Research Article

Psychological Science
2020, Vol. 31(2) 193–201

DOI: 10.1177/0956797619896270
www.psychologicalscience.org/PS

Same Data Set, Different Conclusions: Preschool Delay of Gratification Predicts Later Behavioral Outcomes in a Preregistered Study

Laura E. Michaelson[1] and Yuko Munakata[2,3]
[1]American Institutes for Research, Washington, D.C.; [2]Department of Psychology, University of California, Davis; and [3]Center for Mind and Brain, University of California, Davis

Abstract
One simple marshmallow test in preschool children predicts an array of important life outcomes, according to multiple studies spanning several decades. However, a recent conceptual replication casts doubt on these famous findings. We conducted an independent, preregistered secondary analysis to test whether previously observed longitudinal associations between preschool delay of gratification and adolescent outcomes would be conceptually replicated. Associations were significant for three of the five outcomes we tested using the analytic approach employed in the original studies of the marshmallow test. Relationships between delay of gratification and problem behavior held in bivariate, multivariate, and multilevel models; in contrast, no significant relationships between delay and problem behavior were found in the other recent replication, even though both studies used the same data set. These relationships were better explained by social support than by self-control, suggesting that the marshmallow test is predictive because it reflects aspects of a child's early environment that are important over the long term. This novel interpretation of the classic findings points to new directions for intervention.

Keywords
delay of gratification, marshmallow test, problem behavior, self-control, social support, open data, preregistered

Received 7/5/18; Revision accepted 10/22/19

The marshmallow test is one of the most famous tasks in psychology. Children sit alone with a tempting treat (e.g., two marshmallows). Whether and when they eat the treat indexes their ability to delay gratification. This simple measure predicts important outcomes throughout the life span (e.g., Duckworth, Tsukayama, & Kirby, 2013; Mischel, Shoda, & Peake, 1988; Shoda, Mischel, & Peake, 1990) and has inspired research and policies aimed at improving delay of gratification to promote later benefits.

However, a recent replication casts doubt on this finding. Using a data set containing a larger and more diverse participant sample, Watts, Duncan, and Quan (2018) found associations between marshmallow-delay time and adolescent outcomes that were "much smaller and rarely statistically significant" (p. 1159). These findings have been viewed as a failure to replicate the results of previous studies demonstrating the predictive validity of the preschool marshmallow test for adolescent outcomes, creating skepticism about attempts to promote development by targeting delay of gratification.

We independently analyzed the same data set to conceptually replicate and extend the marshmallow-test associations. Our study was preregistered in 2017 (https://osf.io/vjmkz) and was completed prior to the publication of Watts et al.'s (2018) study. We made different decisions regarding measurement and analysis, which led to different results and conclusions.

Research Aims

Our first aim was to replicate the predictive validity of the marshmallow test. We selected five outcomes to

Corresponding Author:
Laura E. Michaelson, American Institutes for Research, 1000 Thomas Jefferson St. NW, Washington, D.C. 20007
E-mail: lmichaelson@air.org

conceptually match those tested in prior studies (e.g., Mischel et al., 1988; Shoda et al., 1990): academic achievement, emotion regulation, social skills, personality traits, and problem behavior. We tested simple bivariate models to match those used in the original studies and multivariate models that included a common covariate adjustment not possible in the original studies.

Our second aim was to evaluate two explanations for longitudinal associations, both of which have distinct implications for intervention. Prominent theories emphasize cognitive abilities such as self-control (e.g., Casey et al., 2011; Duckworth et al., 2013; Mischel et al., 2011). Individuals vary in their ability to regulate behavioral, emotional, and attentional impulses to achieve long-term goals, and these individual differences show stability across the life span (Friedman, Miyake, Robinson, & Hewitt, 2011; Moffitt et al., 2011). Thus, the marshmallow test may be predictive because preschoolers who can engage self-control to wait for two marshmallows become individuals who can engage self-control to study, exercise, get along with others, and save money.

