Evaluating the Impact of Preschool on Patience, Time Inconsistency and Commitment Demand

Amanda Chuan, Anya Samek, and Shreemayi Samujjwala*

January 23, 2024

Abstract

We evaluate patience, time inconsistency and commitment demand among children. We first show that patience at ages 5-10 predicts reading scores up to 4 years later, even after controlling for prior cognition and executive function. Time inconsistency and commitment demand do not predict reading scores. Second, we evaluate whether preschool affects patience. We leverage a field experiment that randomized children to different preschool curricula. We find that the preschool curriculum focused on self-regulation improved patience overall relative to the control group. Further, preschool helps children use commitment devices to manage their time inconsistency, but does not affect time inconsistency directly.

Keywords: preschool, human capital, time preferences, patience, inconsistency
JEL Classifications: I21, I24, J24, I26

*Chuan: Michigan State University. 368 Farm Lane S435, East Lansing, MI 48824. achuan@msu.edu, 517-353-1696. Samek: University of California at San Diego Rady School of Management. 9500 Gilman Drive 0553, La Jolla, CA 92093. asamek@ucsd.edu, 765-498-9425. Samujjwala: University of Pennsylvania. 133 South 36th Street, Philadelphia, PA 19104. ssamujjwala@gmail.com. Declarations of interest: none. Deidentified data will be made available to the public.

1 Introduction

While the traditional view of preschool focuses on teaching the ABCs and the 123s, a growing literature has identified the importance of preschool in shaping higher-order cognition and executive function (Heckman, 2008; Heckman and Carneiro, 2003).[1] In response, there are now preschool curricula focused on improving such skills (Jones and Bouffard, 2012). For example, the *Tools of the Mind* curriculum aims to develop self-regulation and executive functioning skills (i.e., inhibitory control, working memory and cognitive flexibility) through activities like Play Planning, which encourages children to develop independent plans for their activities (Bodrova and Leong, 2006). Experimental studies find that *Tools of the Mind* improves self-regulation, executive functioning skills and academic skills (Blair and Raver, 2014; Diamond et al., 2007; Diamond et al., 2019). More generally, preschool programs are shown to improve academic skills and socio-emotional skills (Hahn et al., 2016; Ramon et al., 2018).

We propose that preschool may also influence time preferences. Understanding the influence of preschool on time preferences is important since related work shows that time preferences predict academic attainment and health (Castillo et al., 2018; Khwaja et al., 2006; Mørkbak et al., 2017). Activities in preschool, such as sitting still for story time or taking turns playing with toys, may build patience, or the willingness to delay rewards. Further, Play Planning and related activities in programs that focus on self-regulation like *Tools of the Mind* give children practice with making and following through on plans. This may in turn reduce time inconsistency, a feature of some time preference models that explains why people tend to make patient plans but fail to follow through in their actions (Herrnstein et al., 1993; Laibson, 1997; O'Donoghue and Rabin, 1999, 2006; Strotz, 1955; Thaler and Shefrin, 1981).

In this paper, we evaluate the importance of early childhood time preferences on later life outcomes, and measure the causal impact of preschool on time preferences. To do so, we conduct follow-up experiments with 388 children who participated in the Chicago Heights Early Childhood Center (CHECC) field experiment. When the children were 3-4 years old, they enrolled in the Chicago Heights Early Childhood Center (CHECC) and were randomized to a preschool for 1-2

[1] Additional work in this area includes Belfield et al. (2006), Campbell et al. (2001), Campbell et al. (2002), Cappelen et al. (2020), Castillo et al. (2020), Fryer et al. (2015), Heckman et al. (2010b), and Heckman et al. (2006), Masse and Barnett (2002), Nores et al. (2005).

years or to a control group that did not receive preschool (Fryer et al., 2015; Fryer et al., 2020). The preschools varied in their focus. One used the *Tools of the Mind* curriculum, which specialized in building self-regulation and executive function skills; another used *Literacy Express and Math*, which focused on early literacy and math skills; the third used *CogX*, which combined instruction on literacy, math, and executive function.[2]

When children were 5-10 years old, they participated in a two-week task that elicited multiple dimensions of time preferences, including patience, time inconsistency and commitment demand. We measure these elements because we believe they play necessary and distinct roles in children's developmental trajectories.[3] A child who is patient may prefer to excel in school instead of watching TV. However, in order to excel in school, the child must also exhibit time consistency by repeatedly choosing to study even when watching TV provides her with immediate gratification. That is, she must make and follow through with patient plans. Individuals who foresee their future selves struggling to delay gratification may value a commitment device that helps them stick to patient plans (O'Donoghue and Rabin, 1999). To test if this sophistication emerges at early ages, we offer a commitment device and measure its take-up among our sample.

Given these considerations, our task is designed to investigate decisions over time, rather than focusing on one-shot choices in isolation. We give children the opportunity to make patient or impatient choices over two weeks, since *stability of patience* – i.e., making a patient decision in multiple periods – is an important but overlooked component of time preferences. Prior experimental literature relies on one-shot decisions to elicit patient preferences, implicitly assuming that patience is a stable trait. We find that patience is not necessarily stable,[4] and that stability is itself an object of interest. Our results show that stability of patience predicts children's later achievement in ways that one-shot measures of patience do not.

[2] *Tools of the Mind* is a commercially available curriculum developed by Drs. Elena Bodrova and Deborah Leong (Bodrova and Leong, 2006). For more information, see `https://toolsofthemind.org`. *Literacy Express* is a commercially available curriculum developed by Drs. Christopher Lonigan, Jeanine Clancy-Menchetti, Beth Phillips, and colleagues ("Literacy Express", 2010). For more information, see `https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_lit_express_072710.pdf`. *CogX* is not a commercially marketed product, but its components are available by request from the authors of this paper (Fryer et al., 2020).

[3] More broadly, different measures of time preferences may measure different aspects of behavior. For example, Castillo et al. (2020) show that patience as measured by an incentivized economics experiment loads onto a different factor compared to measures of self-control and delay-of-gratification from psychology.

[4] Our result that patience is not necessarily stable over time is consistent with Meier and Sprenger (2015), who explore temporal instability among low and middle income adults in Boston.

Our task elicits time preferences using consumption decisions (Augenblick et al., 2015; Augenblick and Rabin, 2019; Sadoff et al., 2020). On Day 1 of Week 1, children choose between a smaller, earlier reward on Day 2, or a larger, later reward on Day 3. We refer to this as the "Advance" Decision. We return to the children on Day 2 and surprise them with the opportunity to change their mind. Therefore, on Day 2 children choose again between the smaller, earlier reward given out the same day, or the later, larger reward given out the next day, Day 3. We refer to this as the "Immediate" Decision. We identify patience by the willingness to choose the larger, later reward in the Advance Decision. We identify time inconsistency based on whether children switch their choice between the Advance and Immediate Decisions.

Week 2 is designed to evaluate the stability of time preferences across the two weeks and assess children's demand for commitment. In Week 2, children again make Advance and Immediate Decisions. We classify children who make the same choice in the Advance Decision in Week 1 and Week 2 as having stable preferences, and children who select different options as having unstable preferences. After making their Advance Decision on Day 1, children decide whether to commit to their Advance Decision or to retain the option to change their mind the next day.

It is important to emphasize that our focus is not on initial choices. It is easy for children to behave patiently in the initial choice – indeed, the vast majority of them (80%) do. Rather, our primary aim is a holistic investigation of decisions over the course of the two week task. We find that patience drops to 75% in Week 2. About 20% of children are inconsistent, and 50-75% are impulsive - they choose the later reward in the Advance Decision and switch to the earlier reward in the Immediate Decision. About half of the children take up the commitment device and restrict themselves to their Advance Decision in Week 2. Cognitive and executive functioning skills measured contemporaneously explain little of children's choices in the time preference elicitation task.

We first investigate the importance of time preferences for human capital development by correlating our measures of time preferences at ages 5-10 with reading and math test scores up to 5 years later. We find that patience is correlated with future reading test scores, even when controlling for cognitive and executive function skills taken prior to the experiments. This is consistent with prior work using CHECC data that shows that patience predicts disciplinary referrals in elementary school (Castillo et al., 2020), as well as work with other samples which link early self-regulation and

literacy up to second grade (Skibbe et al., 2019). It joins Benjamin et al. (2020), who investigate how delaying gratification during preschool predicts midlife wealth outcomes.

Second, we investigate the causal impact of preschool on time preferences. We find no treatment group differences in patience or time inconsistency in Week 1. However, compared to Week 1, in Week 2 children randomized to *Tools of the Mind* remain patient while children randomized to the control group become significantly less patient. In other words, *Tools of the Mind* children display significantly lower *instability in patience* than control children. Children in the other two preschool programs – *Literacy Express* and *CogX* – behave similarly to the control group. Further, the *Tools of the Mind* group exhibits sophisticated use of the commitment device whereas the other groups do not. That is, *Tools of the Mind* children are more likely to take up commitment in Week 2 if they were impulsive in Week 1.

Finally, we evaluate potential explanations for the difference in patience. First, children may not make the patient choice in Week 2 if they were impulsive in Week 1 and learned from their past behavior. We do not find evidence for this explanation. Second, children may adjust their behavior in Week 2 after observing the behavior of peers in Week 1. We find evidence for this second explanation: children in the control group behaved more impatiently if they happened to be assigned to a classroom with more impatient classmates, whereas children in the *Tools of the Mind* program do not.

Our paper makes contributions along several dimensions. We are the first to document that early childhood education can change time preferences up to five years later. Andreoni et al. (2019) and Castillo et al. (2020) both examine patience among children in the CHECC sample, but find suggestive evidence at best.[5] The difference in our results arises because we use a different manipulation to separate patience from time inconsistency, impulsivity, and commitment demand.[6] Furthermore, we elicit these measures over two weeks, yielding more stable measures beyond the one-shot elicitations used by Andreoni et al. (2019), Castillo et al. (2020), and the rest of the literature on children.

[5]Other papers that use the CHECC study have explored the impact of programs on cognitive ability and executive function Fryer et al. (2015), Fryer et al. (2020), behaviors regarding fairness Cappelen et al. (2020) and Chuan et al. (2021), reading versus math performance Chuan et al. (2022), cheating behavior Houser et al. (2015), charitable giving Ben-Ner et al. (2017), Cowell et al. (2015), Cox et al. (2016), List and Samak (2013), List et al. (2018), and Samek and Sheremeta (2017), competitiveness Samak (2013), and food choice Sadoff and Samek (2019).

[6]However, our patience measure predicts the patience measures used by Andreoni et al. (2019) well, even though we designed it to measure a greater range of time preferences than patience alone.

We can therefore speak directly to whether future skill formation depends on patience, impulsivity, or both. Seminal work involving the "marshmallow test" argues for the importance of early childhood time preferences on future adult success, but does not experimentally distinguish between whether it is patience or impulse control that matters (Mischel et al., 1972; Mischel et al., 1988; Mischel et al., 1989; Shoda et al., 1990). Our task first elicits patience in the Advance Decision and then impulsivity in whether the Immediate Decision matches the Advance Decision. This design allows us to compare the importance of patience versus impulsivity on future skill formation. We find that patience significantly predicts future reading, while impulsivity, inconsistency, and commitment demand do not.

Our paper is also similar to Alan and Ertac (2015, 2018) and Alan et al. (2021), who evaluate an 8-week intervention focused on teaching waiting strategies to third and fourth-graders, and Lührmann et al. (2018), who find that financial literacy training reduces the time inconsistency of adolescents. However, these are short interventions that narrowly target time preferences, whereas we still know little about the more general curricula experienced by most children in preschool. We show that preschool programs that teach a holistic set of skills can change time preferences among children as early as 3-5 years old.

Lastly, we contribute to the policy debate on the benefits of preschool. Evaluations of the Perry Preschool Program, Abecedarian Project, and Head Start show that preschool improves long-term outcomes among low socio-economic status (SES) children (Anders et al., 2023; Belfield et al., 2006; Campbell and Ramey, 1994; Carneiro and Ginja, 2014; Currie and Thomas, 1995; Garcıéa et al., 2020; Heckman et al., 2010a; Heckman, 2008; Heckman et al., 2010b; Nores et al., 2005). Importantly, these long-term gains are unlikely to come from IQ alone (Heckman, 2008), suggesting that preschool benefits can come through building soft skills. Our results indicate that time preference formation may be one such skill, which can aid in the development of further skills. Improving patience may thus equip students with the tools to invest in their own human capital in later years.

Furthermore, our sample consists mainly of minority, low-SES children. Understanding how different curricula affect time preferences is important since studies show that low-SES children are more impatient and impulsive than high-SES counterparts (Delaney and Doyle, 2012; Falk et al., 2021). Such differences may contribute to later gaps in skill acquisition and wealth during

adulthood. Our paper provides evidence that preschool may narrow these adult gaps through narrowing the gap in time preferences during childhood.

In what follows, Section 2 describes our experimental design. Section 3 characterizes the time preferences of the children. Section 4 describes the treatment effects and our interpretation of them. Section 5 concludes.

