lambda Cog_{iEARLY} + \gamma Home_{iEARLY} + \varepsilon_i$$

where Y_{iLATE} is an outcome measure of either academic achievement or behavioral functioning taken at a later time point (both grade 1 and age 15), DOG_{i54} is the 54-month measure of delay of gratification ("DOG") measured in minutes waited, $Back\&Fam_{iEARLY}$ is a vector of family and child demographic characteristics (e.g., family income, gender, race) assessed prior to the 54-month survey, $Child_{iEARLY}$ is a vector of early measures of the child's personal characteristics (e.g., temperament, birth weight), Cog_{iEARLY} is a vector of early measures of cognitive functioning, and $Home_{iEARLY}$ is a set of measures of the home environment captured by the 36-month HOME battery. Finally, the error term, ε_i, will be uncorrelated with our key estimate, β_1, only if our control variables perfectly capture all of the possible sources of confounding variation.

In addition, we tested models that added controls assessed at age 54 months in order to project how changes in delay of gratification ability, holding constant other concurrent cognitive and behavioral abilities, might affect later development:

(2) $Y_{iLATE} = \alpha_0 + \beta_1 DOG_{i54} + \emptyset Back\&Fam_{iEARLY} + \delta Child_{iEARLY} + \lambda Cog_{iEARLY} + \gamma Home_{iEARLY} + \chi Cog\&Beh_{i54} + \varepsilon_i$

where all model parameters are defined as with Equation 1, but a vector of cognitive and behavioral measures assessed at age 54 months, $Cog\&Beh_{i54}$, is added. Although it might be argued that the estimates from Equation 2 provide the best projections for how a delay of gratification intervention might affect later functioning, we recognize that the concurrent timing of their measurement with the gratification delay task may risk over-controlling for capacities that are themselves a product of past emotional self-regulation.

All continuous variables were standardized so that coefficients can be likened to effect sizes, and all models with control variables included a set of dummy variables for each site to adjust for any between-site differences. Finally, in order to account for missing data on control variables, we used SEM with Full Information Maximum Likelihood in Stata 13.0 to estimate all analytic models.

Additional Results

Additional correlational results. In Table S1, we a correlation matrix among all analysis variables for the children of mothers who did not graduate college (n = 551), and in Table S2, we present the same matrix for the children of college completing mothers (n= 365).

Alternative Models

Because models that relied on the children of college completing mothers yielded unreliable results, we focus only on the children of mothers without college degrees (i.e., the lower-SES sample) for supplemental models shown in Tables S3 through S8.

Continuous performance task. Table S3 presents results from models that added controls for age 54-month attention and impulsivity, as measured by the continuous performance task (CPT). These models also included controls measures prior to age 54 months (compare with models 2, 5, 8 and 11 from Table 3). We found that when CPT measures were added, coefficients for gratification delay in achievement models dropped slightly (approximately 21-27%). Thus, we did not find that a direct measure of impulse control completely explained the effect for gratification delay. Further, we only found the CPT measures to be predictive of later outcomes in the grade 1 achievement model.

Self-control. Table S4 also presents results from models with controls measured prior to age 54-months with our composite score of mother and teacher reported self-control also added. Similar to the CPT results, we found only partial mediation. However, our self-control measure significantly predicted achievement and behavior at both grades 1 and age 15. These results again indicate that the gratification delay measure may tap into processes distinct from self-control. For a more complete investigation of this issue, see Duckworth and colleagues (2013), as they had a more robust examination of this hypothesis using the same dataset and the same measure of self-control.

Additional behavioral outcomes. In Table S5, we present descriptive results for additional behavioral outcomes assessed at age 15.

Table S6 presents associations between gratification delay and measures of risk taking; two measures came from the stoplight task and one from the youth-reported Risky Behavior Questionnaire. Aside from the bivariate association between waiting for 7 minutes and youth-reported risk taking (β = -0.38, SE = 0.12, p = .002), we found no significant associations between gratification delay and measures of risk taking. In Table S7, we display associations

between youth-reported measures of behavior problems and delay of gratification and we again found no significant results.

Disaggregated measures of achievement and behavior problems. In the main text, we averaged together measures of math and reading achievement, and measures of externalizing and internalizing, to create composite measures of achievement and behavior, respectively. In Table S8, we present disaggregated results in which we regressed individual measures of age-15 math, reading, externalizing, and internalizing on gratification delay. Results closely mirrored the results shown in Table 4 of the main text. In models with controls measured prior to age 54-months, we found associations between the gratification delay dummy variables and measures of math and reading achievement that ranged between 0.19 and 0.29.

Compared with the composite score of behavioral problems, we found slightly larger associations between gratification delay and externalizing behavior, but no associations were statistically significant in models that contained controls. We found no associations between gratification delay and internalizing behavior.

References

Chein, J., Albert, D., O'Brien, L., Uckert, K., & Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science, 14*(2).

Conger, R.D. & Elder, G.H. (1994). *Families in troubled times: Adapting to change in rural America*. New York: Aldine de Gruyter.

Duncan, G.J., & Gibson, C. (2000). Selection and attrition in the NICHD childcare study's analyses of the impacts of childcare quality on child outcomes. *Unpublished paper. Evanston, IL: Northwestern University*. Retrieved from: http://www.gse.uci.edu/person/duncan_g/docs/Duncan_Gibson_NICHD_childcare_quality_review_2000.pdf

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology, 41*(4), 625.

Halpern-Felsher, B. L., Cornell, J. L., Kropp, R. Y., & Tschann, J. M. (2005). Oral versus vaginal sex among adolescents: Perceptions, attitudes, and behavior. *Pediatrics, 115*(4), 845-851.

Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., & Kellam, S. G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review, 2*, 109-145.

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*(4), 687.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*(4907), 933-938.

NICHD Early Child Care Research Network. (2002). Early childcare and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 39*(1), 133-164.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. Merrill-Palmer Quarterly, 40, 21-39.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26*(6), 978.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., … Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691–697. doi:10.1177/0956797612440101

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Najarian, M., & Hausken, E. G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) combined user's manual for the ECLS-K Eighth-Grade and K-8 full sample data files and electronic codebook*. Washington, DC: U.S Department of Education.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352-360.3.

Weinberger, D.A., Schwartz, G.E., (1990). Distress and Restraint as Superordinate Dimensions of Self-Reported Adjustment: A Typological Perspective, Journal of Personality, 58, 381-417.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.