Alternatively, the marshmallow test may be predictive because it captures social and situational factors. Preschoolers and adults are less willing to delay gratification with untrustworthy individuals or when delaying is not normative or rational (Doebel & Munakata, 2018; Jachimowicz, Chafik, Munrat, Prabhu, & Weber, 2017; Kidd, Palmeri, & Aslin, 2013; Lee & Carlson, 2015; McGuire & Kable, 2012, 2013; Michaelson, de la Vega, Chatham, & Munakata, 2013; Michaelson & Munakata, 2016; Pepper & Nettle, 2017; Schneider, Peters, Peth, & Büchel, 2014). Thus, a child's ability to wait might be less important than the social and environmental circumstances influencing their willingness to wait.

Method

Data

Data for our study were drawn from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD), a prospective longitudinal study of children and their families conducted in the United States from 1991 through 2006 (NICHD Early Child Care Research Network, 2002). This data set contains a variety of psychosocial and environmental measures from 1,364 participants recruited at birth and assessed approximately annually until age 15 (n = 1,009 retained at the final assessment). The sample is not nationally representative but reflects the demographic diversity of the 10 recruitment sites at the time of recruitment (76% White; 13% African American; 6% Hispanic; and 5% Asian, Native American, or other race).

Measures

We preregistered all measures of study variables before testing our research questions and used only ratified variables created by NICHD investigators in our analysis. Table 1 lists all study variables, along with timing of assessments and descriptive information. We measured delay of gratification using a modified version of the marshmallow test administered to study participants as preschoolers, and we created composite variables for outcome variables and explanatory factors. Measured variables to be included in composites were identified using a systematic search-and-selection process (for details, see the Supplemental Material available online). Only measures that demonstrated significant correlations with all other retained measures were included in each composite, and the number of retained measures was maximized; these measures were z scored and averaged to form composites.

Delay of gratification. Participants completed a modified version of the classic marshmallow test when they were 4 years old. An experimenter presented children with a snack treat and told them they would be left alone with the treat for 7 min. Children could either wait to eat the treat until the experimenter returned—in which case they would receive an additional portion as a reward for waiting—or choose not to wait and receive the amount originally presented. The delay-of-gratification measure was the number of seconds waited.

Wait times were not normally distributed because of right censoring of the distribution from the 55% of children who waited until the experimenter returned. We therefore dummy-coded delay of gratification to indicate whether the child waited the full delay period (1 = delayed full 7 min, 0 = did not delay full 7 min). This approach avoided anchoring at 7 min for the children whose wait times were not observed. Watts et al. (2018) addressed this issue by dividing wait times into four discrete intervals using three dummy variables to explore nonlinearities. They then separately analyzed subgroups based on parental education, because children whose mothers had not completed college were less likely to have hit the 7-min ceiling. In both the present study and Watts et al.'s study, the children who waited more than 7 min were treated in equivalent ways, but our approach matched that used in previous research (Duckworth et al., 2013) and allowed us to estimate relationships using the full sample rather than smaller subgroups.

Outcomes. Our outcome variables were academic achievement, emotion regulation, personality, problem behavior, and social skills. We selected these outcomes

Table 1. Descriptive Statistics for Study Variables

Variable and measurement age	*n*	*M*	*SD*	Minimum	Maximum	Skewness	Kurtosis
Academic achievement							
4 years	908	0.00	0.68	−5.66	1.35	−1.09	6.03
15 years	668	0.00	0.74	−1.61	1.98	0.24	−0.36
Emotion regulation							
4 years	646	0.00	0.78	−1.64	3.02	0.33	−0.29
15 years	875	0.00	0.90	−3.06	1.65	−0.58	−0.12
Personality							
4 years	842	0.00	0.71	−2.89	2.27	0.01	0.47
15 years	866	0.00	0.86	−3.89	1.88	−0.47	0.28
Problem behavior							
4 years	903	0.00	0.89	−2.08	3.02	0.33	−0.22
15 years	847	0.00	0.75	−1.28	4.24	1.15	2.12
Social skills							
4 years	895	0.00	0.75	−2.95	1.98	−0.31	0.10
15 years	839	0.00	0.82	−3.20	1.52	−0.46	−0.16
Delay of gratification							
4 years	840	4.55	2.98	0.00	7.00	−0.54	−1.55
Self-control							
4–15 years	953	0.00	0.69	−2.35	1.94	−0.32	0.02
Social support							
4–15 years	951	0.00	0.76	−3.06	1.87	−0.82	0.02