2 Experimental Design and Procedures

2.1 Chicago Heights Early Childhood Center

The children in our experiment are part of the Chicago Heights Early Childhood Center (CHECC). CHECC was a randomized study implemented in 2010-14 in Chicago Heights, Illinois that focused on investigating types of early childhood investments that can reduce the academic achievement gap. In Chicago Heights, 29% of residents live below the poverty line and median income is $43,102. Households in the suburb are 41.5% Black and 33.9% Hispanic or Latino.[7]

CHECC recruited 3-4 year old children annually for 4 years, generating 4 cohorts of participants. Upon enrolling, children were randomized to preschool or control for 1-2 years.[8] All children participated in baseline, endline and follow-up assessments measuring cognitive skills (reading, writing and math) and executive functions (inhibitory control, working memory and attention shifting).[9] The preschool programs were free, full day programs that differed in curricula, as follows:[10]

- Preschool: *Tools of the Mind*, Cohorts 1 and 2: Used the *Tools of the Mind* curriculum with a strong focus on executive functioning skills (Bodrova and Leong, 2006).

[7] U.S. Census Bureau Quick Facts: `http://quickfacts.census.gov/qfd/states/17/1714026.html`.

[8] Two other programs were also part of CHECC: Parent Academy included classes for parents but did not include a child intervention (see Fryer et al., 2015), while Kinderprep included a shorter, 2-month long preschool program utilizing the CogX curriculum. Due to budget constraints, we made the difficult choice of focusing only on the control and Preschool groups when we collected the time preference data described in this paper.

[9] These tests include the Woodcock-Johnson III Test of Achievement (Woodcock, McGrew and Mather, 2001) which comprises of: (i) Applied Problems (WJA), (ii) Quantitative Concepts (WJQ), (iii) WJ-Letter and Word Identification (WJL) and (iv) Spelling (WJS), the Peabody Picture Vocabulary Test III which measures receptive vocabulary (Dunn et al., 1965) and the Clancy and Blair (2006a, 2006b) tests of executive functioning.

[10] We compare between each preschool group and the control group based on children's random assignment to experimental condition, since using attendance would introduce selection into treatment compliance based on parental motivation. We therefore take an intent-to-treat approach.

- Preschool: *Literacy Express and Math*, Cohorts 1 and 2: Used the *Literacy Express* curriculum that focused on building literacy skills (Lonigan et al., 2011; Lonigan et al., 2006), supplemented with math instruction.

- Preschool: *CogX*, Cohorts 3 and 4: Used a new curriculum combining elements of executive functioning and literacy, half-day instruction and half-day daycare. It included a parental component where parents met with teachers bi-weekly to learn how to scaffold their child's learning at home.

- Control, Cohorts 1-4: Did not receive any CHECC program features.

Children in Cohorts 1 and 2 were randomized to *Tools of the Mind* or *Literacy Express* conditional on receiving preschool assignment. Instruction for the two curricula occurred on different campuses. Cohorts 3 and 4 all received the *CogX* curriculum if they received preschool assignment. The control group only participated in mid-year and end-of-year assessments. They were invited to 2-3 CHECC parties per year (e.g., the holiday party and Easter party). They did not receive any preschool instruction, information on parenting, or other program features. It is unlikely that control group children could receive the type of quality instruction provided by CHECC from other sources. Chicago Heights is an underserved community, and alternative resources for early childhood education could not match CHECC's estimated value of $7,000 per child per year. Moreover, parents had limited resources in securing private child care or education for their preschool-aged children.

At age 5, children enrolled in kindergarten in District 170 or in surrounding districts. Children from all treatment arms were assigned to classrooms independent of prior CHECC treatment, meaning that children in the CHECC control and preschool groups often ended up in the same class. We had an agreement with District 170 to continue conducting assessments during school hours. As such, our sample consists of all children who were in kindergarten through fourth grade in District 170 elementary schools in 2015.

There are benefits and drawbacks to assessing children during school hours. One benefit is that it reduces the attrition that may arise if we asked parents to voluntarily bring their children for an assessment. This type of attrition could be problematic if parents of treatment kids have stronger social ties with CHECC than parents of control kids, which is correlated with willingness

to participate in our time preference task and other unobservable characteristics. A drawback is that we only assess District 170 children, so our analysis sample captures 23.3% of the full CHECC sample.[11] Because children typically attend school based on neighborhood of residence, the sample in this paper differs demographically relative to the overall sample. To account for the possibility that CHECC assignment influenced the likelihood of remaining in District 170, in robustness checks we inverse weight for the probability of entering District 170 conditional on CHECC assignment.

2.2 Time Preference Task

The time preference task was conducted at the beginning of the 2015 school year. It is based on the task from Sadoff et al. (2020) but is adapted for children. It lasted two weeks, with two decision-making days in each week (see Table 1). We conducted the task during the last 15-20 minutes of the school day so that children could receive their incentives right away. We pulled children from the CHECC control and treatment groups out of the classroom and asked them to complete the task in small groups. Experimenters (blind to CHECC treatment assignment) read the instructions out loud to the group, and children completed a quiz on understanding. Children marked their choices in private on an activity sheet (see Appendix B).

A concern with time preference elicitation is that researchers can measure choice but usually cannot measure consumption. If consumption does not immediately follow choice, then it may be difficult to measure time inconsistency since the earlier choice may not be sufficiently tempting.[12] We used glow-in-the-dark bracelets as our incentive. Glow-in-the-dark bracelets overcome this limitation by fading within 24 hours of receipt, meaning that children must consume them quickly.[13] We pre-tested the appropriateness of glow bracelets in a survey and found that children preferred them over other incentives; further, there were no significant differences in preference for them by age or gender.[14]

[11] Out of the 1,664 children who were in a CHECC control or preschool group, 388 completed at least one decision in our glow bracelet task. The breakdowns per CHECC group are: 29.3% (39/133) of all children in the *Tools of the Mind* curriculum; 28.7% (39/136) of all children in the *Literacy Express* curriculum; 22.4% (59/263) of all children in the *CogX* curriculum; and 22.2% (251/1132) of all children in the control group.

[12] This is a drawback of experiments that use time-dated monetary rewards (for reviews, see Sprenger, 2015 and Cohen et al., 2020). Since money is not consumed at the time of receipt, researchers have attempted to use non-monetary measures such as choices over perishable foods (Sadoff et al., 2020).

[13] Children were reminded about this feature of the bracelets by the experimenters on each day, so that they would not have the expectation of hoarding them across weeks. When we gave children their bracelets, we cracked the bracelet to activate its glow immediately beforehand.

[14] 79 children ages 5-10 years old participated in a survey in which they were asked to rate their preferences for

Table 1 summarizes the procedure, while Table 2 relates how each time preference is measured by the task. On Day 1, we measured initial **patience**. Children chose between receiving one bracelet on Day 2 versus two bracelets on Day 3. We refer to the choice made on Day 1 as the "Advance Decision" since it is made with a front-end delay. We classify children as patient if they chose the larger, later reward and impatient if they chose the smaller, earlier reward. Importantly, the goal of our study is to examine how choices evolve over time, rather than the initial choices themselves. We therefore designed the initial choice to be an easy decision, so that most children would make the patient decision. This leaves us with room to test whether children change their patient choices later.

Table 1: Time Preference Task Design and Procedures

	Week 1		
	Day 1	**Day 2**	**Day 3**
Decisions	* *Advance Decision:* Choose between 1 reward on Day 2 or 2 rewards on Day 3	* *Immediate Decision:* Can change Day 1 decision	
Rewards		* Receive reward if chosen for Day 2	* Receive reward if chosen for Day 3
	Week 2		
	Day 1	**Day 2**	**Day 3**
Decisions	* *Advance Decision:* Choose between 1 reward on Day 2 or 2 rewards on Day 3 **Commitment Decision:* Commit to Advance Decision as final decision	* *Immediate Decision:* Can change Day 1 decision, if didn't commit	
Rewards		* Receive reward if chosen for Day 2	* Receive reward if chosen for Day 3

Notes: Summary of events on each week and each day of the time preference task.

On Day 2, we surprised children with the option to change their mind. This allows us to measure inconsistency and impulsivity. That is, after viewing their activity sheet from the previous day, children were given a new activity sheet and selected again between two bracelets on Day 3 versus one bracelet on Day 2. We operationalized the re-allocation in this way to ensure equal

glow bracelets, temporary tattoos and similar prizes. We chose glow bracelets because they were relatively well-liked, and their likeability did not vary by age or gender (unlike e.g., glow sticks, which were favored by younger children, and glow necklaces, which were favored by girls).

effort between remaining with the previous choice versus switching to a new choice. Additionally, this protected the privacy of decisions, as every child received and returned an activity sheet, regardless of decision made. We refer to the decision as the "Immediate Decision", since children could receive one bracelet immediately if they chose the earlier option. We classify children as **time inconsistent** if their choice differed between the Advance and Immediate Decision, and **impulsive** if they chose the later option in the Advance Decision and the earlier option in the Immediate Decision. Assuming that children are attentive,[15] we take impulsive behavior to be indicative of present-biased preferences.

Table 2: Time Preference Measures

A: Measures within week

Patience	**Inconsistency**	**Impulsivity**	**Commitment**
Choosing 2 rewards on Day 3 over 1 reward on Day 2 in Advance Decision	Changing between Advance and Immediate Decision	Changing from later option in Advance Decision to earlier option in Immediate Decision	Committing to Advance Decision as final decision

B: Stability of time preferences, measures across weeks

Stable in patience	**Stable in impatience**	**Unstable in patience**	**Unstable in impatience**
Choose later option in both weeks	Choose earlier option in both weeks	Choose later option in Week 1, then switching to earlier option in Week 2	Choose earlier option in Week 1, then switching to later option in Week 2

Note: Summary of how time preferences are measured by the glow bracelet task. Panel A explains how the task measures patience, inconsistency, impulsivity, and commitment take-up, which are measures taken within each week. Panel B explains how the task measures the change in time preferences across weeks: stability of patience, stability of impatience, instability of patience, and instability of impatience.

Week 2 was designed to measure stability of preferences and to elicit commitment demand. It posed the same Advance and Immediate Decisions as in Week 1. However, after making the Advance Decision on Day 1 of Week 2, children chose between committing to their decision, or keeping open the possibility of changing their mind in the Immediate Decision the next day.[16] We operationalized the **commitment** decision by asking children to state whether they would like their current decision sheet to be the final decision, or whether they would like to make their final decision the next day on a new sheet.

[15] The experiment involves pre-tests for comprehension.

[16] We presented the commitment device after the Advance Decision so that the commitment device could only affect inconsistency and impulsivity in Week 2, but not patience.

As Table 2 panel B describes, we classify children as being **stable in patience** if they choose the later option in both weeks and **stable in impatience** if they choose the earlier option in both weeks. Children are **unstable in patience** if they chose the later option in Week 1 and switched to the earlier option in Week 2. They are **unstable in impatience** if they chose the earlier option in Week 1 and switched to the later option in Week 2.

One concern is that children could possess diminishing marginal utility over the glow bracelets. This is a problem with any intertemporal preference elicitation where children make repeated choices, no matter the good. Glow bracelets are advantageous in that they diminish this concern, since their appeal expires after 24 hours. Nevertheless, it is still possible that children's valuations of the bracelets could diminish in Week 2 after they received the bracelets in Week 1. In this case, our setup allows us to detect a lower bound on "stable patience". We will show that this stable patience significantly differs between preschool and control, suggesting that preschool changes how children make intertemporal choices.

Another possibility is that children may observe bracelets worn by other children or exchange bracelets with friends in Week 1, which will affect how they choose bracelets in Week 2. If so, we would conclude that peer effects influence time preferences just as it influences risk preferences, occupational preferences, and other important domains (Ahern et al., 2014; Browne et al., 2021; Eckel et al., 2012; Gaviria and Raphael, 2001; Lundborg, 2006; Mani and Mullin, 2004; Sacerdote, 2001). We return to this point in Section 4.2.

2.3 Sample Characteristics

Table 3 column (1) presents the summary statistics of the full sample. During the bracelet task, children were 7-8 years old on average, although ages ranged from 5 to 10 years. About half of our sample is female. Our sample is 76% Hispanic, 24% Black, and 3% White.[17] At the baseline assessment, children's cognitive skills were in the 27th percentile adjusted for age, while the average percent correct on executive function tasks was 45%. We split household income into three categories. Almost a third of our sample made less than $25,000 per year (32%), 13% of our sample earned $26,000-60,000, and only 1% of our sample earned over $60,000 per year. The

[17] These proportions do not add up to 1, since families can identify as more than one race. For example, 11 children are both Hispanic and Black.

remaining families did not report annual income.