Table S1

Correlations Between All Analysis Variables for Children of Mothers Who Did Not Complete College

PANEL 1	1	2	3	4	5	6	7	8	9	10
Gratification Delay (54)										
1 Continuous	1.00									
2 <0.333 min.	-0.69	1.00								
3 0.333- 2 min.	-0.43	-0.24	1.00							
4 2 to 7 min.	0.01	-0.24	-0.19	1.00						
5 7 min.	0.89	-0.50	-0.39	-0.39	1.00					
Related Measures										
6 Self-control (54)	0.21	-0.12	-0.10	-0.03	0.20	1.00				
7 Attention (54)	0.22	-0.16	-0.08	-0.09	0.26	0.10	1.00			
8 Impulsivity (54)	-0.28	0.26	0.02	0.04	-0.26	-0.21	-0.22	1.00		
Outcome Measures										
9 Achievement (G1)	0.30	-0.26	-0.07	-0.01	0.27	0.30	0.29	-0.26	1.00	
10 Achievement (15)	0.29	-0.27	-0.04	0.02	0.24	0.24	0.18	-0.20	0.60	1.00
11 Behavior (G1)	-0.06	0.03	0.05	-0.01	-0.05	-0.33	-0.06	0.06	-0.09	-0.10
12 Behavior (15)	-0.06	0.07	-0.00	-0.02	-0.05	-0.22	-0.03	0.02	-0.08	-0.09
Demographic Controls										
13 Male	-0.06	0.07	-0.01	-0.01	-0.05	-0.22	-0.01	0.24	0.01	0.12
14 Black	-0.23	0.21	0.05	0.02	-0.23	-0.11	-0.10	0.19	-0.29	-0.33
15 Hispanic	-0.01	-0.02	0.05	-0.01	-0.01	-0.02	-0.01	0.01	-0.04	-0.00
16 Other	-0.01	0.01	-0.02	0.03	-0.02	-0.00	-0.01	-0.05	-0.01	-0.05
17 Age	0.04	-0.04	0.06	-0.07	0.05	0.03	0.02	-0.01	0.02	-0.00
18 Log of Income	0.25	-0.23	-0.04	-0.00	0.22	0.15	0.20	-0.14	0.35	0.32
19 Mother's Age	0.18	-0.17	-0.02	0.01	0.15	0.15	0.12	-0.13	0.21	0.26
20 Mother's Ed (yrs)	0.12	-0.11	-0.01	-0.01	0.11	0.13	0.11	-0.11	0.27	0.21
21 Mother PPVT	0.19	-0.20	0.01	-0.04	0.19	0.21	0.09	-0.14	0.34	0.38
22 Site 1	-0.06	0.03	0.01	0.08	-0.09	-0.06	0.08	-0.06	0.05	-0.16
23 Site 2	0.01	-0.09	0.05	0.03	0.01	0.06	0.05	-0.05	0.02	0.06
24 Site 3	0.09	-0.06	-0.04	-0.02	0.09	-0.05	-0.01	-0.10	-0.03	-0.05
25 Site 4	0.01	0.02	-0.00	-0.04	0.01	0.05	0.02	0.13	-0.01	0.08
26 Site 5	-0.05	0.04	0.05	-0.03	-0.05	0.05	0.00	0.04	-0.03	-0.03
27 Site 6	-0.04	0.06	-0.03	-0.05	-0.00	-0.04	0.01	-0.03	-0.03	0.00
28 Site 7	-0.05	0.02	0.02	0.02	-0.05	-0.01	-0.08	0.13	-0.11	-0.04
29 Site 8	0.02	-0.00	0.00	-0.07	0.05	0.10	-0.02	-0.07	0.01	0.00
30 Site 9	0.01	-0.03	0.01	0.04	-0.01	-0.05	0.02	-0.04	0.15	0.16
31 Birthweight (g's)	0.01	-0.00	0.00	-0.06	0.04	-0.04	0.07	-0.01	0.10	0.15
32 Bracken	0.25	-0.18	-0.10	-0.03	0.25	0.25	0.23	-0.26	0.54	0.47
33 Bayley	0.30	-0.23	-0.08	-0.05	0.28	0.24	0.24	-0.16	0.43	0.37
34 Temperament	-0.11	0.14	-0.04	-0.00	-0.09	-0.16	-0.05	0.11	-0.13	-0.16
H.O.M.E. Controls										
35 Learn. Mater.	0.28	-0.21	-0.10	-0.02	0.26	0.25	0.14	-0.18	0.38	0.36
36 Lang. Stim.	0.20	-0.14	-0.06	-0.05	0.20	0.12	0.09	-0.11	0.27	0.19
37 Phys. Env.	0.21	-0.12	-0.14	0.02	0.19	0.16	0.15	-0.12	0.29	0.24
38 Responsivity	0.16	-0.10	-0.08	-0.05	0.18	0.16	0.15	-0.10	0.25	0.17
39 Academ. Stim.	0.19	-0.15	-0.05	-0.03	0.18	0.12	0.08	-0.13	0.27	0.22
40 Modeling	0.14	-0.07	-0.08	-0.03	0.14	0.12	0.11	-0.05	0.23	0.23
41 Variety	0.22	-0.10	-0.14	-0.05	0.22	0.14	0.14	-0.19	0.29	0.27
42 Acceptance	0.09	-0.02	-0.09	-0.02	0.09	0.16	0.12	-0.14	0.13	0.18

43	Respons. Emp.	0.17	-0.10	-0.09	-0.04	0.17	0.13	0.11	-0.06	0.22	0.13
54-month Controls											
44	Letter Word (54)	0.25	-0.19	-0.08	-0.02	0.23	0.24	0.25	-0.22	0.56	0.44
45	App. Prob. (54)	0.34	-0.27	-0.11	0.02	0.30	0.29	0.31	-0.28	0.63	0.50
46	Pic. Vocab. (54)	0.24	-0.18	-0.03	-0.09	0.24	0.17	0.17	-0.13	0.37	0.46
47	Mem. Sent. (54)	0.26	-0.24	-0.06	0.03	0.23	0.24	0.21	-0.18	0.42	0.41
48	Inc. Words (54)	0.20	-0.16	-0.04	-0.04	0.20	0.10	0.22	-0.14	0.40	0.37
49	Internalizing (54)	-0.04	0.03	0.02	-0.00	-0.03	-0.17	-0.02	0.09	-0.08	-0.08
50	Externalizing (54)	-0.12	0.08	0.07	-0.02	-0.10	-0.39	-0.06	0.07	-0.10	-0.13