Note: All variables except delay of gratification are composites created from multiple measures. Measures were *z* scored with a mean of 0 and a standard deviation of 1 before being averaged to form composites. Descriptive information is reported for the continuous delay-of-gratification measure, which was dummy coded prior to analysis to address right censoring in the distribution.

to match those examined in prior follow-up studies that also tested 15-year-olds (Mischel et al., 1988; Shoda et al., 1990) and to reflect the broader array of adult behaviors and attributes that have been associated with the preschool marshmallow test. We created composite variables for each outcome in preschool (age 4) and adolescence (age 15). The adolescent outcomes were our dependent variables, and the age-4 outcomes were used as control variables in multivariate models to help isolate adolescent-specific variance in the dependent variables.

Explanatory factors. Our explanatory factors were self-control and social support. We selected these factors for their theoretical and empirical relevance to the delay of gratification and its long-term correlates and their distinct directions for intervention. We created one composite variable for each explanatory factor from multiple observations spanning preschool through adolescence (ages 4–15) in order to test relationships across the same developmental period as the longitudinal associations between preschool delay of gratification and adolescent outcomes.

Analytic approach

Predictive validity of the preschool marshmallow test. We used bivariate and multivariate regression models to test for conceptual replication of the predictive validity of the preschool marshmallow test for each of our five outcome variables. The bivariate models estimated direct correlations between delay of gratification at age 4 and each outcome variable at age 15, as in the original studies by Mischel et al. (1988) and Shoda et al. (1990). The multivariate models estimated the same correlations while controlling for age-4 outcome levels (i.e., preschool levels of the adolescent outcome constructs). This latter approach was not possible in the original studies because baseline outcomes were not measured, but it tests the same question—that is, whether the preschool marshmallow test predicts outcomes in adolescence—while isolating adolescent-specific variance in the dependent variables. All models were restricted to complete cases to address missing data on predictor and outcome variables.

Explanatory factors in longitudinal associations. To address possible explanations for relationships between the

Table 2. Results of the Bivariate Regression Analysis Predicting Adolescent Outcomes From Preschool Delay of Gratification

Outcome	*n*	β	*SE*	*F*	*p*
Academic achievement	602	0.27	0.08	$F(1, 600) = 11.22$	< .001
Emotion regulation	777	−0.02	0.07	$F(1, 775) = 0.08$	.78
Personality	770	0.09	0.07	$F(1, 768) = 1.57$	.21
Problem behavior	756	−0.23	0.07	$F(1, 754) = 9.38$	.002
Social skills	748	0.18	0.07	$F(1, 746) = 5.78$	.02

Note: Preschool delay of gratification was dummy coded for marshmallow-test behavior (1 = delayed full 7 min, 0 = did not delay full 7 min).

marshmallow test and later outcomes, we took advantage of the 15 years of longitudinal data available in the SECCYD data set and created multilevel models of individual change trajectories in behavioral outcomes across childhood. Multilevel modeling accounts for repeated measures nested within individuals over time while allowing explicit estimation and modeling of variance within and between individuals. We used two-level models with restricted maximum-likelihood estimation; the Level 1 models described how individuals' outcomes change across development, and the Level 2 models described how changes differed across individuals (Raudenbush & Bryk, 2002). Time was represented continuously at Level 1 and zero-centered around age 15, yielding a set of intercepts reflecting the predicted adolescent outcome for each individual. These intercepts were our dependent variables at Level 2, which we predicted from independent variables—preschool delay of gratification, self-control, and social support.

For our focal questions about possible explanations, we evaluated the significance of our Level 2 independent variables using nested model comparisons, which test whether a given variable explains significant and unique variance in the outcome given the other variables in the model. (This is akin to examining the significance of an individual predictor in an ordinary least squares regression.) We then compared the Level 2 variance explained by self-control versus social support after accounting for preschool delay of gratification to see which variable better explains longitudinal associations between preschool delay and adolescent outcomes.