We used a lottery to randomize children to preschool or control, but children's assignment could influence their likelihood of staying in District 170 and participating in the glow bracelet task. We examine differential attrition by testing for joint orthogonality of covariates with preschool assignment in columns (2)-(4).[18] This is preferred over t-tests comparing characteristics by treatment

Table 3: Regression of Treatment on Demographics

	Summary Statistics	Logit Regression (coefficients)		
	All	Tools	Literacy	CogX
	(1)	(2)	(3)	(4)
Child age	7.726	0.618*	0.327	-0.146
	(0.056)	(0.321)	(0.341)	(0.443)
Female	0.490	0.385	0.537	0.301
	(0.025)	(0.444)	(0.479)	(0.402)
Black	0.240	-1.920	-2.363	1.373*
	(0.022)	(1.312)	(1.479)	(0.822)
Hispanic	0.758	-1.578	-1.809	1.391
	(0.022)	(1.110)	(1.311)	(0.958)
White	0.029	0.000	0.000	0.888
	(0.009)	(0.000)	(0.000)	(1.752)
Baseline cognitive	27.354	0.013	0.036**	-0.008
	(1.006)	(0.016)	(0.016)	(0.014)
Baseline executive function	45.391	-0.007	-0.005	0.013
	(1.511)	(0.011)	(0.012)	(0.009)
% classmates with 2 bracelets in Week 1	0.696	-0.022	-0.485	0.070
	(0.016)	(0.898)	(0.871)	(0.792)
Income				
0-25K	0.322	-0.119	-0.392	0.555
	(0.024)	(0.564)	(0.568)	(0.453)
26-60K	0.126	0.992	0.194	-0.310
	(0.017)	(0.684)	(0.835)	(0.702)
61K or more	0.010	-0.265	0.000	0.000
	(0.005)	(1.665)	(0.000)	(0.000)
School				
Garfield	0.206	0.683	2.167*	0.924
	(0.021)	(0.963)	(1.176)	(0.802)
Grant	0.121	0.921	0.116	-0.327
	(0.017)	(0.885)	(1.404)	(0.917)
Greenbriar	0.085	0.444	0.361	0.827
	(0.014)	(1.037)	(1.407)	(0.989)
Jefferson	0.064	0.006	0.430	0.687
	(0.012)	(1.327)	(1.539)	(0.945)
Lincoln	0.041	2.107	0.000	-0.243
	(0.010)	(1.295)	(0.000)	(1.395)
Wilson	0.067	0.364	2.849**	0.235
	(0.013)	(1.366)	(1.346)	(0.965)
Kennedy	0.095	1.560	1.748	0.760
	(0.015)	(1.053)	(1.304)	(0.938)
Roosevelt	0.206	1.286	1.582	0.663
	(0.021)	(0.857)	(1.177)	(0.737)
Observations	388	158	146	133
Joint χ^2 test p-value		0.448	0.357	0.894

Notes: Column (1) shows summary statistics for the full sample. Columns (2)-(4) show logit regression of treatment vs. control on predictors. Estimates represent coefficients of the logit regression. χ^2-test for joint orthogonality of covariates indicates no systematic correlation with treatment status. Control: 251. Preschool tools: 39. Preschool literacy: 39. Preschool CogX: 59. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[18] Ghanem et al. (2023) argue that to estimate the average treatment effect, the relevant concern is not whether attrition occurs at different rates across experimental conditions, but whether characteristics correlated with the outcomes systematically differ across experimental conditions.

and control, which may report significant differences even when assignment is truly random, especially with small samples like ours (Altman, 1985; Bruhn and McKenzie, 2009; Imai et al., 2008; Senn, 1994).[19] Rather, the χ^2 statistic from the joint orthogonality test is preferred in testing for systematic correlation between observables and experimental condition (McKenzie, 2015).

In choosing which variables to include, prior work in statistics and economics recommends testing on only covariates that would directly impact the outcomes of interest. Afterwards, the main regression models can exclude or include additional controls to determine whether this changes estimates (Altman, 1985; Bruhn and McKenzie, 2009; Imai et al., 2008). We therefore only test for joint orthogonality of covariates based on variables that may directly influence time preferences and test scores (child demographic characteristics, baseline assessment scores, family income, and school). Based on these variables, the χ^2 statistics indicate no significant correlation with treatment (p-values are 0.45 for Tools, 0.36 for Literacy, and 0.89 for CogX). In our regression models in Section 3-4, we later add controls for maternal characteristics (age, high school graduation status, full-time employment status) and show that coefficient estimates barely change with these additional controls.

As mentioned above, one limitation of our data is the small sample size of each preschool curriculum. However, aggregating all treatments into a single preschool group would prevent us from comparing curricula that emphasize self-regulation skills with those that emphasize traditional reading and math skills. Given this trade-off, we chose to explore results by curriculum. We report wild bootstrap-t confidence intervals to adjust our inference for the small number of treated observations (Cameron et al., 2008; Rao, 2019; Roodman et al., 2019; Wu, 1986).[20] In robustness checks, we present p-values after conducting multiple hypothesis tests.

[19] Prior work argues that demonstrating covariate balance is neither necessary nor sufficient in testing for orthogonality between covariates and treatment assignment. It does not guide in assessing the validity of the results when a few covariates differ between experimental conditions, nor can it detect cases where small, insignificant differences in covariates lead to important differences in outcomes (Altman, 1985; Bruhn and McKenzie, 2009; Imai et al., 2008).

[20] We use the boottest package in Stata to implement the wild bootstrap procedure following Roodman et al. (2019). We use 999 bootstrap replications.

3 Characterizing Time Preferences

3.1 Aggregate Choices

We measure patience by whether children chose the later option in the Advance Decision. In the full sample of 388 children, 80.8% were patient in Week 1 and 75.7% were patient in Week 2. Patience declined between Week 1 and Week 2 ($p < 0.05$). This leads us to our first result.

Result 1. *In the full sample, children exhibit high rates of patience in Week 1 and a significant decline in patience in Week 2.*

We investigate this decline by examining stability of patience, measured by whether Advance Decisions changed between Week 1 and 2. We find that 55.7% (216) were stable in patience and chose the later option in both weeks. 7.2% (28) were stable in impatience and chose the earlier option in both weeks. 23.7% (92) exhibited instability, changing their choice across weeks.[21] Among those who made unstable choices, 58.7% switched from the patient to the impatient decision. The decline in patience from Week 1 to Week 2 is driven by the 13.9% of children (54) who were *unstable in their patience*, behaving patiently in Week 1 but impatiently in Week 2.

Using the difference between the Advance and Immediate Decision in Week 1, we find that 21.4% (77) of children were time inconsistent. Of these children, 74.0% (57) were impulsive, behaving patiently and then impatiently between the Advance and Immediate Decisions. Rates of time inconsistency among our sample are comparable to those reported for adults (Augenblick et al., 2015; Sadoff et al., 2020).[22] We do not observe statistically significant differences in time inconsistency between Week 1 and 2 (p=0.12), but we do observe that impulsivity is lower in Week 2 ($p < 0.05$). This could happen because children use the commitment device, or because patience is lower in Week 2 to begin with, leaving children with fewer opportunities to act impulsively.[23]

Children may not make the patient choice in Week 2 if they were impulsive in Week 1 and learned from their past behavior. That is, children who behaved impulsively in Week 1 may become less confident that they can wait out for the later option in Week 2. We test for this explanation

[21] These numbers do not add to 100% because 71 children (18.3%) were not present for both weeks of the task.

[22] Inconsistency rates among our sample fall within the 15-30% reported by Augenblick et al. (2015) and below the 40-50% reported by Sadoff et al. (2020).

[23] In Week 2, 18% (29) of children are inconsistent, with only 55% of them (16) behaving impulsively. Of the children in Week 2 who are patient in the Immediate Decision in Week 2, 55% use the commitment device.

by evaluating whether children classified as impulsive in Week 1 are less likely to delay in the Week 2 Advance Decision. In the full sample, we find no evidence that this is the case, as there is no significant difference in the proportion who delay among impulsive children (15/57=26.3%) relative to non-impulsive children (67/303=22.1%). Similarly, regression results in Appendix Table A.1 show no significant correlation between patience in Week 2 and impulsive behavior in Week 1.

Table 4: Correlations with Cognitive and Executive Function

	Decision in Glow Bracelet Task	
	(1)	(2)
	A: Dep. Var. = Stable in Patience	
Cognitive	0.019	
	[-0.022,0.060]	
Executive Function		0.005
		[-0.046,0.055]
Observations	313	313
	B: Dep. Var. = Inconsistent, Week 1	
Cognitive	-0.035**	
	[-0.065,-0.005]	
Executive Function		-0.018
		[-0.060,0.025]
Observations	335	335
	C: Dep. Var. = Impulsive, Week 1	
Cognitive	-0.023	
	[-0.049,0.002]	
Executive Function		-0.008
		[-0.041,0.024]
Observations	335	335
	D: Dep. Var. = Commit	
Cognitive	0.005	
	[-0.040,0.048]	
Executive Function		0.027
		[-0.026,0.078]
Observations	315	315

Notes: Correlations between time preferences and assessment scores. Brackets show 90% CI after wild bootstrapping. Cognitive and executive function measures constructed using exploratory factor analysis of test scores from 2010 to 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

If children are aware of their self-control problems, they may take up commitment in Week 2. 51% (135) of children take up commitment. However, in the overall sample, they do not appear to use the commitment device to manage their self-control problems, as there is no correlation between impulsivity in Week 1 and commitment take-up in Week 2 (p-value from a regression of commitment and Week 1 impulsivity is 0.654, see Table 7). The related literature with adults on

commitment take-up is mixed: some studies show that commitment take-up is positively correlated with prior inconsistency, indicating that individuals are sophisticated about their inconsistency and use commitment devices to reduce it (Ariely and Wertenbroch, 2002; Ashraf et al., 2006; Augenblick et al., 2015; Kaur et al., 2010). However, other papers show a negative correlation (Sadoff et al., 2020), suggesting that consistent individuals are more likely to use commitment. In Section 4, we explore commitment device use by preschool curriculum. We find that take-up rises with past impulsive behavior only in the *Tools* group. Among this group, it appears our commitment device was used by those most likely to need it.

3.2 Association of Time Preferences with Prior Skills

We next evaluate the association of our time preference measures with scores on cognitive and executive function tests taken prior to the time preference task. For each year, we use composite scores that combine measures of reading, writing and math (cognitive) and working memory, inhibitory control and attention shifting (executive functions). We then use exploratory factor analysis to extract a summary measure of cognitive and executive function skills across all years prior to the time preference task (2010-14). Table 4 presents estimated correlations with the composite cognitive and executive function measures, while Appendix Tables A.2 and A.3 present correlation estimates with each sub-test. We report 90% wild bootstrap confidence intervals in brackets.

Overall, we find remarkably little predictive power of prior test scores on behaviors in the time preference task. Both cognitive score and executive function show no significant correlation with stability of patience, which equals 1 if children delay in the Advance Decision of both weeks and 0 otherwise. Similarly, they show no significant correlation with commitment take-up. The estimates suggest that children's behaviors in the time preference task represent something that is distinct from what is captured by cognitive and executive function assessment scores. This finding is consistent with Watts et al. (2018), who find that parent- and teacher-reported self-regulation measures do little to predict children's ability to delay gratification, and Castillo et al. (2020), who find that behaviors in economic experiments have a weak association with cognitive and executive function assessment scores.

Result 2. *Stability of patience is poorly explained by existing cognitive and executive function*

measures.

3.3 Association of Time Preferences with Schooling Outcomes in Middle Childhood

In Table 5, we evaluate the association of each of our measures with standardized test scores measured 1-4 years later. All regressions include demographic controls (described in Section 4). We also include controls for preschool curriculum and the interaction of time preferences with preschool curriculum. The coefficient on time preferences can therefore be interpreted as the estimated correlation with later academic performance for the control group.[24] We assign cognitive and executive function scores to 0 when they are missing and use missing dummies as controls in our regression.

We find that stability of patience predicts reading (column 1) but does not significantly predict math (column 2). Children who behave patiently in both weeks have 0.2 standard deviations higher reading scores in the ensuing four years (column 1, $p < 0.05$). To situate this magnitude, stability of patience predicts reading scores roughly as well as executive function scores and about half as well as cognitive scores. This estimate is statistically significant after controlling for underlying differences in cognitive performance and executive function, indicating that the association between stability of patience and reading performance cannot be explained by the history of these skills. In the next section, we will show that the main treatment effect of preschool is on the stability of patience. Putting the two results together, our findings suggest that preschool has the potential to change time preferences in ways that improve later academic performance.

Result 3. *After accounting for the roles of cognitive and executive function, stability of patience uniquely predicts reading up to four years later, but not math scores.*

In column (3), we show that our stability of patience measure predicts future patience, elicited in 2018-2020. These patience tasks were performed in an entirely different environment, for an entirely different purpose. They focused only on quantifying patience using a continuous measure, rather than aiming to distinguish patience from time inconsistency, impulsivity, and commitment demand. These tasks involved asking children to choose between the number of earlier rewards

[24] Appendix Table A.6 conducts similar regressions for only the control group, without interactions with preschool curriculum.

versus later rewards. Our results show that our patience measure in the glow bracelet task reliably captures these 2018-2020 patience measures, even though they were conducted 3-5 years later.