PANEL 2		11	12	13	14	15	16	17	18	19	20
11	Behavior (G1)	1.00									
12	Behavior (15)	0.54	1.00								
Demographic Controls											
13	Male	0.01	-0.04	1.00							
14	Black	0.06	-0.04	-0.01	1.00						
15	Hispanic	0.00	0.04	0.05	-0.12	1.00					
16	Other	-0.06	-0.01	-0.07	-0.09	-0.06	1.00				
17	Age	0.04	0.04	0.01	0.02	0.02	-0.02	1.00			
18	Log of Income	-0.15	-0.16	-0.01	-0.35	-0.04	-0.05	0.01	1.00		
19	Mother's Age	-0.24	-0.23	-0.05	-0.25	-0.06	-0.05	-0.03	0.48	1.00	
20	Mother's Ed (yrs)	-0.08	-0.10	-0.05	-0.07	-0.08	-0.08	0.03	0.45	0.36	1.00
21	Mother PPVT	-0.09	-0.02	-0.04	-0.39	-0.05	-0.05	-0.02	0.39	0.34	0.39
22	Site 1	0.09	0.08	-0.02	0.13	-0.09	-0.07	-0.07	-0.05	-0.07	0.02
23	Site 2	-0.06	-0.07	-0.01	-0.14	0.32	0.02	-0.18	0.20	0.08	0.07
24	Site 3	0.01	0.08	-0.07	-0.01	0.04	-0.01	0.09	-0.03	-0.06	0.01
25	Site 4	-0.10	-0.06	0.07	-0.09	-0.07	0.04	-0.02	0.06	0.11	-0.01
26	Site 5	-0.03	-0.06	0.00	0.19	-0.07	-0.01	0.22	-0.10	0.05	0.05
27	Site 6	0.02	-0.05	0.03	0.20	-0.04	-0.05	-0.11	0.02	0.04	0.05
28	Site 7	-0.02	-0.03	0.03	0.01	-0.02	0.04	0.20	-0.12	-0.10	-0.16
29	Site 8	-0.09	0.02	0.01	-0.04	0.03	0.07	-0.18	0.02	0.07	0.03
30	Site 9	0.01	0.00	-0.00	-0.10	-0.08	-0.06	0.04	0.04	-0.03	-0.07
31	Birthweight (g's)	0.05	0.08	0.11	-0.17	0.08	-0.10	0.06	0.04	0.06	0.08
32	Bracken	-0.09	-0.06	-0.14	-0.30	-0.00	-0.05	0.03	0.37	0.27	0.28
33	Bayley	-0.05	-0.11	-0.16	-0.31	-0.06	0.02	-0.01	0.35	0.19	0.24
34	Temperament	0.12	0.14	-0.02	0.19	-0.01	0.06	0.01	-0.22	-0.21	-0.07
H.O.M.E. Controls											
35	Learn. Mater.	-0.08	-0.08	-0.05	-0.37	-0.06	-0.09	0.05	0.45	0.28	0.38
36	Lang. Stim.	0.05	-0.05	-0.02	-0.09	-0.09	-0.13	0.04	0.27	0.07	0.20
37	Phys. Env.	-0.09	-0.08	-0.01	-0.24	0.01	-0.01	-0.04	0.31	0.19	0.19
38	Responsivity	-0.04	-0.02	-0.01	-0.22	-0.00	-0.07	-0.10	0.32	0.22	0.24
39	Academ. Stim.	0.03	-0.00	-0.01	-0.17	-0.07	-0.03	0.03	0.23	0.07	0.23
40	Modeling	-0.02	0.02	-0.04	-0.13	-0.03	-0.07	-0.03	0.26	0.19	0.29
41	Variety	-0.07	-0.03	-0.04	-0.26	-0.05	-0.05	0.01	0.38	0.21	0.27
42	Acceptance	-0.17	-0.15	-0.04	-0.07	0.02	-0.03	-0.07	0.16	0.18	0.15
43	Respons. Emp.	-0.03	-0.02	-0.03	-0.16	-0.03	-0.04	-0.06	0.31	0.15	0.24
54-month Controls											
44	Letter Word (54)	-0.03	-0.03	-0.07	-0.18	-0.06	0.04	-0.03	0.31	0.14	0.26
45	App. Prob. (54)	-0.02	-0.10	-0.10	-0.30	-0.06	0.01	-0.02	0.38	0.22	0.26
46	Pic. Vocab. (54)	-0.05	0.01	0.11	-0.34	-0.06	-0.02	0.03	0.34	0.28	0.26

47	Mem. Sent. (54)	-0.08	-0.05	-0.05	-0.16	-0.13	0.06	0.02	0.26	0.18	0.17
48	Inc. Words (54)	0.01	-0.10	0.00	-0.16	-0.09	0.05	0.04	0.23	0.15	0.14
49	Internalizing (54)	0.52	0.36	0.04	0.06	0.00	0.01	0.14	-0.07	-0.13	-0.11
50	Externalizing (54)	0.64	0.45	-0.06	0.05	0.01	-0.01	0.03	-0.11	-0.17	-0.13

PANEL 3		21	22	23	24	25	26	27	28	29	30
21	Mother PPVT	1.00									
22	Site 1	-0.11	1.00								
23	Site 2	0.12	-0.11	1.00							
24	Site 3	-0.01	-0.11	-0.10	1.00						
25	Site 4	0.00	-0.12	-0.11	-0.11	1.00					
26	Site 5	-0.04	-0.12	-0.11	-0.11	-0.12	1.00				
27	Site 6	-0.02	-0.09	-0.08	-0.08	-0.09	-0.09	1.00			
28	Site 7	-0.05	-0.12	-0.11	-0.11	-0.12	-0.12	-0.09	1.00		
29	Site 8	0.10	-0.10	-0.09	-0.09	-0.10	-0.10	-0.07	-0.10	1.00	
30	Site 9	-0.06	-0.13	-0.12	-0.12	-0.13	-0.13	-0.10	-0.14	-0.11	1.00
31	Birthweight (g's)	0.11	-0.05	0.01	-0.07	0.07	0.01	-0.00	0.00	-0.03	0.01
32	Bracken	0.36	-0.02	0.10	0.01	0.09	0.04	-0.06	-0.19	0.04	0.03
33	Bayley	0.30	-0.04	0.11	0.02	0.16	0.04	-0.10	-0.24	-0.09	0.06
34	Temperament	-0.19	0.07	-0.12	-0.02	-0.02	-0.03	0.06	0.04	0.00	0.03
H.O.M.E. Controls											
35	Learn. Mater.	0.42	-0.03	-0.06	-0.03	0.02	-0.05	0.00	-0.10	-0.02	0.12
36	Lang. Stim.	0.22	0.10	-0.18	-0.19	-0.18	-0.00	0.17	-0.02	0.07	0.13
37	Phys. Env.	0.23	-0.05	-0.06	-0.07	0.05	0.07	-0.03	-0.27	-0.08	0.22
38	Responsivity	0.25	-0.08	-0.01	-0.09	-0.01	-0.11	0.25	-0.32	0.18	0.11
39	Academ. Stim.	0.23	-0.01	-0.16	-0.07	-0.04	0.02	0.10	-0.14	0.05	0.12
40	Modeling	0.23	0.06	0.01	-0.09	-0.08	-0.02	0.09	-0.10	0.00	-0.02
41	Variety	0.29	0.03	-0.12	-0.06	0.01	-0.05	0.08	-0.11	0.05	0.15
42	Acceptance	0.15	-0.07	0.05	-0.07	0.03	-0.03	0.09	0.03	0.01	-0.13
43	Respons. Emp.	0.22	0.08	-0.04	-0.11	-0.01	-0.17	0.15	-0.14	0.14	0.05
54-month Controls											
44	Letter Word (54)	0.31	-0.00	0.00	0.03	0.08	0.01	0.01	-0.15	0.01	0.00
45	App. Prob. (54)	0.37	-0.05	0.03	0.02	0.12	-0.03	0.02	-0.11	-0.01	0.05
46	Pic. Vocab. (54)	0.40	-0.12	0.01	0.03	0.14	-0.02	-0.05	-0.07	0.04	0.02
47	Mem. Sent. (54)	0.23	-0.06	-0.07	-0.05	0.11	0.11	-0.02	-0.13	0.01	0.11
48	Inc. Words (54)	0.24	-0.08	-0.07	-0.06	0.07	0.06	-0.07	-0.04	-0.05	0.20
49	Internalizing (54)	-0.12	0.01	-0.03	0.00	0.06	0.00	0.03	0.02	-0.07	-0.07
50	Externalizing (54)	-0.10	0.04	0.01	0.04	-0.04	-0.01	0.01	0.00	-0.08	-0.00