We limited these analyses to problem behavior, given the theoretical and empirical relevance of this construct to delaying gratification in multiple applied domains, including psychopathology (Ayduk et al., 2008; Wulfert, Block, Santa Ana, Rodriguez, & Colsman, 2002), addiction (Bickel, Koffarnus, Moody, & Wilson, 2014; Yi, Carter, & Landes, 2012), and criminal offending (Arantes, Berg, Lawlor, & Grace, 2013; Lee, Derefinko, Milich, Lynam, & DeWall, 2017). To improve interpretation, we used a single measure of problem behavior in multilevel analyses, rather than a composite variable, because standardized variables do not reflect change over time. We chose the parent-reported Total Problems scores on the Child Behavior Checklist (CBCL; Achenbach, 1991), given the extensive psychometric information indicating the high reliability, internal consistency, and validity of this measure.

Results

Predictive validity of the preschool marshmallow test

We found significant bivariate associations between the preschool marshmallow test and three out of the five adolescent outcomes we tested. Preschoolers who delayed gratification on the marshmallow test had significantly better academic achievement, $\beta = 0.27$, $F(1, 600) = 11.22$; fewer problem behaviors, $\beta = -0.22$, $F(1, 754) = -9.38$; and better social skills, $\beta = 0.18$, $F(1, 746) = 5.78$, in adolescence than those who did not delay gratification (all ps < .05; see Table 2). Bivariate associations between preschool delay of gratification and adolescent emotion regulation and personality traits were not statistically distinguishable from zero.

In multivariate models that adjusted for preschool levels of the adolescent outcome constructs (Table 3), preschool delay of gratification continued to predict significantly less problem behavior in adolescence, $\beta = -0.18$, $F(1, 741) = 6.45$, $p = .01$. Relationships between delay of gratification and academic achievement and between delay of gratification and social skills were no longer significant. This weakening of relationships may reflect a combination of (a) concurrent associations between delay of gratification and academic and social functioning in preschool and (b) correlations between repeated measures of the same outcome over time, suggesting that the marshmallow test predicts adolescent academic and social outcomes because it predicts those outcomes in preschool and because academic and social functioning show relative stability across development.

Table 3. Results of the Multivariate Regression Analysis Predicting Adolescent Outcomes From Preschool Level of Adolescent Outcome Construct and Preschool Delay of Gratification

		Preschool level of outcome				Delay of Gratification			
Outcome	*n*	β	*SE*	*F*	*p*	β	*SE*	*F*	*p*
Academic achievement	600	0.37	0.04	$F(1, 597) = 91.82$	< .001	0.04	0.08	$F(1, 597) = 0.30$	.58
Emotion regulation	546	0.35	0.04	$F(1, 543) = 75.53$	< .001	−0.10	0.08	$F(1, 543) = 1.40$	.24
Personality	699	−0.04	0.04	$F(1, 696) = 0.94$	.33	0.09	0.08	$F(1, 696) = 0.06$	.23
Problem behavior	744	0.39	0.04	$F(1, 741) = 125.13$	< .001	−0.18	0.07	$F(1, 741) = 6.45$	.01
Social skills	732	0.31	0.04	$F(1, 729) = 67.70$	< .001	0.08	0.07	$F(1, 729) = 1.27$	.26

Note: Preschool delay of gratification was dummy coded for marshmallow-test behavior (1 = delayed full 7 min, 0 = did not delay full 7 min).

Explanatory factors in longitudinal associations

Our first multilevel model (Table 4, Model 1) contained a variable for time at Level 1 but no variables at Level 2. Random effects for intercepts and slopes were significant, indicating substantial variance in the development of problem behavior from preschool through adolescence. All subsequent models were identical at Level 1. Model-specification details and coefficients for all fixed and random effects are shown in Tables S2 through S4 in the Supplemental Material.