Table 5: Regression of Future Outcomes on Glow Bracelet Task Choices

	Reading	Math	Patience
	(1)	(2)	(3)
	A: Indep. Var. = Stable in Patience		
Stable in Patience	0.206**	0.103	0.344**
	[0.055,0.360]	[-0.028,0.236]	[0.100,0.594]
Cognitive	0.353***	0.346***	0.012
	[0.283,0.430]	[0.272,0.424]	[-0.103,0.130]
Executive Function	0.226***	0.192***	0.037
	[0.144,0.308]	[0.117,0.269]	[-0.084,0.158]
Observations	324	324	303
	B: Indep. Var. = Inconsistent, Week 1		
Inconsistent, Week 1	-0.009	-0.037	-0.259
	[-0.193,0.180]	[-0.211,0.136]	[-0.571,0.054]
Cognitive	0.341***	0.327***	0.005
	[0.272,0.411]	[0.260,0.398]	[-0.106,0.120]
Executive Function	0.249***	0.209***	0.017
	[0.172,0.328]	[0.136,0.283]	[-0.095,0.131]
Observations	346	346	325
	C: Indep. Var. = Impulsive, Week 1		
Impulsive, Week 1	-0.013	-0.056	-0.282
	[-0.242,0.213]	[-0.261,0.159]	[-0.656,0.090]
Cognitive	0.339***	0.327***	0.011
	[0.271,0.408]	[0.260,0.397]	[-0.098,0.125]
Executive Function	0.245***	0.209***	0.034
	[0.171,0.325]	[0.136,0.281]	[-0.079,0.147]
Observations	346	346	325
	D: Indep. Var. = Commit		
Commit	0.045	-0.005	-0.067
	[-0.115,0.201]	[-0.154,0.143]	[-0.295,0.159]
Cognitive	0.350***	0.340***	0.027
	[0.277,0.428]	[0.268,0.415]	[-0.078,0.140]
Executive Function	0.220***	0.193***	0.060
	[0.137,0.304]	[0.116,0.270]	[-0.058,0.178]
Observations	326	326	305

Notes: Brackets show 90% CI after wild bootstrapping. Outcome variables are factors derived from exploratory factor analysis using quarterly standardized math and reading assessment scores from 2015 to 2019 in Columns (1) and (2) and from annual patience elicitation tasks from 2018 to 2020. Cognitive and executive function measures constructed using exploratory factor analysis of test scores from 2010 to 2014. Controls: age, gender, race, and proportion of classmates with 2 rather than 1 bracelet in Week 1. Coefficients on preschool treatment, interaction terms between delay of gratification behavior and preschool treatment suppressed. Coefficients on dummies for missing data suppressed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Turning to panels B-D, inconsistency, impulsivity, and commitment demand do not appear to significantly predict later reading or math scores. The coefficients on inconsistency and impulsivity are negative across all specifications, which would be expected if time inconsistent children exhibit self-control problems when studying and score lower on examinations. However, the point estimates are quite small, and the bootstrapped 90% confidence intervals indicate the negative estimate could

have arisen due to chance. Similarly, the point estimates on commitment are also close to 0.

Appendix Table A.5 reproduces Table 5 after incorporating inverse probability weights and restricting to only children who were present for all decisions. It shows that the association between stability of patience and later reading scores remains significant after controlling for cognitive and executive function and after correcting for multiple hypothesis testing ($p < 0.10$). Similarly, Appendix Table A.6 reproduces Table 5 for only the control group, without interaction terms between time preferences and preschool assignment. Again, the association between stability of patience and later reading scores remains strong and significant ($p < 0.05$).

4 Preschool Treatment Effects

4.1 Summary of Effects

Figure 1a shows predicted rates of patience, inconsistency, and commitment use by CHECC treatment group. It displays 90% confidence intervals after a wild bootstrap t-test that compares each treatment group with the control group (Roodman et al., 2019). For this and all following figures, significant differences occur between treatment and control if the control mean falls outside the 90% confidence interval for a treatment group. The regression controls for child demographics, maternal characteristics, baseline cognitive, baseline executive function, and the fraction of classmates in Week 1 with two bracelets rather than one (see explanation after Equation 1). The results are similar to the raw results on the choice to delay, shown in Appendix Figure A.1.

We report our results with a focus on the holistic change in choices over time, rather than handling each one-shot choice in isolation. We find no treatment effect on patience in Week 1, which is unsurprising since we designed the experiment so that most children would make the patient choice initially. However, in Week 2, patience rates significantly fall among children randomized to the control group (81.6% in Week 1 vs. 71.6% in Week 2, $p < 0.01$) but not the *Tools of the Mind* group (81.3% in Week 1 vs. 93.2% in Week 2, $p = 0.31$). The net effect is greater patience in Week 2 among *Tools* children (21.7% gap in patience, $p < 0.01$). Children in the other two preschool programs – *Literacy Express* and *CogX* – behave similarly to the control group.

Figure 1b then classifies children based on the stability of their time preferences across the two weeks. We find that the decline in patience in the control group but not the *Tools* group is explained

by the *Tools* group's lower instability in patience: compared to the control group, *Tools* children are 6 percentage points less likely to choose the patient option in Week 1 and then switch to the impatient option in Week 2. There is no significant difference between *Tools* and control in the proportion who are stable in their patience, who choose the patient option in both weeks. Similarly, there is no significant difference in the fraction of children who are unstable in their impatience, who choose the impatient option in Week 1 and the patient option in Week 2 (a small fraction of children). We note that there is a difference in stability of impatience, in that *Tools* children are 3 percentage points less likely than control children to choose the impatient option in both weeks ($p < 0.05$). However, this cannot explain the decline in patience among the control group.

Result 4. *The Tools group exhibits lower instability of patience relative to the Control group.*

Figure 1a also demonstrates no significant differences in inconsistency, impulsivity, or commitment uptake between the control group and any treatment group. We find some evidence of a significant decline in inconsistency among children randomized to the *Literacy Express* curriculum. However, this result is only marginally significant in our regression specifications ($p < 0.10$) and not robust to alternative specifications (inverse propensity weighting, limiting to only children present for all days, and multiple hypothesis test correction).

Table 6 further investigates the results on patience using the following regression specification:

$$\text{choice}_{it} = \alpha_0 + \alpha_1 \mathbf{1}(\text{week}_t = 2) + \alpha_2 D_i + \alpha_3 D_i \mathbf{1}(\text{week}_t = 2) + \alpha_4 X_i + \epsilon_{it} \tag{1}$$

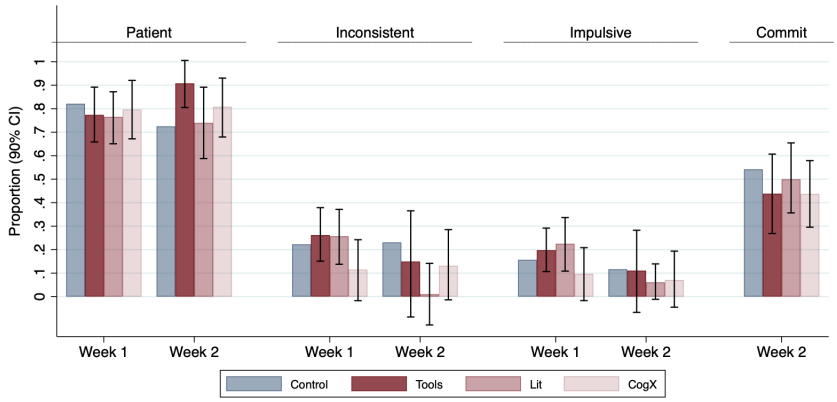
where choice_{it} equals 1 if child i delays by choosing the later option and 0 if child i chooses the earlier option in the Advance Decision in Week t. We regress choice_{it} on a dummy for the second week of the task, a matrix of dummies for preschool curriculum D_i, and their interaction. Under this specification, the coefficient α_1 estimates the change in patience between Week 2 and Week 1 for the control group. Since children were assigned to CHECC condition individually, we do not cluster by elementary school classroom (Abadie et al., 2023).[25]

The vector α_2 represents the difference in patience between each preschool curriculum relative to the control group in Week 1. Finally, the vector of coefficients α_3 represents the difference in

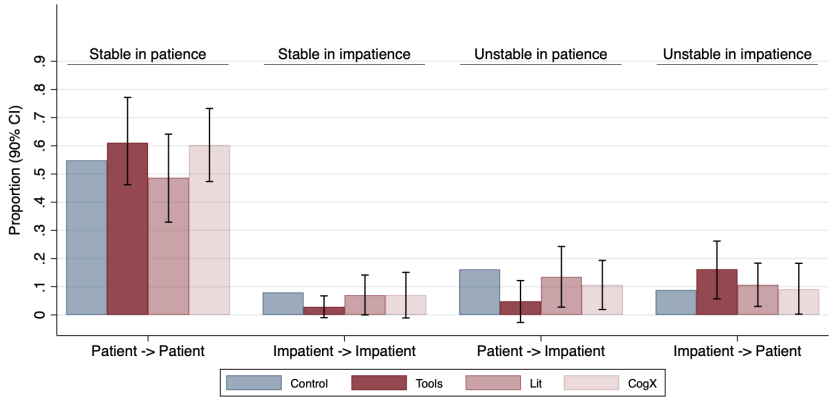
[25] We cluster by elementary school classroom in our Kitagawa-Oaxaca-Blinder decomposition, which examines how classmates influence patient decisions (Appendix Table A.9).

Figure 1: Time preferences across weeks

(a) Patience, inconsistency, impulsivity, and commitment



(b) Stability of time preferences



Notes: Panel A displays rates of patience, time inconsistency, impulsivity, and commitment use across CHECC conditions. Panel B classifies children by whether they are stable in patience, stable in impatience, unstable in patience, or unstable in impatience. 90% confidence intervals of wild bootstrap test comparing treatment to control. If the confidence intervals fall outside of the control mean, there is a significant difference between the treatment group and the control group.

patience between each preschool curriculum in Week 2 and the control group in Week 1. The linear combination of $\alpha_2 + \alpha_3$ represents the difference between control and treatment in Week 2. The linear combination of $\alpha_1 + \alpha_3$ represents the change from Week 1 to Week 2 within treatment groups.

In the matrix X_i, we incorporate controls including age during the glow bracelet task, year(s) of CHECC participation, gender, race (columns 2-4). In columns (3)-(4), we add controls for cognitive and executive function test scores prior to randomization into CHECC treatment.[26] In column (4)

[26]We use baseline scores since all other measures of cognitive and executive function are endogenous to CHECC treatment assignment.

we add maternal controls for age, high school graduation status, and full time employment status. In column (4), we also control for the fraction of classmates with 2 bracelets as opposed to 1 bracelet by the end of Week 1, which Section 4.2 will show to be an important determinant of Week 2 patience. In all regression results we will show, the point estimates barely change across specifications, even after column (4) adds controls for maternal characteristics or the fraction of classmates with 2 bracelets in Week 1. As with Tables 4-5 and Figure 1a, we report 90% wild bootstrapped confidence intervals in brackets.

Table 6: Regression of Advance Decisions on CHECC Treatment

	Chose Later Option in Advance Decision			
	(1)	(2)	(3)	(4)
Week 2	-0.092**	-0.094**	-0.095**	-0.094**
	[-0.158,-0.027]	[-0.159,-0.029]	[-0.160,-0.030]	[-0.159,-0.032]
Tools	-0.014	-0.024	-0.020	0.017
	[-0.121,0.095]	[-0.133,0.087]	[-0.130,0.095]	[-0.093,0.126]
Tools × Week 2	0.180**	0.182**	0.183**	0.183**
	[0.052,0.310]	[0.057,0.310]	[0.056,0.312]	[0.051,0.314]
Literacy	-0.015	-0.028	-0.032	0.021
	[-0.126,0.098]	[-0.136,0.085]	[-0.139,0.083]	[-0.092,0.140]
Literacy × Week 2	0.014	0.014	0.013	0.014
	[-0.153,0.179]	[-0.156,0.185]	[-0.156,0.182]	[-0.149,0.177]
CogX	-0.038	-0.043	-0.037	-0.047
	[-0.154,0.077]	[-0.159,0.071]	[-0.155,0.079]	[-0.161,0.068]
CogX × Week 2	0.149	0.153	0.153	0.149
	[-0.007,0.307]	[-0.003,0.309]	[-0.005,0.311]	[-0.005,0.303]
Observations	707	707	707	707
Child Gender and Race		Yes	Yes	Yes
Pre-assessment scores			Yes	Yes
Parental Controls				Yes

Notes: All regressions control for year of CHECC participation and child age. Parental controls include age, HS graduate status, and full-time employment of mother and father. Dummies for missing variables suppressed. 90% confidence intervals use wild bootstrap-t method from Roodman et al. (2019). P-values in brackets use Sidak correction for multiple hypothesis testing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6 reports results for patience in the Advance Decision. Again, we find no differences across groups in the likelihood of making the patient choice in Week 1. However, patience rates decline by 9-10 percentage points from Week 1 to Week 2 among control children, as shown by the negative coefficient on Week 2 ($p < 0.01$). *Literacy* children exhibit similar behavior as control children across both decisions. In contrast, children in the *Tools* group experience no decline in patience, making them 16 percentage points more likely to choose the patient option in Week 2 compared to the control group (adding the coefficients on *Tools* and *Tools*$\times$*Week 2*, $p < 0.01$). For children in the *CogX* group, point estimates are similar to those in the *Tools* group in both weeks.

However, we are underpowered to detect whether the trajectory of patience among *CogX* children differed from that of control children ($p = 0.15$).

We next turn to take-up of the commitment device. Figure 1a demonstrates no direct treatment effects on commitment use in Week 2. This is to be expected, since commitment devices are designed to help with a particular form of inconsistency: impulsivity. We therefore focus on the relationship between impulsivity and commitment use, rather than expecting the commitment device to affect all children. The commitment device helps only children who prefer the later option, are impulsive and are aware of their impulsivity. Among children who demonstrate this sort of sophistication, we would expect greater take-up rates among those who behaved impulsively in Week 1.