PANEL 4		31	32	33	34	35	36	37	38	39	40
31	Birthweight (g's)	1.00									
32	Bracken	0.11	1.00								
33	Bayley	0.07	0.46	1.00							
34	Temperament	-0.05	-0.15	-0.10	1.00						
H.O.M.E. Controls											
35	Learn. Mater.	0.08	0.39	0.41	-0.13	1.00					
36	Lang. Stim.	0.07	0.24	0.19	-0.04	0.46	1.00				
37	Phys. Env.	0.04	0.26	0.27	-0.09	0.44	0.31	1.00			
38	Responsivity	0.09	0.27	0.21	-0.11	0.34	0.38	0.25	1.00		

39	Academ. Stim.	0.11	0.30	0.25	-0.02	0.55	0.55	0.32	0.33	1.00	
40	Modeling	0.05	0.18	0.22	-0.14	0.34	0.27	0.24	0.23	0.25	1.00
41	Variety	0.05	0.30	0.30	-0.08	0.56	0.41	0.37	0.28	0.46	0.31
42	Acceptance	0.05	0.17	0.12	-0.06	0.23	0.15	0.17	0.16	0.12	0.30
43	Respons. Emp.	0.04	0.20	0.17	-0.11	0.31	0.48	0.24	0.79	0.28	0.23
54-month Controls											
44	Letter Word (54)	0.09	0.55	0.37	-0.09	0.35	0.27	0.28	0.24	0.32	0.21
45	App. Prob. (54)	0.12	0.54	0.55	-0.15	0.40	0.24	0.30	0.23	0.23	0.16
46	Pic. Vocab. (54)	0.11	0.43	0.38	-0.13	0.39	0.20	0.25	0.25	0.28	0.16
47	Mem. Sent. (54)	0.10	0.38	0.42	-0.06	0.30	0.19	0.26	0.15	0.21	0.16
48	Inc. Words (54)	0.11	0.30	0.39	-0.09	0.28	0.25	0.27	0.13	0.23	0.15
49	Internalizing (54)	0.05	-0.04	0.02	0.14	-0.07	-0.00	-0.05	-0.01	0.03	-0.02
50	Externalizing (54)	0.04	-0.09	-0.01	0.13	-0.10	-0.05	-0.11	-0.10	-0.07	-0.10

PANEL 5		41	42	43	44	45	46	47	48	49	50
41	Variety	1.00									
42	Acceptance	0.16	1.00								
43	Respons. Emp.	0.25	0.18	1.00							
54-month Controls											
44	Letter Word (54)	0.31	0.17	0.19	1.00						
45	App. Prob. (54)	0.28	0.21	0.19	0.57	1.00					
46	Pic. Vocab. (54)	0.32	0.11	0.18	0.44	0.47	1.00				
47	Mem. Sent. (54)	0.31	0.11	0.09	0.40	0.47	0.42	1.00			
48	Inc. Words (54)	0.28	0.11	0.15	0.37	0.46	0.39	0.49	1.00		
49	Internalizing (54)	-0.06	-0.09	-0.03	0.02	-0.03	-0.03	-0.02	-0.04	1.00	
50	Externalizing (54)	-0.09	-0.15	-0.13	-0.03	-0.03	-0.05	-0.10	-0.05	0.59	1.00

Note. n=552. Only children from mothers who had not completed college were included here, and all non-missing cases for each pairwise correlation were included. Tables 2 and 3 include the full variable names and labels. The "G1" abbreviation stands for "grade 1," "15" stands for "age 15," and "54" stands for 54 months.

Table S2

Correlations Between All Analysis Variables for Children of Mothers Completed College