Model 2 included preschool delay of gratification in the equation for the Level 2 intercept and explained approximately 3% of the variance in adolescent problem behavior. This model was conceptually similar to the multivariate regression model for adolescent problem behavior that controlled for preschool problem-behavior levels, but Model 2 provided a more robust adjustment by controlling for developmental trajectories in problem behavior rather than a single preschool observation. Comparisons revealed that Model 2 provided a significantly better fit to the data than Model 1 (Table 5), consistent with the significance of preschool delay of gratification in the multivariate regression model for problem behavior.

Model 3 included self-control in the equation for the Level 2 intercept, alongside preschool delay of gratification. In Model 4, self-control was exchanged for social support. These models provided a significantly better fit to the data than Model 2, which contained delay of gratification alone. This suggests that both self-control and social support constitute plausible explanations for the predictive validity of the marshmallow test. Notably, 30% of the variance in adolescent problem-behavior intercepts was explained by social support and delay of gratification in Model 4, compared with 17% explained by self-control and delay of gratification in Model 3. Thus, social support accounted for nearly twice as much variance as self-control.

Discussion

This preregistered secondary analysis supports the predictive validity of the preschool marshmallow test for life success, with significant bivariate longitudinal associations for three of five outcomes tested. Relationships between delay of gratification and problem behavior held across multiple models, and social support provided a more powerful explanation for these relationships than self-control. Differences between our study and an independent analysis illustrate how researchers'

Table 4. Variance Explained in Adolescent Problem Behavior

Model	Equation	Variance	Variance explained
1	$\pi_{0i} = \beta_{00} + \mu_{0i}$	74.64	
2	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \mu_{0i}$	73.61	2.7%
3	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{self-control}_i + \mu_{0i}$	61.84	17.1%
4	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{social support}_i + \mu_{0i}$	52.08	30.2%

Note: Variance refers to the intercept variance at Level 2, with the variance computed as a percentage of the total Level 2 intercept variance, $\widehat{\tau_{00}}(\text{Model 1}) - \widehat{\tau_{00}}(\text{Models 2 – 4}) / \widehat{\tau_{00}}(\text{Model 1})$ "Delayer" is a dummy variable coded as 1 for individuals who waited the full 7 min in the marshmallow test and 0 for individuals who did not.

Table 5. Nested Comparisons for Adolescent Problem-Behavior Intercept Models

Nested model	Augmented model	χ^2	*p*
$\pi_{0i} = \beta_{00} + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \mu_{0i}$	8.98	.003
$\pi_{0i} = \beta_{00} + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{self-control}_i + \mu_{0i}$	244.51	< .001
$\pi_{0i} = \beta_{00} + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{social support}_i + \mu_{0i}$	466.34	< .001
$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{self-control}_i + \mu_{0i}$	213.79	< .001
$\pi_{0i} = \beta_{00} + \beta_{01}\text{self-control}_i + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{self-control}_i + \mu_{0i}$	0.08	.779
$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{social support}_i + \mu_{0i}$	396.95	< .001
$\pi_{0i} = \beta_{00} + \beta_{01}\text{social support}_i + \mu_{0i}$	$\pi_{0i} = \beta_{00} + \beta_{01}\text{delayer}_i + \beta_{02}\text{social support}_i + \mu_{0i}$	50.19	.667

Note: The *p* values reflect a test of the difference in the deviance statistics for each pair of nested models, which is distributed as approximately χ^2 with 2 degrees of freedom. A *p* value less than .05 indicates that the augmented model provides a significantly better fit to the data than the nested model. "Delayer" is a dummy variable coded as 1 for individuals who waited the full 7 min in the marshmallow test and 0 for individuals who did not.

decisions shape results and conclusions even when they use the same data set (Silberzahn et al., 2018).

Comparisons with Watts et al.'s (2018) study

We replicated the majority of longitudinal associations tested, in contrast with Watts et al. (2018), whose findings are viewed as a failure to replicate the results of previous studies demonstrating the predictive validity of the preschool marshmallow test for adolescent outcomes.[1] Notably, we found significant relationships between preschool delay of gratification and adolescent problem behavior in bivariate, multivariate, and multilevel models, whereas Watts et al. found no relationships between delay of gratification and problem behavior.