Table 7: Correlation between Impulsive Behavior in Week 1 and Commitment Take-up in Week 2

	Commitment in Week 2			
	(1)	(2)	(3)	(4)
Impulsive in week 1	0.047	0.013	0.022	0.011
	[-0.124,0.217]	[-0.160,0.186]	[-0.152,0.192]	[-0.165,0.191]
PK Tools	-0.140	-0.144	-0.152	-0.119
	[-0.318,0.035]	[-0.324,0.035]	[-0.334,0.027]	[-0.311,0.067]
PK Literacy	-0.018	-0.013	-0.035	-0.027
	[-0.187,0.157]	[-0.179,0.163]	[-0.204,0.148]	[-0.202,0.165]
PK CogX	-0.047	-0.058	-0.036	-0.058
	[-0.205,0.113]	[-0.216,0.102]	[-0.199,0.130]	[-0.223,0.111]
Impulsive in week 1 × PK Tools	0.413*	0.502*	0.492*	0.517*
	[0.015,0.762]	[0.043,0.873]	[0.044,0.854]	[0.078,0.942]
Impulsive in week 1 × PK Lit	-0.167	-0.118	-0.109	-0.110
	[-0.575,0.253]	[-0.529,0.314]	[-0.534,0.333]	[-0.528,0.344]
Impulsive in week 1 × PK CogX	-0.291	-0.212	-0.221	-0.190
	[-0.624,0.040]	[-0.533,0.097]	[-0.537,0.084]	[-0.517,0.122]
Observations	335	335	335	335
Child Gender and Race		Yes	Yes	Yes
Pre-assessment scores			Yes	Yes
Parental Controls				Yes

Notes: All regressions control for year of CHECC participation and child age. Parental controls include age, high school graduate status, and full-time employment of mother and father. Dummies for missing variables suppressed. Brackets display 90% confidence intervals using wild bootstrap-t method from Roodman et al. (2019). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The *Tools* group is the only group that exhibits this type of sophistication. Table 7 estimates the correlation between impulsive behavior in Week 1 and commitment take-up in Week 2, controlling for preschool assignment. It is intended to document a pattern of behavior that, while not necessarily causal, shows that *Tools* children who behave impulsively in Week 1 are 41-50% more likely to use commitment in Week 2 ($p < 0.10$). It appears that for the *Tools* group, commitment was higher among the very children who were likely to benefit from it: those who failed to wait for the later option despite preferring it initially. Bolstering this descriptive evidence, Appendix Table A.7 shows greater commitment take-up among *Tools* children who failed to delay in the Immediate

Decision than those who successfully delayed ($p < 0.10$). These results are consistent with prior work documenting greater commitment use among time inconsistent adults (Ariely and Wertenbroch, 2002; Ashraf et al., 2006; Augenblick et al., 2015; Kaur et al., 2010), but contrast with other studies that find greater commitment demand among those who are time consistent (Sadoff et al., 2020 for adults; Alan and Ertac, 2015 for children). We find no significant correlation between Week 1 behavior and commitment use among control, *Literacy*, or *CogX* children.

Result 5. *Relative to the control group, Tools children are more likely to exhibit sophisticated use of the commitment device.*

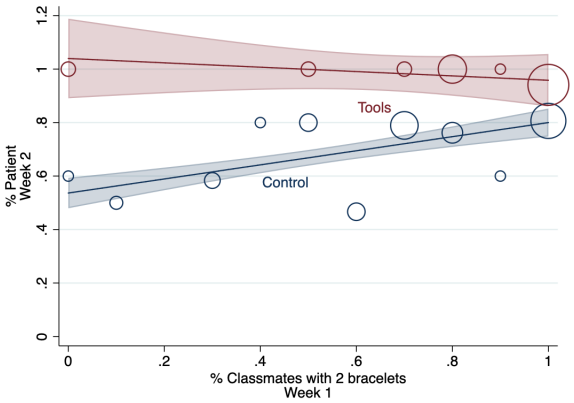
4.2 Interpretation

We next evaluate potential explanations for the treatment effects on stability of patience. Since children across treatment groups are similar in their patience, inconsistency and impulsivity in Week 1, differences in experiences in Week 1 cannot explain our findings. However, children may adjust their behavior in Week 2 based on the behavior of peers, and this adjustment may differ by treatment. To evaluate this explanation, we exploit quasi-random variation in the assignment of children to elementary classrooms. Following participation in CHECC, children enter elementary school in the Chicago Heights School District. Differences across classrooms should be orthogonal to CHECC treatment status, since assignment to classrooms is made without reference to prior preschool curriculum assignment. Indeed, we find no significant correlation between preschool treatment assignment and elementary school classroom, with and without demographic controls (Appendix Table A.8). Furthermore, our analysis (described below) finds no differences in characteristics between the control group and any treatment group.

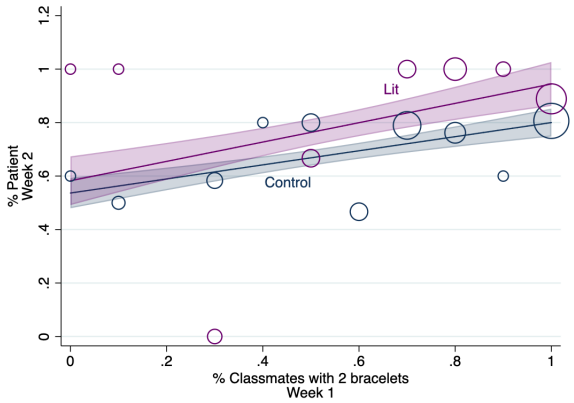
Figure 2 plots the linear prediction of Week 2 patience based on the share of classmates who ended Week 1 with the later option of two bracelets. Across the three panels of Figure 2, we compare control with *Tools* (panel 2a), *Literacy* (panel 2b), and *CogX* (panel 2c). The slope is positive and steep for the control group and *Literacy*, suggesting that these groups are responsive to the behavior of peers. On the other hand, *Tools of the Mind* - and to a lesser extent *CogX* - show flat slopes, suggesting that these groups are not responsive to the behavior of peers. In Appendix Table A.9, we perform a Kitagawa-Oaxaca-Blinder decomposition to distinguish whether the gaps

Figure 2: Patience in Week 2 and Classmates' Outcomes in Week 1

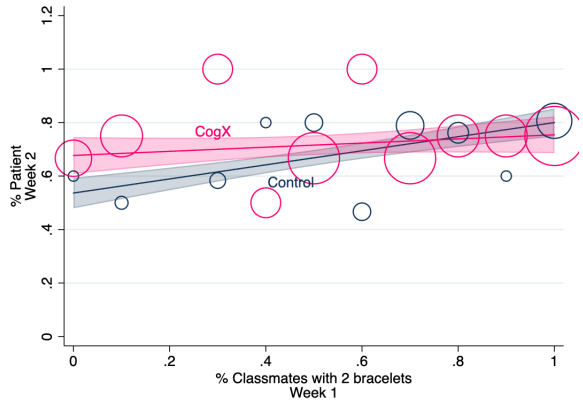
(a) Tools of the Mind



(b) Literacy Express



(c) CogX



Notes: Own patience in Week 2 vs. proportion of classmates with 2 bracelets at end of Week 1. Panel 2a presents 95% CI of linear predictions for control and Tools. Panel 2b presents linear predictions for control and Literacy. Panel 2c presents linear predictions for control and CogX.

in patience between *Tools of the Mind* and the control group stem from differences in responsiveness to peers or differences in the share of peers who are patient in Week 1 (Blinder, 1973; Kitagawa, 1955; Oaxaca, 1973). The decomposition also allows us to compare the relative importance of peer influence with other predictors, including age, gender, race, maternal characteristics, baseline cognitive score, and baseline executive function. We estimate that 94.4% of the 21.3 percentage point gap in patience is driven by differences in the coefficients on explanatory variables. That is, almost all of the gap is due to differences in how *Tools* and control children respond to the same explanatory variables, rather than differences in the explanatory variables themselves. Among all explanatory variables, differences in susceptibility to peer influence is the only significant predictor of the gap in patience. If we were to eliminate differences in susceptibility to peer influence between the two groups, the patience gap would reduce from 21.3 to 7.8 percentage points.

We take our results as evidence that social environment influences time preference formation, just as it influences preferences for risk (Ahern et al., 2014; Browne et al., 2021; Eckel et al., 2012), alcohol and drug use (Gaviria and Raphael, 2001; Lundborg, 2006; Sacerdote, 2001), and occupations (Mani and Mullin, 2004). The finding that *Tools of the Mind* children are less susceptible to peer influence is consistent with prior research which shows that *Tools of the Mind* reduces bullying, peer ostracism and clique formation (Diamond et al., 2019). Children from supportive preschool environments could perceive greater freedom to pursue their own preferences, rather than conform to peer pressure.

Another interpretation is that the glow bracelet task did not isolate time preferences, but reflects a combination of time preferences and social concerns. It is a long-recognized problem that preference elicitations depend on the external environment, as it is impossible to isolate pure preferences in a vacuum (Cohen et al., 2020; Sprenger, 2015). Nevertheless, we improve on the extant literature by taking repeated measurements of time preferences, rather than relying on one-shot choices. To that end, Table 5 column (3) shows positive and significant correlations between behavior in the glow bracelet task and patience measures taken 3-5 years later in 2018-2020, after children changed classrooms and schools. While it is always possible that elicited behaviors do not perfectly capture underlying preferences, we interpret this as evidence that our glow bracelet measures capture components of patience that transcend the social context in which they are elicited.

4.3 Robustness Tests

We report robustness tests in Appendix Figure A.2, Appendix Tables A.4-A.5 and Appendix Tables A.10-A.11. We first show that our results replicate when using inverse probability weights to account for the possibility that preschool assignment influenced the likelihood of remaining in District 170, and therefore participating in the glow bracelet task. Second, some children are absent on some task days, so our results could be driven by differences in the sample across days. We show that our results replicate when using only children who were present for all decisions (81.7%, $N = 317$).

Finally, we conduct multiple hypothesis testing (List et al., 2019; Roodman et al., 2019). Appendix Tables A.4-A.7 and A.10-A.11 present p-values after adjusting for testing of multiple equations and multiple explanatory variables. Almost all main results remain significant. Stability of patience continues to predict reading scores in the ensuing four years (Appendix Table A.5). Patience significantly declines from Week 1 to Week 2 in the control group, and the *Tools* group continues to exhibit greater patience than the control group in Week 2 (Appendix Figure A.2, Appendix Table A.10). The exception is that the correlation between commitment demand and impulsive behavior in Week 1 becomes insignificant after multiple hypothesis correction ($p = 0.41 - 0.48$, Appendix Table A.11).[27]

5 Conclusion

Early life skills are particularly important for later-life outcomes (Heckman, 2008). Our study focuses on patience, time inconsistency and commitment demand at ages 5-10. Despite initially high rates of patience in Week 1 (80.8%), patience tapers off in Week 2 for the full sample (75.7%). This indicates that patience is not necessarily stable, and that repeated elicitations more reliably capture time preferences than one-shot measures. We characterize the stability of patience as the ability to maintain patient choices throughout both weeks of our glow bracelet task. We find that it is explained by neither traditional cognitive measures nor by executive function measures. Even

[27]Finally, we explore whether our main treatment effects hold when we combine the two preschool curricula that teach executive function, *Tools of the Mind* and *CogX*. Appendix Table A.10 demonstrates a significant effect on patience in Week 2 ($p < 0.05$), but this effect becomes insignificant after multiple hypothesis correction ($p = 0.15$). Appendix Table A.11 shows no significant effect on commitment demand in Week 2. Both results may stem from weaker impacts in the *CogX* group, which devoted less than half of its curriculum to executive function and focused more on literacy skills.

after controlling for cognitive and executive function skills, stability of patience positively predicts reading scores in the ensuing four years. This result joins landmark studies in psychology which show that delaying gratification at ages 3-5 increases the probability of success in adulthood (Mischel et al., 1972; Mischel et al., 1988; Mischel et al., 1989; Shoda et al., 1990). It also joins work by Castillo et al. (2020) and Skibbe et al. (2019) which link early time preferences and self-regulation to improvements in discipline and literacy years later.

The next question is whether interventions affect the stability of patience. We leverage a field experiment which randomized children into a preschool or control group. We find greater stability of patience among children who participated in the *Tools of the Mind* curriculum, which focused on improving decision-making skills. For all other children, we present evidence that susceptibility to peer influence contributed to the decline in patient behavior from Week 1 to Week 2. We also find that *Tools of the Mind* children are more likely to display sophisticated use of the commitment device, whereas other groups display no evidence of sophistication.

We are the first to show that preschool can improve stability of patience, which predicts later reading achievement. In particular, our results suggest that children who can repeatedly delay rewards are better equipped to invest in reading. By showing that time preferences can be taught, our intervention demonstrates one pathway through which preschools can equip disadvantaged children with the tools to better invest in their human capital. Our paper illuminates one way early childhood interventions can help narrow achievement gaps between those born poor and those born rich.