PANEL 1		1	2	3	4	5	6	7	8	9	10
Gratification Delay (54)											
1	Continuous	1.00									
2	<0.333 min.	-0.66	1.00								
3	0.333- 2 min.	-0.54	-0.11	1.00							
4	2 to 7 min.	-0.20	-0.12	-0.12	1.00						
5	7 min.	0.91	-0.48	-0.48	-0.55	1.00					
Related Measures											
6	Self-control (54)	0.17	-0.10	-0.17	0.01	0.17	1.00				
7	Attention (54)	0.15	-0.15	-0.02	-0.05	0.14	0.17	1.00			
8	Impulsivity (54)	-0.26	0.18	0.12	0.06	-0.24	-0.33	-0.29	1.00		
Outcome Measures											
9	Achievement (G1)	0.17	-0.16	-0.05	-0.03	0.15	0.25	0.25	-0.17	1.00	
10	Achievement (15)	0.15	-0.11	-0.08	-0.04	0.14	0.27	0.14	-0.15	0.59	1.00
11	Behavior (G1)	-0.05	0.06	0.04	-0.04	-0.03	-0.20	-0.08	-0.01	-0.02	-0.05
12	Behavior (15)	0.03	0.02	-0.00	-0.11	0.07	-0.18	-0.05	0.06	-0.06	-0.07
Demographic Controls											
13	Male	-0.03	0.04	-0.04	0.07	-0.05	-0.16	-0.01	0.22	-0.02	0.00
14	Black	-0.16	0.02	0.09	0.13	-0.16	-0.11	-0.05	0.02	-0.14	-0.18
15	Hispanic	-0.02	0.00	0.06	-0.06	0.01	-0.01	-0.01	0.01	0.02	-0.00
16	Other	-0.09	-0.02	0.12	0.05	-0.10	0.01	0.08	0.03	0.08	0.12
17	Age	0.03	-0.06	-0.01	0.06	0.00	0.05	0.14	-0.06	0.10	-0.08
18	Log of Income	0.14	-0.14	-0.03	0.00	0.11	0.15	0.01	-0.05	0.09	0.10
19	Mother's Age	-0.04	-0.00	0.01	0.04	-0.03	-0.08	-0.06	0.09	-0.09	0.04
20	Mother's Ed (yrs)	0.08	-0.06	-0.07	0.01	0.07	0.05	-0.02	-0.07	0.06	0.05
21	Mother PPVT	0.21	-0.10	-0.18	-0.09	0.24	0.16	0.02	-0.05	0.12	0.32
22	Site 1	0.06	-0.05	-0.05	0.01	0.06	-0.01	0.05	0.05	0.06	-0.06
23	Site 2	-0.02	0.01	0.04	-0.02	-0.02	-0.01	-0.01	0.01	0.12	0.15
24	Site 3	0.05	-0.02	-0.02	-0.02	0.04	-0.03	0.05	-0.06	-0.08	-0.13
25	Site 4	0.00	0.00	-0.03	0.04	-0.01	-0.00	0.11	-0.03	0.00	0.04
26	Site 5	-0.08	-0.00	0.09	0.05	-0.09	-0.03	0.07	-0.05	-0.02	-0.10
27	Site 6	0.04	-0.06	-0.03	0.04	0.02	0.10	0.02	0.01	0.02	0.06
28	Site 7	-0.02	0.06	-0.04	-0.00	-0.01	-0.02	-0.12	0.09	0.08	0.14
29	Site 8	0.07	0.05	-0.10	-0.10	0.10	0.05	-0.01	-0.05	-0.09	0.00
30	Site 9	-0.09	0.02	0.10	0.04	-0.10	-0.04	-0.03	0.07	-0.06	-0.10
31	Birthweight (g's)	-0.08	0.09	0.05	-0.06	-0.05	-0.04	-0.00	0.04	0.09	0.02
32	Bracken	0.14	-0.17	-0.01	0.01	0.11	0.26	0.22	-0.18	0.40	0.33
33	Bayley	0.26	-0.25	-0.01	-0.06	0.21	0.22	0.16	-0.27	0.24	0.22
34	Temperament	0.04	-0.03	0.04	-0.08	0.05	-0.04	0.02	-0.06	0.01	0.05
H.O.M.E. Controls											
35	Learn. Mater.	0.09	-0.12	-0.04	0.06	0.06	0.17	0.05	-0.15	0.13	0.13
36	Lang. Stim.	0.14	-0.19	0.02	0.00	0.11	0.15	0.00	-0.12	0.08	0.07
37	Phys. Env.	0.06	-0.07	-0.03	0.07	0.01	-0.03	0.02	-0.01	-0.06	-0.10
38	Responsivity	0.11	-0.11	-0.04	-0.04	0.12	0.06	0.04	-0.05	-0.09	-0.04
39	Academ. Stim.	0.13	-0.14	-0.04	0.06	0.07	0.06	-0.08	-0.10	0.04	-0.02
40	Modeling	0.06	-0.08	0.05	-0.05	0.06	0.08	0.01	0.03	0.03	0.06
41	Variety	0.12	-0.11	-0.06	0.02	0.10	0.14	-0.04	-0.09	0.02	0.04
42	Acceptance	0.08	-0.10	0.04	-0.07	0.09	0.17	0.06	-0.14	0.07	0.06

43	Respons. Emp.	0.15	-0.17	-0.01	-0.06	0.15	0.08	0.06	-0.07	-0.01	0.04
54-month Controls											
44	Letter Word (54)	0.16	-0.16	-0.04	-0.02	0.14	0.19	0.15	-0.14	0.57	0.39
45	App. Prob. (54)	0.26	-0.15	-0.19	-0.03	0.23	0.28	0.27	-0.28	0.48	0.50
46	Pic. Vocab. (54)	0.20	-0.13	-0.11	-0.06	0.19	0.21	0.20	-0.13	0.33	0.37
47	Mem. Sent. (54)	0.22	-0.19	-0.08	-0.06	0.22	0.23	0.17	-0.17	0.31	0.32
48	Inc. Words (54)	0.17	-0.07	-0.11	-0.07	0.16	0.11	0.07	-0.13	0.27	0.18
49	Internalizing (54)	-0.02	0.03	0.01	-0.02	-0.01	-0.15	-0.07	0.04	-0.00	-0.01
50	Externalizing (54)	-0.03	0.01	0.07	-0.03	-0.03	-0.37	-0.07	0.10	-0.03	-0.04