Analytic decisions may explain this discrepancy. For example, we used the established, validated index of problem behavior from the CBCL, whereas Watts et al. (2018) created a "behavior composite" by averaging two subscales that included only five of eight syndromes. We used multilevel models containing multiple repeated observations of problem behavior over time, allowing us to estimate trajectories and partition within- and between-individuals variance, whereas Watts et al. estimated their models from fewer observations and did not differentiate the variance between levels. We analyzed the sample as a whole to maximize power, whereas Watts et al. focused on children of mothers without college degrees. Each set of decisions may be justifiable, but we used validated measures, more observations, and a larger sample and therefore may have obtained more robust estimates.

Even when results were similar, conclusions differed. Both our study and Watts et al.'s (2018) study found that significant relationships between preschool delay of gratification and adolescent outcomes diminished after including control variables. Watts et al. treated their more than 30 controls as confounding and extraneous, interpreting the resulting weakening of relationships as reflecting bias in the bivariate associations. Our approach to control variables was theory driven and minimalist, so we interpret weakening relationships in terms of explanations. For example, we hypothesized that the predictive validity of the marshmallow test derives from its capturing of social and cognitive factors that shape development. The weakening relationships between delay of gratification and problem behavior after we controlled for social support and self-control supports this hypothesis. Our approach provides insight into the classic findings, whereas Watts et al.'s interpretation implies that the findings were spurious.

Both approaches to using and interpreting control variables may be reasonable, but our approach is consistent with best practices (e.g., avoiding the use of control variables for purification purposes) and generates new, testable predictions (Greenhoot & Dowsett, 2012; McCall & Appelbaum, 1991; Spector & Brannick, 2011). Adding control variables will not necessarily improve causal inference and could lead to erroneous conclusions—for example, if estimation issues are neglected or if the controls inadvertently remove part of the effect of interest (Becker, 2005; Doebel, Michaelson, & Munakata, 2020; Rohrer, 2018; Spector & Brannick, 2011; Westfall & Yarkoni, 2016). Watts et al.'s (2018) extensive covariate adjustment divorced delaying gratification from its psychological and social antecedents, obscuring the meaning of residual associations (Doebel et al., 2020).

A new understanding of the marshmallow test

The marshmallow test captures characteristics and circumstances of early childhood essential for life success.

Our study provides the first evidence that the test may be predictive because it reflects social factors. Children who grow up in supportive environments might be more willing to delay gratification because of a greater trust that events will unfold as expected (Kidd et al., 2013; Michaelson & Munakata, 2016) and because of norms around future-oriented decisions (Doebel & Munakata, 2018; Lamm et al., 2018; Pepper & Nettle, 2017). These experiences with delaying gratification may in turn lead to greater willingness and ability to delay by providing opportunities for children to practice delaying and experience its benefits (Doebel, Michaelson, & Munakata, 2017). Children in supportive environments may thus increasingly delay gratification to foster behavioral, social, and academic success across development. Our findings support this new perspective by demonstrating that social factors can explain why a simple marshmallow test predicts life success.

This account departs from prominent theories and intervention approaches. Most delay-of-gratification interventions target cognitive mechanisms (e.g., Murray, Theakston, & Wells, 2016; Peake, Hebl, & Mischel, 2002). Our results suggest that interventions might instead target social factors. Supporting parents and teachers in developing norms around delaying gratification and providing children with experiences of delayed rewards being delivered may increase children's willingness to wait and may support children's learning about the value of future-oriented decisions. A community-trust intervention with adults supported the effectiveness of such approaches (Jachimowicz et al., 2017), but these approaches have not been tested with children.