Acknowledgements. *We thank John List and the team at the Chicago Heights Early Childhood Center (CHECC), especially Edie Dobrez and Kristin Troutman, for making this work possible. We also thank Charles Sprenger for invaluable input on the experimental design. The field work research was carried out while Samek was a visiting professor at the University of Chicago and field work and analysis was funded in part by the University of Chicago departmental research funds, the Kenneth and Anne Griffin Foundation and NIH grant 1R01DK114238. This research is approved by the Institutional Review Boards (IRB) at the University of Chicago, the University of Southern California and the University of California-San Diego. The study is registered in the AEA RCT Registry, AEARCTR-0010891. We are grateful for feedback from seminar participants at the Association for Education Finance and Policy, University of California at San Diego Rady School of Business, Society of Labor Economists, Stanford Institute for Theoretical Economics, University of Alabama, Economic Science Association, Advances with Field Experiments, Tilburg University, Queen's University, Columbia University Teachers College, and Michigan State University.*

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, *138*(1), 1–35.

Ahern, K. R., Duchin, R., and Shumway, T. (2014). Peer effects in risk aversion and trust. *The Review of Financial Studies*, *27*(11), 3213–3240.

Alan, S., and Ertac, S. (2015). Patience, self-control and the demand for commitment: Evidence from a large-scale field experiment. *Journal of Economic Behavior and Organization*, *115*, 111–122.

Alan, S., and Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, *126*(5), 1865–1911.

Alan, S., Ertac, S., and Gumus, I. (2021). Does a forward-looking perspective affect self-control and the demand for commitment? results from an educational intervention. *Economic Inquiry*, *59*(4), 1533–1546.

Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society Series D: The Statistician*, *34*(1), 125–136.

Anders, J., Barr, A., and Smith, A. (2023). The effect of early childhood education on adult criminality: Evidence from the 1960s through 1990s. *American Economic Journal: Economic Policy*.

Andreoni, J., Kuhn, M. A., List, J. A., Samek, A., Sokal, K., and Sprenger, C. (2019). Toward an understanding of the development of time preferences: Evidence from field experiments. *Journal of Public Economics*, *177*, 104039.

Ariely, D., and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science*, *13*(3), 219–224.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, *121*(2), 635–672.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, *130*(3), 1067–1115.

Augenblick, N., and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, *86*(3), 941–975.

Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. (2006). The high/scope perry preschool program cost–benefit analysis using data from the age-40 followup. *Journal of Human Resources*, *41*(1), 162–190.

Benjamin, D. J., Laibson, D., Mischel, W., Peake, P. K., Shoda, Y., Wellsjo, A. S., and Wilson, N. L. (2020). Predicting mid-life capital formation with pre-school delay of gratification and life-course measures of self-regulation. *Journal of Economic Behavior and Organization*, *179*, 743–756.

Ben-Ner, A., List, J. A., Putterman, L., and Samek, A. (2017). Learned generosity? an artefactual field experiment with parents and their children. *Journal of Economic Behavior and Organization*, *143*, 28–44.

Blair, C., and Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PloS one*, *9*(11), e112393.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 436–455.

Bodrova, E., and Leong, D. J. (2006). *Tools of the mind*. Pearson Australia Pty Limited Upper Saddle River, NJ.

Browne, M. J., Hofmann, A., Richter, A., Roth, S.-M., and Steinorth, P. (2021). Peer effects in risk preferences: Evidence from germany. *Annals of Operations Research*, *299*, 1129–1163.

Bruhn, M., and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, *1*(4), 200–232.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., and Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*(2), 231.

Campbell, F. A., and Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child development*, *65*(2), 684–698.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., and Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the abecedarian project. *Applied Developmental Science*, *6*(1), 42–57.

Cappelen, A., List, J., Samek, A., and Tungodden, B. (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy*, *128*(7), 2739–2758.

Carneiro, P., and Ginja, R. (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from head start. *American Economic Journal: Economic Policy*, *6*(4), 135–173.

Castillo, M., Jordan, J. L., and Petrie, R. (2018). Children's rationality, risk attitudes and field behavior. *European Economic Review*, *102*, 62–81.

Castillo, M., List, J. A., Petrie, R., and Samek, A. (2020). Detecting drivers of behavior at an early age: Evidence from a longitudinal field experiment.

Chuan, A., List, J., and Samek, A. (2021). Do financial incentives aimed at decreasing inter-household inequality increase intrahousehold inequality? *Journal of Public Economics*, *196*, 104382.

Chuan, A., List, J. A., Samek, A., and Samujjwala, S. (2022). Parental investments in early childhood and the gender gap in math and literacy. *AEA Papers and Proceedings*, *112*, 603–08.

Cohen, J., Ericson, K. M., Laibson, D., and White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, *58*(2), 299–347.

Cowell, J. M., Samek, A., List, J., and Decety, J. (2015). The curious relation between theory of mind and sharing in preschool age children. *PloS one*, *10*(2), e0117947.

Cox, J. C., List, J. A., Price, M., Sadiraj, V., and Samek, A. (2016). Moral costs and rational choice: Theory and experimental evidence.

Currie, J., and Thomas, D. (1995). Does head start make a difference? *The American Economic Review*, *85*(3), 341–364.

Delaney, L., and Doyle, O. (2012). Socioeconomic differences in early childhood time preferences. *Journal of Economic Psychology*, *33*(1), 237–247.

Diamond, A., Barnett, W. S., Thomas, J., and Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388.

Diamond, A., Lee, C., Senften, P., Lam, A., and Abbott, D. (2019). Randomized control trial of tools of the mind: Marked benefits to kindergarten children and their teachers. *PloS one*, *14*(9), e0222447.

Eckel, C. C., Grossman, P. J., Johnson, C. A., De Oliveira, A. C., Rojas, C., and Wilson, R. K. (2012). School environment and risk preferences: Experimental evidence. *Journal of Risk and Uncertainty*, *45*, 265–292.

Falk, A., Kosse, F., Pinger, P., Schildberg-Hörisch, H., and Deckers, T. (2021). Socioeconomic status and inequalities in children's iq and economic preferences. *Journal of Political Economy*, *129*(9), 2504–2545.

Fryer, R. G., Levitt, S. D., List, J. A., et al. (2015). Parental incentives and early childhood achievement: A field experiment in chicago heights.

Fryer, R. G., Levitt, S. D., List, J. A., and Samek, A. (2020). Introducing cogx: A new preschool education program combining parent and child interventions.

Garcıéa, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy*, *128*(7), 2502–2541.

Gaviria, A., and Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, *83*(2), 257–268.

Ghanem, D., Hirshleifer, S., and Ortiz-Beccera, K. (2023). Testing attrition bias in field experiments. *Journal of Human Resources*.

Hahn, R. A., Barnett, W. S., Knopf, J. A., Truman, B. I., Johnson, R. L., Fielding, J. E., Muntaner, C., Jones, C. P., Fullilove, M. T., Hunt, P. C., et al. (2016). Early childhood education to promote health equity: A community guide systematic review. *Journal of Public Health Management and Practice: JPHMP*, *22*(5), E1.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010a). Analyzing social experiments as implemented: A reexamination of the evidence from the highscope perry preschool program. *Quantitative economics*, *1*(1), 1–46.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3), 289–324.

Heckman, J. J., and Carneiro, P. (2003). Human capital policy.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010b). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, *94*(1-2), 114–128.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., and Vaughan Jr, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, *6*(3), 149–185.

Houser, D., List, J. A., Piovesan, M., Samek, A. S., and Winter, J. (2015). On the origins of dishonesty: From parents to children.

Imai, K., King, G., and Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *171*(2), 481–502.

Jones, S. M., and Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies. social policy report. volume 26, number 4. *Society for Research in Child Development*.

Kaur, S., Kremer, M., and Mullainathan, S. (2010). Self-control and the development of work arrangements. *American Economic Review*, *100*(2), 624–628.

Khwaja, A., Sloan, F., and Salm, M. (2006). Evidence on preferences and subjective beliefs of risk takers: The case of smokers. *International Journal of Industrial Organization*, *24*(4), 667–682.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, *50*(272), 1168–1194.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.

List, J. A., and Samak, A. C. (2013). Exploring the origins of charitable acts: Evidence from an artefactual field experiment with young children. *Economics Letters*, *118*(3), 431–434.

List, J. A., Samek, A., and Suskind, D. L. (2018). Combining behavioral economics and field experiments to reimagine early childhood education. *Behavioural Public Policy*, *2*(1), 1–21.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*, 773–793.

Literacy express. (2010). https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_lit_express_072710.pdf

Lonigan, C. J., Farver, J. M., Phillips, B. M., and Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, *24*, 305–337.

Lonigan, C., Phillips, B., and Menchetti, J. (2006). Impact of preschool literacy curricula: Results of a randomized trial. *Annual Meeting of the Society for the Scientific Study of Reading, Vancouver, British Columbia, Canada.*

Lührmann, M., Serra-Garcia, M., and Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, *10*(3), 309–332.

Lundborg, P. (2006). Having the wrong friends? peer effects in adolescent substance use. *Journal of Health Economics*, *25*(2), 214–233.

Mani, A., and Mullin, C. H. (2004). Choosing the right pond: Social approval and occupational choice. *Journal of Labor Economics*, *22*(4), 835–861.

Masse, L. N., and Barnett, W. S. (2002). A benefit cost analysis of the abecedarian early childhood intervention.

McKenzie, D. (2015). *Tools of the trade: A joint test of orthogonality when testing for balance.* Retrieved November 2, 2023, from https://blogs.worldbank.org/impactevaluations/tools-trade-joint-test-orthogonality-when-testing-balance

Meier, S., and Sprenger, C. D. (2015). Temporal stability of time preferences. *Review of Economics and Statistics*, *97*(2), 273–286.

Mischel, W., Ebbesen, E. B., and Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, *21*(2), 204.

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

Mørkbak, M. R., Gyrd-Hansen, D., and Kjær, T. (2017). Can present biasedness explain early onset of diabetes and subsequent disease progression? exploring causal inference by linking survey and register data. *Social Science and Medicine*, *186*, 34–42.

Nores, M., Belfield, C. R., Barnett, W. S., and Schweinhart, L. (2005). Updating the economic impacts of the high/scope perry preschool program. *Educational Evaluation and Policy Analysis*, *27*(3), 245–261.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 693–709.

O'Donoghue, T., and Rabin, M. (1999). Doing it now or later. *American Economic Review*, *89*(1), 103–124.

O'Donoghue, T., and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, *90*(10-11), 1825–1849.

Ramon, I., Chattopadhyay, S. K., Barnett, W. S., and Hahn, R. A. (2018). Early childhood education to promote health equity: A community guide economic review. *Journal of Public Health Management and Practice: JPHMP*, *24*(1), e8.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, *109*(3), 774–809.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, *19*(1), 4–60.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Sadoff, S., and Samek, A. (2019). Can interventions affect commitment demand? a field experiment on food choice. *Journal of Economic Behavior and Organization*, *158*, 90–109.

Sadoff, S., Samek, A., and Sprenger, C. (2020). Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies*, *87*(4), 1954–1988.

Samak, A. C. (2013). Is there a gender gap in preschoolers' competitiveness? an experiment in the us. *Journal of Economic Behavior and Organization*, *92*, 22–31.

Samek, A., and Sheremeta, R. M. (2017). Selective recognition: How to recognize donors to increase charitable giving. *Economic Inquiry*, *55*(3), 1489–1496.

Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in medicine*, *13*(17), 1715–1726.

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

Skibbe, L. E., Montroy, J. J., Bowles, R. P., and Morrison, F. J. (2019). Self-regulation and the development of literacy and language achievement from preschool through second grade. *Early childhood research quarterly*, *46*, 240–251.

Sprenger, C. (2015). Judging experimental evidence on dynamic inconsistency. *American Economic Review*, *105*(5), 280–285.

Strotz, R. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, *23*(3), 165–180.

Thaler, R. H., and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, *89*(2), 392–406.

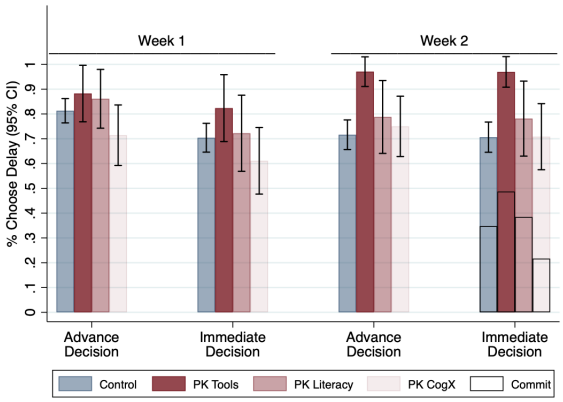
Watts, T. W., Duncan, G. J., and Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159–1177.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*(4), 1261–1295.

Online Appendix - For Online Publication

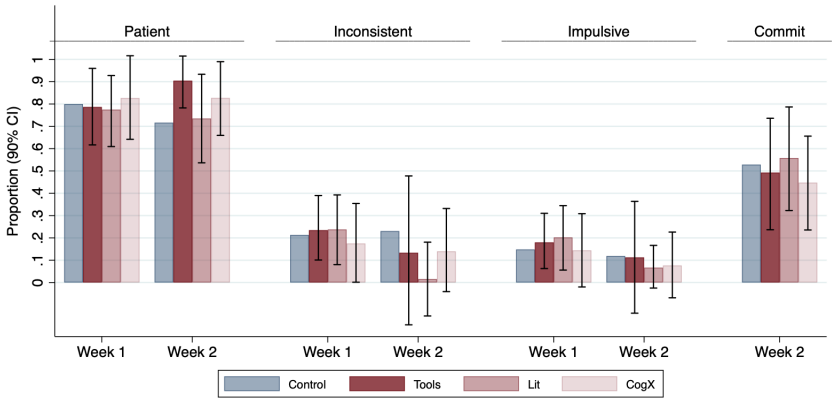
A Appendix Tables and Figures

Figure A.1: Proportion choose delay in each decision



Notes: Rates of delay in each decision across the two weeks of the glow bracelet task, raw results. 95% confidence intervals in brackets. If the confidence intervals fall outside of the control mean, there is a significant difference between the treatment group and the control group.