PANEL 2		11	12	13	14	15	16	17	18	19	20
11	Behavior (G1)	1.00									
12	Behavior (15)	0.56	1.00								
Demographic Controls											
13	Male	-0.03	-0.04	1.00							
14	Black	-0.07	-0.02	-0.01	1.00						
15	Hispanic	-0.00	-0.02	-0.02	-0.02	1.00					
16	Other	0.08	-0.02	0.03	-0.03	-0.03	1.00				
17	Age	0.02	0.02	-0.02	0.07	-0.02	-0.07	1.00			
18	Log of Income	-0.07	-0.03	-0.01	-0.03	-0.08	0.08	0.00	1.00		
19	Mother's Age	0.05	-0.04	0.01	-0.02	-0.12	-0.04	-0.03	0.22	1.00	
20	Mother's Ed (yrs)	-0.09	-0.10	-0.01	-0.01	-0.05	0.02	0.05	0.24	0.28	1.00
21	Mother PPVT	0.01	0.06	0.06	-0.16	-0.14	-0.18	-0.04	0.21	0.31	0.28
22	Site 1	0.07	0.02	0.03	-0.04	-0.04	-0.05	0.05	-0.06	-0.05	-0.02
23	Site 2	-0.04	-0.07	0.02	-0.05	0.05	-0.07	-0.17	0.13	0.01	-0.11
24	Site 3	0.08	-0.05	0.04	-0.05	0.16	-0.07	0.12	-0.19	-0.06	-0.01
25	Site 4	-0.03	-0.01	-0.06	0.03	0.01	0.04	0.04	0.12	0.05	-0.03
26	Site 5	0.00	0.05	0.04	0.01	-0.06	-0.03	0.23	-0.02	-0.01	-0.03
27	Site 6	-0.15	-0.07	-0.07	0.11	-0.07	0.03	-0.09	0.12	0.03	0.02
28	Site 7	-0.03	-0.03	0.01	-0.04	0.06	-0.02	0.05	-0.01	0.02	0.06
29	Site 8	0.10	0.11	-0.01	0.00	-0.02	0.16	-0.20	0.07	0.06	0.05
30	Site 9	0.04	0.04	0.04	0.04	-0.04	-0.05	0.03	-0.15	-0.06	-0.05
31	Birthweight (g's)	-0.00	0.13	0.15	-0.03	-0.03	-0.03	-0.07	-0.04	-0.03	-0.02
32	Bracken	0.00	-0.03	-0.19	-0.14	-0.14	0.02	-0.01	0.20	-0.04	0.05
33	Bayley	-0.05	-0.06	-0.20	-0.14	-0.05	-0.03	-0.03	0.15	-0.05	0.07
34	Temperament	0.10	0.12	-0.07	0.02	-0.06	0.02	-0.06	-0.03	-0.05	-0.00
H.O.M.E. Controls											
35	Learn. Mater.	-0.05	-0.01	-0.05	-0.18	-0.20	-0.07	0.04	0.09	0.02	0.05
36	Lang. Stim.	-0.08	-0.02	-0.02	0.04	-0.16	-0.07	-0.02	0.05	-0.03	0.05
37	Phys. Env.	-0.03	0.03	0.09	-0.03	0.05	0.05	-0.01	-0.06	-0.14	-0.02
38	Responsivity	-0.12	-0.07	-0.02	-0.00	-0.17	-0.08	-0.07	0.09	0.10	0.06
39	Academ. Stim.	0.02	0.05	-0.07	-0.03	-0.08	-0.05	-0.02	-0.01	-0.04	0.01
40	Modeling	-0.10	-0.05	-0.06	0.04	-0.05	-0.04	-0.06	0.09	-0.01	0.01
41	Variety	-0.08	-0.01	-0.00	-0.04	-0.12	-0.07	0.04	0.07	-0.01	0.08
42	Acceptance	-0.10	-0.06	-0.07	0.01	-0.06	0.07	0.02	0.12	0.04	0.09
43	Respons. Emp.	-0.07	-0.04	0.03	-0.06	-0.10	-0.05	-0.05	0.10	0.08	0.09
54-month Controls											
44	Letter Word (54)	-0.05	-0.03	-0.15	-0.02	-0.04	0.07	0.03	0.17	-0.12	0.07
45	App. Prob. (54)	0.01	-0.04	-0.10	-0.18	-0.01	0.00	0.01	0.11	0.01	0.13
46	Pic. Vocab. (54)	-0.09	0.02	0.12	-0.12	-0.12	0.01	0.04	0.20	0.02	0.13

47	Mem. Sent. (54)	-0.10	-0.08	-0.02	-0.06	-0.02	0.02	0.05	0.09	0.02	0.08
48	Inc. Words (54)	-0.09	-0.09	0.00	-0.08	-0.06	-0.05	0.03	-0.02	-0.03	0.07
49	Internalizing (54)	0.56	0.38	0.02	-0.10	-0.05	0.05	0.01	0.02	0.05	-0.05
50	Externalizing (54)	0.59	0.50	-0.10	-0.04	-0.00	0.00	-0.03	-0.05	0.04	-0.08

PANEL 3		21	22	23	24	25	26	27	28	29	30
21	Mother PPVT	1.00									
22	Site 1	-0.01	1.00								
23	Site 2	-0.07	-0.09	1.00							
24	Site 3	-0.08	-0.09	-0.12	1.00						
25	Site 4	0.01	-0.08	-0.10	-0.10	1.00					
26	Site 5	-0.01	-0.09	-0.12	-0.12	-0.11	1.00				
27	Site 6	0.08	-0.11	-0.14	-0.14	-0.12	-0.15	1.00			
28	Site 7	0.09	-0.08	-0.11	-0.11	-0.09	-0.11	-0.13	1.00		
29	Site 8	0.11	-0.10	-0.14	-0.14	-0.12	-0.14	-0.16	-0.13	1.00	
30	Site 9	-0.10	-0.07	-0.09	-0.09	-0.08	-0.10	-0.11	-0.08	-0.11	1.00
31	Birthweight (g's)	0.11	0.05	0.05	-0.05	-0.00	-0.06	-0.08	-0.03	0.10	0.05
32	Bracken	0.11	-0.05	0.08	-0.04	0.00	-0.06	0.06	0.00	-0.04	-0.05
33	Bayley	0.08	0.00	0.07	-0.10	0.09	-0.01	0.05	-0.06	-0.11	-0.06
34	Temperament	-0.09	0.03	-0.02	-0.00	-0.01	-0.07	0.02	-0.00	0.08	0.00
H.O.M.E. Controls											
35	Learn. Mater.	0.19	-0.02	-0.19	-0.01	-0.13	0.05	0.05	0.13	-0.09	0.05
36	Lang. Stim.	0.10	0.01	-0.15	-0.11	-0.22	0.04	0.17	0.09	-0.00	0.08
37	Phys. Env.	-0.14	-0.06	-0.08	-0.01	-0.02	0.14	0.03	-0.18	-0.07	0.17
38	Responsivity	0.09	-0.13	-0.01	-0.04	-0.02	-0.12	0.34	-0.27	-0.02	0.05
39	Academ. Stim.	0.07	0.03	-0.24	-0.05	-0.13	0.07	0.11	-0.04	-0.00	0.10
40	Modeling	0.10	-0.09	0.02	-0.11	-0.07	0.00	0.15	-0.04	0.00	-0.08
41	Variety	0.19	0.09	-0.20	-0.08	-0.09	-0.02	0.17	-0.02	0.02	0.06
42	Acceptance	0.08	-0.11	-0.13	-0.06	-0.01	-0.09	0.17	0.08	0.04	-0.08
43	Respons. Emp.	0.14	-0.02	-0.03	0.00	-0.04	-0.13	0.17	-0.03	-0.06	0.00
54-month Controls											
44	Letter Word (54)	0.08	0.02	-0.01	-0.08	-0.00	-0.05	0.12	0.06	-0.02	-0.08
45	App. Prob. (54)	0.26	-0.07	0.02	-0.02	0.02	-0.03	0.09	0.03	0.03	-0.10
46	Pic. Vocab. (54)	0.34	-0.06	-0.00	-0.03	-0.01	-0.03	0.06	0.03	0.12	-0.02
47	Mem. Sent. (54)	0.19	-0.07	-0.06	-0.08	0.05	0.05	0.07	0.08	0.02	-0.03
48	Inc. Words (54)	0.14	-0.01	-0.06	-0.03	-0.04	0.03	0.10	0.11	-0.10	0.08
49	Internalizing (54)	-0.04	-0.00	-0.05	-0.06	-0.03	0.12	-0.12	0.00	0.06	0.03
50	Externalizing (54)	-0.06	0.04	0.01	0.02	-0.02	0.04	-0.04	-0.04	-0.01	-0.00