Limitations and future directions

To further test the replicability of the marshmallow-test findings and possible explanations, researchers should explore different measures in future studies. We created composite variables from multiple measures identified through a systematic search-and-selection process. This had the advantage of minimizing our data-informed analytic decisions, but creating latent factors would have been more effective in correcting for potential measurement error. Additionally, like most researchers, we cannot rule out potential differences in the quality of our composites. However, we used only ratified variables created by NICHD investigators in our analysis. Of the 38 reliability estimates reported in the SECCYD documentation from measures we included, 28 were highly reliable ($\alpha > .80$), and 8 of the remaining 10 were adequate (αs = .69–.79). Less reliable measures may have limited our power to detect significant relationships, but they were always combined with other highly reliable measures, and all within-composite correlations between measures were significant. Furthermore, problems caused by bias due to uncorrected measurement error are less likely in models with few variables and paths such as ours (Cole & Preacher, 2014). Thus, measurement error is unlikely to have affected our substantive conclusions.

To guide intervention development in future studies, researchers should examine the timing of and relationships among developmental processes linking social support and delay of gratification to later outcomes. This will help determine moderator and mediator relationships. We focused on preschool delay, given prior evidence that the marshmallow test is predictive, and we aggregated data across repeated observations measured within individuals for both social support and self-control to match the time span of the longitudinal associations of interest. Utilizing more waves of data with higher temporal resolution could help determine whether protective effects of early social support on later outcomes are mediated by delaying gratification. This, in turn, could inform whether interventions focused on social support should also improve delaying gratification to produce benefits. Including additional theoretically informed control variables could address alternative causal explanations and target interventions—for example, by controlling for within-individuals changes in potential confounds and testing moderators of the effects of social support and delay of gratification.

General conclusions

Our results support the validity of the preschool marshmallow test as an early indicator of academic, social, and behavioral functioning in adolescence. The relative importance of social support over self-control for understanding this predictive validity suggests novel interpretations and points to new directions for intervention.

Transparency

Action Editor: D. Stephen Lindsay
Editor: D. Stephen Lindsay
Author Contributions

L. E. Michaelson and Y. Munakata conceived and designed the study. L. E. Michaelson obtained access to the data and performed the analysis. Both authors wrote the manuscript and approved the final version for submission.

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 National Institutes of Health Grant No. R01 HD078532.

Open Practices

Access to the archival data analyzed in this study is restricted, but data are available on request from the Inter-University Consortium for Political and Social Research (ICPSR). All of our analysis scripts have been made publicly available on the Open Science Framework at https://osf.io/j7hzt/. Our hypotheses, design, method, and analysis plan were preregistered on the Open Science Framework prior to data analysis (https://osf.io/vjmkz). The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619896270. This article has received the badges for Open Data and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Laura E. Michaelson https://orcid.org/0000-0003-0321-025X

Acknowledgments

Part of this work was completed while both authors were affiliated with the Department of Psychology and Neuroscience at the University of Colorado Boulder.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619896270

Note

1. Our study and that of Watts et al. (2018) both found significant bivariate relationships between delay of gratification in the marshmallow test and adolescent academic achievement, although the relationships were smaller than in the original studies. We also found significant relationships between delay of gratification and adolescent social skills; Watts et al. did not test this outcome. Duckworth et al. (2013) also replicated associations between delay of gratification and both adolescent academic achievement and other adolescent outcomes, such as body mass index.

References

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

Arantes, J., Berg, M. E., Lawlor, D., & Grace, R. C. (2013). Offenders have higher delay-discounting rates than non-offenders after controlling for differences in drug and alcohol abuse. *Legal and Criminological Psychology*, *18*, 240–253. doi:10.1111/j.2044-8333.2012.02052.x

Ayduk, Ö., Zayas, V., Downey, G., Cole, A. B., Shoda, Y., & Mischel, W. (2008). Rejection sensitivity and executive control: Joint predictors of borderline personality features. *Journal of Research in Personality*, *42*, 151–168. doi:10.1016/j.jrp.2007.04.002

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*, 274–289.

Bickel, W. K., Koffarnus, M. N., Moody, L., & Wilson, A. G. (2014). The behavioral- and neuro-economic process of temporal discounting: A candidate behavioral marker of addiction. *Neuropharmacology*, *76*, 518–527.

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

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*, 300–315. doi:10.1037/a0033805

Doebel, S., Michaelson, L. E., & Munakata, Y. (2017). Beyond personal control: The role of developing self-control abilities in the behavioral constellation of deprivation. *Behavioral & Brain Sciences*, *40*, Article e324.