Figure A.2: Patience, inconsistency, and commitment across weeks, robustness checks



Notes: Rates of patience, time inconsistency, and commitment use across CHECC conditions. 90% confidence intervals of wild bootstrap test comparing treatment to control. If the confidence intervals fall outside of the control mean, there is a significant difference between the treatment group and the control group. Confidence intervals estimated after 1) inverse weighting for the probability of remaining in the Chicago Heights School District conditional on CHECC condition; 2) restricting the sample to only children present for all decision days; and 3) adjusting for multiple hypothesis testing.

Table A.1: Correlation between impulsive behavior in Week 1 and patience in Week 2

	Patience in Week 2			
	(1)	(2)	(3)	(4)
Impulsive in week 1	-0.027	-0.037	-0.033	0.019
	[-0.179,0.127]	[-0.192,0.120]	[-0.192,0.128]	[-0.140,0.182]
Tools	0.204	0.184	0.205	0.211
	[0.112,0.294]***	[0.090,0.275]***	[0.107,0.302]***	[0.102,0.312]***
Impulsive in week 1 × Tools	0.054	0.064	0.066	0.134
	[-0.241,0.369]	[-0.259,0.377]	[-0.250,0.409]	[-0.205,0.497]
Literacy	0.053	0.030	0.047	0.069
	[-0.095,0.201]	[-0.122,0.179]	[-0.120,0.208]	[-0.090,0.225]
Impulsive in week 1 × Lit	-0.155	-0.133	-0.166	-0.122
	[-0.768,0.487]	[-0.790,0.574]	[-0.813,0.545]	[-0.793,0.577]
CogX	0.064	0.061	0.071	0.070
	[-0.090,0.214]	[-0.089,0.211]	[-0.094,0.230]	[-0.095,0.234]
Impulsive in week 1 × CogX	0.059	0.091	0.092	0.125
	[-0.259,0.371]	[-0.217,0.406]	[-0.212,0.399]	[-0.186,0.423]
Observations	335	335	335	335
Child Gender and Race		Yes	Yes	Yes
Pre-assessment scores			Yes	Yes
Parental Controls				Yes

Notes: All regressions control for year of CHECC participation and child age. Parental controls include age, high school graduate status, and full-time employment of mother. Dummies for missing variables suppressed. 90% confidence intervals use wild bootstrap-t method from Roodman et al. (2019). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Correlations with Cognitive Scores

	Decision in Glow Bracelet Experiment				
	(1)	(2)	(3)	(4)	(5)
	A: Dep. Var. = Stable in Patience				
PPVT	0.012				
	[-0.027,0.053]				
WJ-A		0.021			
		[-0.019,0.062]			
WJ-L			0.015		
			[-0.027,0.057]		
WJ-Q				0.003	
				[-0.034,0.042]	
WJ-S					0.021
					[-0.019,0.061]
	B: Dep. Var. = Inconsistent, Week 1				
PPVT	-0.032				
	[-0.061,-0.002]*				
WJ-A		-0.038			
		[-0.068,-0.008]**			
WJ-L			-0.019		
			[-0.050,0.011]		
WJ-Q				-0.035	
				[-0.067,-0.005]*	
WJ-S					-0.010
					[-0.038,0.018]
	C: Dep. Var. = Impulsive, Week 1				
PPVT	-0.023				
	[-0.051,0.003]				
WJ-A		-0.027			
		[-0.054,-0.003]*			
WJ-L			-0.016		
			[-0.042,0.011]		
WJ-Q				-0.029	
				[-0.056,-0.002]*	
WJ-S					-0.008
					[-0.032,0.016]
	D: Dep. Var. = Commit				
PPVT	0.022				
	[-0.021,0.064]				
WJ-A		0.004			
		[-0.039,0.044]			
WJ-L			-0.003		
			[-0.046,0.043]		
WJ-Q				0.012	
				[-0.034,0.055]	
WJ-S					-0.025
					[-0.068,0.023]

Notes: Correlations between time preferences and cognitive scores. Brackets show 90% CI after wild bootstrapping. Cognitive and executive function measures constructed using exploratory factor analysis of test scores from 2010 to 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Correlations with Executive Function

	Decision in Glow Bracelet Experiment		
	(1)	(2)	(3)
	A: Dep. Var. = Stable in Patience		
Working Memory	0.001		
	[-0.042,0.048]		
Inhibitory Control		-0.008	
		[-0.053,0.039]	
PSRA			0.018
			[-0.046,0.059]
	B: Dep. Var. = Inconsistent, Week 1		
Working Memory	-0.007		
	[-0.040,0.025]		
Inhibitory Control		0.016	
		[-0.012,0.041]	
PSRA			-0.038
			[-0.087,0.004]
	C: Dep. Var. = Impulsive, Week 1		
Working Memory	-0.011		
	[-0.037,0.013]		
Inhibitory Control		0.018	
		[-0.005,0.037]	
PSRA			-0.025
			[-0.069,0.011]
	D: Dep. Var. = Commit		
Working Memory	0.003		
	[-0.045,0.052]		
Inhibitory Control		0.055	
		[0.007,0.108]*	
PSRA			-0.000
			[-0.045,0.056]

Notes: Correlations between time preferences and executive function score. Brackets show 90% CI after wild bootstrapping. Cognitive and executive function measures constructed using exploratory factor analysis of test scores from 2010 to 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Correlations with Cognitive and Executive Function, Robustness Checks

Inverse Propensity Weights, present for all decisions

	Decision in Glow Bracelet Task	
	(1)	(2)
	A: Dep. Var. = Stable in Patience	
Cognitive	0.018	
	[-0.029,0.064]	
MHT-corrected p-value	[0.822]	
Executive Function		0.002
		[-0.051,0.061]
MHT-corrected p-value		[0.941]
Observations	294	294
	B: Dep. Var. = Inconsistent, Week 1	
Cognitive	-0.048	
	[-0.081,-0.017]**	
MHT-corrected p-value	[0.069]	
Executive Function		-0.030
		[-0.077,0.013]
MHT-corrected p-value		[0.337]
Observations	294	294
	C: Dep. Var. = Commit	
Cognitive	0.021	
	[-0.033,0.074]	
MHT-corrected p-value	[0.475]	
Executive Function		0.051
		[-0.006,0.110]
MHT-corrected p-value		[0.248]
Observations	295	295

Notes: All regressions use inverse weights for the probability of completing the glow bracelet task conditional on CHECC assignment. Sample includes only children who were present for all experimental decisions. Conventional standard errors, clustered by child, in parentheses. First set of brackets display 90% confidence intervals using wild bootstrap-t method from Cameron, Gelbach, and Miller (2008). Second set of brackets display Romano-Wolf adjustment for multiple hypothesis testing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Regression of Future Outcomes on Glow Bracelet Task Choices

Inverse propensity weights, present for all decisions

	Reading	Math	Patience
	(1)	(2)	(3)
	A: Indep. Var. = Stable in Patience		
Stable in Patience	0.224	0.090	0.383
	[0.075,0.372]**	[-0.051,0.227]	[0.158,0.612]**
MHT-corrected p-value	[0.069]	[0.703]	[0.059]
Cognitive	0.340	0.318	0.016
	[0.269,0.412]***	[0.248,0.394]***	[-0.118,0.154]
MHT-corrected p-value	[0.010]	[0.010]	[0.941]
Executive Function	0.234	0.186	-0.030
	[0.154,0.311]***	[0.117,0.254]***	[-0.167,0.104]
MHT-corrected p-value	[0.020]	[0.020]	[0.941]
Observations	306	306	286
	B: Indep. Var. = Inconsistent, Week 1		
Inconsistent, Week 1	0.085	0.032	-0.130
	[-0.119,0.288]	[-0.157,0.226]	[-0.451,0.210]
MHT-corrected p-value	[0.941]	[0.980]	[0.941]
Cognitive	0.344	0.319	0.002
	[0.271,0.420]***	[0.250,0.393]***	[-0.129,0.135]
MHT-corrected p-value	[0.010]	[0.010]	[1.000]
Executive Function	0.237	0.187	-0.041
	[0.158,0.315]***	[0.118,0.256]***	[-0.182,0.094]
MHT-corrected p-value	[0.010]	[0.020]	[0.941]
Observations	306	306	286
	C: Indep. Var. = Impulsive, Week 1		
Impulsive, Week 1	0.093	0.018	-0.062
	[-0.152,0.332]	[-0.217,0.242]	[-0.441,0.342]
MHT-corrected p-value	[0.950]	[1.000]	[1.000]
Cognitive	0.341	0.318	0.011
	[0.269,0.417]***	[0.249,0.392]***	[-0.117,0.144]
MHT-corrected p-value	[0.010]	[0.010]	[1.000]
Executive Function	0.233	0.187	-0.025
	[0.155,0.311]***	[0.118,0.255]***	[-0.163,0.109]
MHT-corrected p-value	[0.020]	[0.030]	[1.000]
Observations	306	306	286
	D: Indep. Var. = Commit		
Commit	0.096	0.028	-0.055
	[-0.065,0.257]	[-0.121,0.176]	[-0.293,0.181]
MHT-corrected p-value	[0.802]	[0.990]	[0.980]
Cognitive	0.336	0.317	0.023
	[0.261,0.414]***	[0.247,0.393]***	[-0.109,0.151]
MHT-corrected p-value	[0.010]	[0.010]	[0.990]
Executive Function	0.228	0.187	-0.010
	[0.144,0.313]***	[0.113,0.262]***	[-0.143,0.119]
MHT-corrected p-value	[0.020]	[0.020]	[0.990]
Observations	307	307	287

Notes: All regressions use inverse propensity weights and include only children present for all decision days. First set of brackets show 90% CI after wild bootstrapping. Second set of brackets use Romano-Wolf adjustment for multiple hypothesis testing. Outcome variables are factors derived from exploratory factor analysis using quarterly standardized math and reading assessment scores from 2015 to 2019 in Columns (1) and (2) and from annual patience elicitation tasks from 2018 to 2020. Cognitive and executive function controls constructed using exploratory factor analysis of test scores from 2010 to 2014. Controls: age, gender, race, and proportion of classmates with 2 rather than 1 bracelet in Week 1. Coefficients on preschool treatment, interaction terms between delay of gratification behavior and preschool treatment suppressed. Coefficients on dummies for missing data suppressed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Regression of Future Outcomes on Glow Bracelet Task Choices (control group only)

	Reading	Math	Patience
	(1)	(2)	(3)
	A: Indep. Var. = Stable in Patience		
Stable in Patience	0.211**	0.107	0.317**
	[0.068,0.355]	[-0.029,0.245]	[0.062,0.571]
Cognitive	0.369***	0.347***	0.035
	[0.272,0.465]	[0.254,0.446]	[-0.112,0.184]
Executive Function	0.276***	0.173***	-0.048
	[0.174,0.375]	[0.077,0.271]	[-0.188,0.090]
Observations	213	213	199
	B: Indep. Var. = Inconsistent, Week 1		
Inconsistent, Week 1	-0.010	-0.060	-0.234
	[-0.198,0.166]	[-0.230,0.109]	[-0.551,0.079]
Cognitive	0.355***	0.342***	0.028
	[0.265,0.446]	[0.249,0.434]	[-0.096,0.157]
Executive Function	0.298***	0.195***	-0.051
	[0.195,0.400]	[0.093,0.297]	[-0.187,0.085]
Observations	231	231	216
	C: Indep. Var. = Impulsive, Week 1		
Impulsive, Week 1	-0.015	-0.080	-0.250
	[-0.240,0.217]	[-0.299,0.136]	[-0.627,0.133]
Cognitive	0.354***	0.341***	0.027
	[0.265,0.444]	[0.249,0.432]	[-0.097,0.155]
Executive Function	0.298***	0.196***	-0.048
	[0.196,0.400]	[0.094,0.297]	[-0.185,0.089]
Observations	231	231	216
	D: Indep. Var. = Commit		
Commit	0.053	0.007	-0.043
	[-0.109,0.211]	[-0.138,0.153]	[-0.279,0.194]
Cognitive	0.362***	0.343***	0.021
	[0.265,0.461]	[0.251,0.441]	[-0.129,0.173]
Executive Function	0.273***	0.172***	-0.052
	[0.170,0.373]	[0.074,0.272]	[-0.193,0.085]
Observations	213	213	199

Notes: Brackets show 90% CI after wild bootstrapping. Outcome variables are factors derived from exploratory factor analysis using quarterly standardized math and reading assessment scores from 2015 to 2019 in Columns (1) and (2) and from annual patience elicitation tasks from 2018 to 2020 in Column (3). Cognitive and executive function measures constructed using exploratory factor analysis of test scores from 2010 to 2014. Controls: age, gender, race, and proportion of classmates with 2 rather than 1 bracelet in Week 1. Sample contains only control group students. Coefficients on dummies for missing data suppressed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Correlation between commitment and Week 1 choices