PANEL 4		31	32	33	34	35	36	37	38	39	40
31	Birthweight (g's)	1.00									
32	Bracken	-0.03	1.00								
33	Bayley	-0.01	0.45	1.00							
34	Temperament	-0.01	-0.02	-0.05	1.00						
H.O.M.E. Controls											
35	Learn. Mater.	-0.03	0.31	0.22	0.02	1.00					
36	Lang. Stim.	0.13	0.21	0.15	0.05	0.53	1.00				
37	Phys. Env.	-0.10	0.02	0.01	0.01	0.16	0.09	1.00			
38	Responsivity	0.00	0.22	0.15	-0.04	0.28	0.28	0.15	1.00		

39	Academ. Stim.	-0.01	0.24	0.16	0.05	0.48	0.48	0.18	0.24	1.00	
40	Modeling	0.06	0.08	0.08	0.03	0.18	0.27	0.18	0.26	0.21	1.00
41	Variety	-0.03	0.18	0.27	-0.02	0.30	0.28	0.10	0.26	0.23	0.22
42	Acceptance	0.01	0.15	0.16	0.05	0.17	0.23	0.02	0.14	0.05	0.23
43	Respons. Emp.	-0.00	0.18	0.08	-0.10	0.30	0.43	0.17	0.68	0.21	0.27
54-month Controls											
44	Letter Word (54)	-0.01	0.55	0.25	0.04	0.21	0.20	-0.03	0.14	0.25	0.07
45	App. Prob. (54)	-0.02	0.44	0.42	0.02	0.19	0.08	0.00	0.09	0.12	-0.03
46	Pic. Vocab. (54)	0.10	0.30	0.36	-0.01	0.23	0.18	0.01	0.14	0.13	0.06
47	Mem. Sent. (54)	0.03	0.26	0.35	-0.03	0.15	0.13	-0.07	0.09	0.13	-0.05
48	Inc. Words (54)	0.03	0.15	0.19	-0.06	0.14	0.12	-0.01	0.09	0.10	-0.02
49	Internalizing (54)	0.02	0.04	-0.08	0.11	-0.00	-0.07	0.08	-0.08	0.04	-0.03
50	Externalizing (54)	0.05	-0.06	-0.06	0.06	-0.08	-0.06	0.00	-0.05	-0.03	-0.07

PANEL 5		41	42	43	44	45	46	47	48	49	50
41	Variety	1.00									
42	Acceptance	0.21	1.00								
43	Respons. Emp.	0.24	0.29	1.00							
54-month Controls											
44	Letter Word (54)	0.09	0.04	0.12	1.00						
45	App. Prob. (54)	0.12	0.07	0.08	0.45	1.00					
46	Pic. Vocab. (54)	0.25	0.07	0.15	0.32	0.44	1.00				
47	Mem. Sent. (54)	0.13	0.10	0.09	0.32	0.36	0.40	1.00			
48	Inc. Words (54)	0.10	-0.02	0.09	0.22	0.32	0.22	0.42	1.00		
49	Internalizing (54)	-0.05	0.00	-0.08	-0.02	-0.01	-0.12	-0.08	-0.02	1.00	
50	Externalizing (54)	-0.08	-0.07	-0.04	0.01	-0.01	-0.08	-0.07	0.00	0.57	1.00

Note. n=366. Only children from mothers completed college were included here, and all non-missing cases for each pairwise correlation were included. Tables 2 and 3 include the full variable names and labels. The "G1" abbreviation stands for "grade 1," "15" stands for "age 15," and "54" stands for 54 months.

Table S3

Delay of Gratification and Later Outcomes with Controls for Attention and Impulsivity Added (Lower-SES sample)

	Grade 1 Achievement	Age 15 Achievement	Grade 1 Behavior	Age 15 Behavior
	(1)	(2)	(3)	(4)
Delay of Gratification (categorical)				
<0.333 minutes	ref	ref	ref	ref
0.333- 2 minutes	0.166	0.200	0.096	-0.073
	(0.104)	(0.105)	(0.139)	(0.150)
2 to 7 minutes	0.186	0.271*	-0.016	-0.110
	(0.103)	(0.105)	(0.138)	(0.146)
7 minutes	0.206*	0.185*	0.006	-0.103
	(0.088)	(0.092)	(0.118)	(0.128)
Continuous Performance Task				
Sustained Attention	0.088*	0.026	-0.013	0.013
	(0.034)	(0.035)	(0.044)	(0.048)
Impulsivity	-0.072*	-0.062	0.011	-0.005
	(0.036)	(0.039)	(0.047)	(0.053)
p-value of test of equality of all categories	.113	.056	.868	.853
p-value of test of equality of 2^{nd}, 3^{rd} and 4^{th} categories	.914	.646	.714	.968
Child demographic and H.O.M.E. controls	Inc.	Inc.	Inc.	Inc.
Concurrent 54-month controls	-	-	-	-

Note. n= 552. Standard errors are in parentheses. These estimates compare to the middle column in each set of estimates in Table 4 (i.e., the estimates with controls measured prior to age 54 months), with standardized measures of age 54-month attention and impulsivity added. See Table 4 note for full explanation of model parameters.