Doebel, S., Michaelson, L. E., & Munakata, Y. (2020). Good things come to those who wait: Delaying gratification likely does matter for later achievement (a commentary on Watts, Duncan, & Quan, 2018). *Psychological Science*, *31*, 97–99. doi:10.1177/0956797619839045

Doebel, S., & Munakata, Y. (2018). Group influences on engaging self-control: Children delay gratification and value it more when their in-group delays and their out-group doesn't. *Psychological Science*, *29*, 738–748.

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. doi:10.1177/0146167213482589

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*, 1410–1430.

Greenhoot, A. F., & Dowsett, C. J. (2012). Secondary data analysis: An important tool for addressing developmental questions. *Journal of Cognition and Development*, *13*(1), 2–18. doi:10.1080/15248372.2012.646613

Jachimowicz, J. M., Chafik, S., Munrat, S., Prabhu, J. C., & Weber, E. U. (2017). Community trust reduces myopic decisions of low-income individuals. *Proceedings of the National Academy of Sciences, USA*, *114*, 5401–5406. doi:10.1073/PNAS.1617395114

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. doi:10.1016/j.cognition.2012.08.004

Lamm, B., Keller, H., Teiser, J., Gudi, H., Yovsi, R. D., Freitag, C., . . . Lohaus, A. (2018). Waiting for the second treat: Developing culture-specific modes of self-regulation.

Child Development, *89*, e261–e277. doi:10.1111/cdev.12847

Lee, C. A., Derefinko, K. J., Milich, R., Lynam, D. R., & DeWall, C. N. (2017). Longitudinal and reciprocal relations between delay discounting and crime. *Personality and Individual Differences*, *111*, 193–198.

Lee, W. S., & Carlson, S. M. (2015). Knowing when to be "rational": Flexible economic decision making and executive function in preschool children. *Child Development*, *86*, 1434–1448. doi:10.1111/cdev.12401

McCall, R. B., & Appelbaum, M. I. (1991). Some issues of conducting secondary analyses. *Developmental Psychology*, *27*, 911–917. doi:10.1037/0012-1649.27.6.911

McGuire, J. T., & Kable, J. W. (2012). Decision makers calibrate behavioral persistence on the basis of time-interval experience. *Cognition*, *29*, 997–1003. doi:10.1016/j.biotechadv.2011.08.021

McGuire, J. T., & Kable, J. W. (2013). Rational temporal predictions can underlie apparent failures to delay gratification. *Psychological Review, 120*, 395–410. doi:10.1037/a0031910

Michaelson, L. E., de la Vega, A., Chatham, C. H., & Munakata, Y. (2013). Delaying gratification depends on social trust. *Frontiers in Psychology*, *4*, Article 355. doi:10.3389/fpsyg.2013.00355

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. doi:10.1111/desc.12388

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., . . . Shoda, Y. (2011). '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.

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.

Peake, P. K., Hebl, M., & Mischel, W. (2002). Strategic attention deployment for delay of gratification in working and waiting situations. *Developmental Psychology*, *38*, 313–326.

Pepper, G. V., & Nettle, D. (2017). The behavioural constellation of deprivation: Causes and consequences. *Behavioral & Brain Sciences*, *40*, Article e314. doi:10.1017/S0140525X1600234X

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*, 27–42.

Schneider, S., Peters, J., Peth, J. M., & Büchel, C. (2014). Parental inconsistency, impulsive choice and neural value representations in healthy adolescents. *Translational Psychiatry*, *4*, Article e382. doi:10.1038/tp.2014.20

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.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*, 337–356.

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*, 287–305.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*, 1159–1177.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE*, *11*(3), Article e0152719. doi:10.1371/journal.pone.0152719

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, *70*, 533–552.

Yi, R., Carter, A. E., & Landes, R. D. (2012). Restricted psychological horizon in active methamphetamine users: Future, past, probability, and social discounting. *Behavioural Pharmacology*, *23*, 358–366. doi:10.1097/FBP.0b013e3283564e11