* p< .05

Table S4

Delay of Gratification and Later Outcomes with Index of Self-Control Added (Lower-SES Sample)

	Grade 1 Achievement	Age 15 Achievement	Grade 1 Behavior	Age 15 Behavior
	(1)	(2)	(3)	(4)
Delay of Gratification (categorical)				
<0.333 minutes	ref	ref	ref	ref
0.333- 2 minutes	0.221*	0.252*	0.050	-0.108
	(0.103)	(0.102)	(0.133)	(0.146)
2 to 7 minutes	0.204*	0.305*	-0.018	-0.111
	(0.102)	(0.102)	(0.132)	(0.142)
7 minutes	0.261*	0.224*	0.040	-0.068
	(0.085)	(0.087)	(0.109)	(0.121)
Self-Control Composite	0.157*	0.114*	-0.303*	-0.212*
	(0.034)	(0.035)	(0.044)	(0.049)
p-value of test of equality of all categories	.018*	.011*	.943	.848
p-value of test of equality of 2^{nd}, 3^{rd} and 4^{th} categories	.800	.680	.861	.926
Child demographic and H.O.M.E. controls	Inc.	Inc.	Inc.	Inc.
Concurrent 54-month controls	-	-	-	-

Note. n= 552. Standard errors are in parentheses. These estimates compare to the middle column in each set of estimates in Table 4 (i.e., the estimates with controls measured prior to age 54 months), with standardized measures of age 54-month self-control added. See Table 4 note for full explanation of model parameters.

Table S5
Descriptive Characteristics of Supplemental Age 15 Behavioral Measures (Lower-SES Sample)

	M (SD)
Age 15 Behavioral Measures	
Stoplight- Brake Applications	4.85 (1.47)
Stoplight- Brake Time (ms)	900.26 (358.33)
Internalizing (self-report)	47.44 (10.38)
Externalizing (self-report)	50.15 (10.02)
Impulse Control	3.48 (0.89)
Risk Taking	6.81 (5.82)
Observations	478

Note. See Table 2 note. Mean values are presented in each cell, and standard deviations are in parentheses.

Table S6

Associations Between Age 54 month Measure of Delay of Gratification and Measures of Age 15 Risk Taking (Lower-SES Sample)

	Stoplight- Brake Applications			Stoplight- Time Waited			Risk Taking		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Delay minutes (categorical)									
<0.333 minutes	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333- 2 minutes	-0.085	-0.182	-0.179	0.073	0.096	0.076	-0.072	0.087	0.085
	(0.154)	(0.156)	(0.156)	(0.155)	(0.153)	(0.153)	(0.151)	(0.147)	(0.146)
2 to 7 minutes	0.234	0.143	0.106	0.111	0.179	0.188	-0.065	0.113	0.120
	(0.156)	(0.156)	(0.158)	(0.156)	(0.153)	(0.155)	(0.151)	(0.145)	(0.146)
7 minutes	0.094	0.032	0.011	0.067	0.089	0.107	-0.377*	-0.129	-0.129
	(0.123)	(0.132)	(0.133)	(0.124)	(0.129)	(0.130)	(0.120)	(0.123)	(0.124)
p-value of test of equality of all categories	.236	.254	.374	.906	.709	.677	.004*	.189	.174
p-value of test of equality of 2^{nd}, 3^{rd} and 4^{th} categories	.157	.131	.213	.949	.793	.765	.015*	.093	.084
Child demographic and H.O.M.E. controls	-	Inc.	Inc.	-	Inc.	Inc.	-	Inc.	Inc.
Concurrent 54-month controls	-	-	Inc.	-	-	Inc.	-	-	Inc.

Note. n = 552. Standard errors are in parentheses. See Table 4 note.

* p< .05

Table S7

Associations Between Age 54 month Measure of Delay of Gratification and Measures of Age 15 Behavior Problems (Lower-SES Sample)

	Internalizing			Externalizing			Impulse Control		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Delay minutes (categorical)									
<0.333 minutes	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333- 2 minutes	0.107	0.196	0.182	-0.011	0.128	0.090	0.040	-0.050	-0.039
	(0.152)	(0.150)	(0.151)	(0.150)	(0.148)	(0.148)	(0.146)	(0.143)	(0.143)
2 to 7 minutes	-0.032	0.047	0.008	-0.124	0.039	0.020	-0.056	-0.150	-0.130
	(0.152)	(0.149)	(0.151)	(0.150)	(0.147)	(0.149)	(0.146)	(0.142)	(0.144)
7 minutes	-0.075	-0.024	-0.038	-0.156	-0.041	-0.068	0.087	-0.041	-0.024
	(0.121)	(0.127)	(0.128)	(0.120)	(0.125)	(0.126)	(0.116)	(0.120)	(0.122)
p-value of test of equality of all categories	.594	.425	.443	.503	.641	.676	.704	.752	.811
p-value of test of equality of 2^{nd}, 3^{rd} and 4^{th} categories	.400	.270	.270	.548	.438	.467	.545	.672	.695
Child demographic and H.O.M.E. controls	-	Inc.	Inc.	-	Inc.	Inc.	-	Inc.	Inc.
Concurrent 54-month controls	-	-	Inc.	-	-	Inc.	-	-	Inc.

Note. n = 552. Standard errors are in parentheses. See Table 4 note.

* p< .05

Table S8

Associations Between Age 54 month Measure of Delay of Gratification and Disaggregated Measures of Age 15 Achievement and Behavior (Compare with Table 4)

	Math Achievement			Reading Achievement			Externalizing			Internalizing		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Delay minutes (categorical)												
<0.333 minutes	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
0.333- 2 minutes	0.305*	0.192	0.143	0.339*	0.222*	0.176	-0.173	-0.125	-0.158	-0.076	0.002	-0.030
	(0.122)	(0.109)	(0.105)	(0.128)	(0.112)	(0.107)	(0.152)	(0.148)	(0.131)	(0.150)	(0.148)	(0.138)
2 to 7 minutes	0.440*	0.287*	0.197	0.400*	0.257*	0.220*	-0.180	-0.114	-0.046	-0.146	-0.080	-0.049
	(0.124)	(0.110)	(0.107)	(0.129)	(0.112)	(0.108)	(0.151)	(0.145)	(0.129)	(0.149)	(0.145)	(0.136)
7 minutes	0.569*	0.222*	0.130	0.615*	0.208*	0.142	-0.307*	-0.196	-0.140	-0.032	0.033	0.060
	(0.098)	(0.093)	(0.090)	(0.103)	(0.095)	(0.092)	(0.120)	(0.123)	(0.109)	(0.119)	(0.123)	(0.116)
p-value of test of equality of all categories	.001*	.042*	.278	.001*	.070	.183	.084	.465	.503	.777	.861	.790
p-value of test of equality of 2nd, 3rd and 4th categories	.046*	.699	.778	.024*	.887	.715	.477	.770	.655	.688	.688	.596
Child demographic and H.O.M.E. controls	-	Inc.	Inc.	-	Inc.	Inc.	-	Inc.	Inc.	-	Inc.	Inc.
Concurrent 54-month controls	-	-	Inc.	-	-	Inc.	-	-	Inc.	-	-	Inc.

Note. n = 552. Standard errors are in parentheses. See Table 4 note.

* p< .05