	Commitment in Week 2			
	(1)	(2)	(3)	(4)
Advance decision, week 1	0.029	0.016	0.010	0.007
	[-0.128,0.186]	[-0.142,0.173]	[-0.151,0.167]	[-0.158,0.167]
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	[1.000]
Immediate decision, week 1	-0.057	-0.009	-0.018	-0.014
	[-0.211,0.095]	[-0.162,0.142]	[-0.170,0.132]	[-0.183,0.155]
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	[1.000]
PK Tools	-0.231	-0.235	-0.264	-0.178
	[-0.756,0.453]	[-0.758,0.452]	[-0.797,0.426]	[-0.742,0.397]
MHT-corrected p-values	[0.999]	[0.999]	[0.998]	[1.000]
PK Literacy	-0.304	-0.272	-0.283	-0.234
	[-0.730,0.301]	[-0.659,0.234]	[-0.685,0.227]	[-0.632,0.294]
MHT-corrected p-values	[0.965]	[0.953]	[0.950]	[0.981]
PK CogX	-0.108	-0.081	-0.065	-0.061
	[-0.412,0.175]	[-0.385,0.214]	[-0.373,0.217]	[-0.375,0.246]
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	[1.000]
Advance decision, week 1 × PK Tools	0.501	0.573*	0.584*	0.561*
	[-0.141,0.997]	[0.019,1.130]	[0.018,1.150]	[0.005,1.142]
MHT-corrected p-values	[0.815]	[0.663]	[0.663]	[0.679]
Advance decision, week 1 × PK Lit	0.110	0.141	0.139	0.093
	[-0.423,0.662]	[-0.371,0.666]	[-0.392,0.670]	[-0.432,0.624]
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	[1.000]
Advance decision, week 1 × PK CogX	-0.137	-0.082	-0.080	-0.077
	[-0.491,0.198]	[-0.435,0.247]	[-0.431,0.244]	[-0.443,0.267]
MHT-corrected p-values	[0.999]	[1.000]	[1.000]	[1.000]
Immediate decision, week 1 × PK Tools	-0.378*	-0.443*	-0.432*	-0.471*
	[-0.670,-0.055]	[-0.763,-0.100]	[-0.754,-0.066]	[-0.831,-0.098]
MHT-corrected p-values	[0.576]	[0.512]	[0.566]	[0.506]
Immediate decision, week 1 × PK Lit	0.228	0.168	0.159	0.155
	[-0.198,0.649]	[-0.270,0.602]	[-0.275,0.607]	[-0.294,0.588]
MHT-corrected p-values	[0.986]	[0.999]	[1.000]	[0.999]
Immediate decision, week 1 × PK CogX	0.168	0.071	0.076	0.044
	[-0.130,0.458]	[-0.223,0.372]	[-0.225,0.378]	[-0.257,0.343]
MHT-corrected p-values	[0.987]	[1.000]	[1.000]	[1.000]
Observations	335	335	335	335
Child Gender and Race		Yes	Yes	Yes
Pre-assessment scores			Yes	Yes
Parental Controls				Yes

Notes: All regressions control for year of CHECC participation and child age. Parental controls include age, high school graduate status, and full-time employment of mother and father. Dummies for missing variables suppressed. First set of brackets present 90% wild bootstrap confidence intervals (Roodman et al., 2019). Second set of brackets show p-values after Sidak correction for multiple hypothesis testing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: No significant correlation between preschool treatment assignment and elementary school classroom

	Tools	Literacy	CogX
	(1)	(2)	(3)
	Panel A: Without controls		
Observations	103	112	140
Joint F-test p-value	0.897	1.000	0.899
	Panel B: With controls		
Observations	98	105	124
Joint F-test p-value	0.973	0.994	0.924

Notes: Logit regression of treatment vs. control on class assignment. Panel A regresses treatment on class assignment without controls. Panel B controls for child age, gender, race, baseline cognitive and non-cognitive test scores, income bracket, and fraction of classmates with 2 bracelets in Week 1. F-test for joint orthogonality of covariates indicates no systematic correlation with treatment status. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Oaxaca decomposition on % classmates with 2 bracelets

	Choose Later Option, Week 2 Advance Choice Tools vs. Control			
	Overall			
	(1)	(2)	(3)	(4)
Control	0.757***			
	(0.0510)			
Tools	0.970***			
	(0.0286)			
Control - Tools	-0.213***			
	(0.0579)			
Endowments	-0.000352			
	(0.0386)			
Coefficients	-0.201**			
	(0.0829)			
Interaction	-0.0119			
	(0.0773)			
Decomposition:		Endowments	Coefficient	Interaction
% classmates with 2 bracelets		0.0183	0.502***	-0.0708
		(0.0199)	(0.187)	(0.0472)
Female		0.0116	0.0974	-0.0141
		(0.0203)	(0.0642)	(0.0247)
Black		0	-0.0517	-0.00722
		(.)	(0.0386)	(0.0264)
Hispanic		0.00160	-0.0861	0.00112
		(0.0146)	(0.126)	(0.0103)
Baseline cognitive (std)		0.0214	0.0660	-0.0236
		(0.0271)	(0.0568)	(0.0309)
Baseline executive function (std)		-0.0176	-0.0185	0.0232
		(0.0184)	(0.0216)	(0.0266)
Mother's age		0.00737	-0.162	0.0346
		(0.0219)	(0.265)	(0.0601)
Mother is HS graduate		0.0161	0.0216	-0.00434
		(0.0218)	(0.0983)	(0.0204)
Mother works full-time		-0.000546	0.00387	-0.0000896
		(0.00768)	(0.0233)	(0.00137)
N		144		

Standard errors clustered by classroom in parentheses. All regressions additionally control for years of CHECC participation and dummies for missing variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Regression of advance decisions on CHECC treatment, robustness checks

	Chose Later Option, Advance Decision			
	(1)	(2)	(3)	(4)
Week 2	-0.185	-0.185	-0.184	-0.184
	[-0.322,-0.047]**	[-0.322,-0.047]**	[-0.340,-0.033]**	[-0.341,-0.033]**
MHT-corrected p-values	[0.068]	[0.075]	[0.138]	[0.101]
Tools	-0.049	-0.036	-0.013	
	[-0.179,0.080]	[-0.165,0.095]	[-0.146,0.129]	
MHT-corrected p-values	[0.996]	[0.999]	[1.000]	
Tools × Week 2	0.279	0.279	0.259	
	[0.103,0.451]***	[0.103,0.451]***	[0.085,0.432]**	
MHT-corrected p-values	[0.021]	[0.021]	[0.081]	
Literacy	-0.045	0.020	0.032	0.027
	[-0.163,0.079]	[-0.094,0.137]	[-0.091,0.161]	[-0.094,0.156]
MHT-corrected p-values	[0.994]	[1.000]	[1.000]	[0.998]
Literacy × Week 2	0.040	0.041	0.089	0.089
	[-0.178,0.243]	[-0.178,0.246]	[-0.122,0.297]	[-0.122,0.297]
MHT-corrected p-values	[1.000]	[1.000]	[0.990]	[0.962]
CogX	-0.032	-0.039	-0.015	
	[-0.152,0.085]	[-0.151,0.073]	[-0.156,0.118]	
MHT-corrected p-values	[0.999]	[0.997]	[1.000]	
CogX × Week 2	0.202	0.207	0.203	
	[0.015,0.388]*	[0.020,0.393]*	[0.004,0.409]*	
MHT-corrected p-values	[0.434]	[0.417]	[0.503]	
Tools or Cogx				-0.014
				[-0.122,0.096]
MHT-corrected p-values				[1.000]
Tools or Cogx × Week 2				0.224
				[0.051,0.395]**
MHT-corrected p-values				[0.150]
Observations	707	707	633	633
Inverse probability weights	Yes	Yes	Yes	Yes
% classmates with 2 bracelets		Yes	Yes	Yes
Present all days			Yes	Yes

Notes: All regressions use inverse weights for the probability of completing the glow bracelet task conditional on CHECC assignment. All regressions control for age, race, gender, baseline cognitive scores, baseline executive function, and year of CHECC participation. Columns (2)-(4) add controls for the number of classmates with 2 bracelets by the end of Week 1. Columns (3) and (4) restrict the sample to only children who were present for all decisions in the glow bracelet task. Dummies for missing variables suppressed. First set of brackets display 90% confidence intervals using wild bootstrap-t method from Roodman et al. (2019). Second set of brackets display Sidak adjustment for multiple hypothesis testing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Correlation between impulsive behavior in Week 1 and commitment take-up in Week 2, robustness checks

	Commitment in Week 2			
	(1)	(2)	(3)	(4)
Impulsive in Week 1	-0.036	0.015	0.003	-0.004
	[-0.242,0.176]	[-0.204,0.232]	[-0.222,0.223]	[-0.224,0.211]
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	[1.000]
Tools	-0.175	-0.195	-0.158	
	[-0.376,0.029]	[-0.403,0.009]	[-0.376,0.053]	
MHT-corrected p-values	[0.666]	[0.558]	[0.812]	
Impulsive in Week 1 × Tools	0.603	0.630	0.634	
	[0.373,0.981]*	[0.339,1.056]*	[0.178,1.260]*	
MHT-corrected p-values	[0.408]	[0.417]	[0.484]	
Literacy	0.034	0.083	0.065	0.077
	[-0.142,0.203]	[-0.106,0.266]	[-0.141,0.272]	[-0.127,0.282]
MHT-corrected p-values	[1.000]	[0.983]	[0.998]	[0.966]
Impulsive in Week 1 × Literacy	-0.192	-0.178	-0.165	-0.170
	[-0.716,0.656]	[-0.723,0.703]	[-1.102,4.707]	[-1.099,4.753]
MHT-corrected p-values	[0.990]	[0.995]	[1.000]	[0.996]
CogX	-0.066	-0.023	0.003	
	[-0.249,0.132]	[-0.220,0.190]	[-0.199,0.217]	
MHT-corrected p-values	[0.996]	[1.000]	[1.000]	
Impulsive in Week 1 × CogX	-0.006	-0.017	-0.046	
	[-0.472,0.369]	[-0.522,0.365]	[-0.565,0.339]	
MHT-corrected p-values	[1.000]	[1.000]	[1.000]	
Tools or CogX				-0.067
				[-0.216,0.087]
MHT-corrected p-values				[0.959]
Impulsive in Week 1× Tools or CogX				0.113
				[-0.240,0.443]
MHT-corrected p-values				[0.985]
Observations	516	516	471	471
Inverse probability weights	Yes	Yes	Yes	Yes
% classmates with 2 bracelets		Yes	Yes	Yes
Present all days			Yes	Yes

Notes: All regressions use inverse weights for the probability of completing the glow bracelet task conditional on CHECC assignment. All regressions control for age, race, gender, baseline cognitive scores, baseline executive function, and year of CHECC participation. Columns (2)-(4) add controls for the number of classmates with 2 bracelets by the end of Week 1. Columns (3) and (4) restrict the sample to only children who were present for all decisions in the glow bracelet task. Dummies for missing variables suppressed. First set of brackets display 90% confidence intervals using wild bootstrap-t method from Roodman et al. (2019). Second set of brackets display Sidak adjustment for multiple hypothesis testing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Experiment Scripts

B.1 Assent (Day 1 only)

Today, I am asking you all to do an activity with me where I will ask you some questions about choices you want to make. You will get some glow bracelets for doing the activity. We will keep your answers to look at later, to help us learn more about how kids decide. You do not have to do this activity. It is up to you if you want to do the activity or not. You can say no now or you can even change your mind later. No one will be upset with you if you don't want to do this activity.

Does anyone have any questions for me? You can ask me questions at any time. Will you do the activity with me?

B.2 Week 1 Day 1

Thanks for doing the activity. First, we are going to sit quietly. It is very important that we do not talk during the activity. If anyone talks, they will have to go back to class and will not get any prizes. Can we practice being quiet? [wait for kids to settle down]. Here are the glow bracelets that we will play with. (Walk around with tray of glow bracelets.)

In the activity, you are going to decide WHEN you want to get glow bracelets. When you get the glow bracelets, they will glow for 1 day. But after 1 day they will stop glowing. Today you will decide whether you want 1 glow bracelet TOMORROW - that's sooner, OR 2 glow bracelets the DAY AFTER TOMORROW - that's later. You will get your glow bracelet in an envelope at the end of the day that you decide to get it, and get a blank envelope on the other day. The choice you make is up to you. You will only get glow bracelets if you follow the rules for the activity sheet. Here is the activity sheet you will have (Show poster 1). Here are the rules:

- If you want 1 glow bracelet TOMORROW - that's SOONER, you should circle this choice. See this shows ONE glow bracelet. It also shows that you will have to sleep for ONE night before tomorrow comes. Tomorrow is TUESDAY/THURSDAY. (point to the 1 glow bracelet sheet as well as each component).

- If you want 2 glow bracelets the DAY AFTER TOMORROW - that's LATER, circle this choice. See this shows TWO glow bracelets. It also shows that you will have to sleep for TWO nights before the day after tomorrow comes. The day after tomorrow IS WEDNESDAY/FRIDAY. (Point to the 2 glow bracelets sheet and each component).

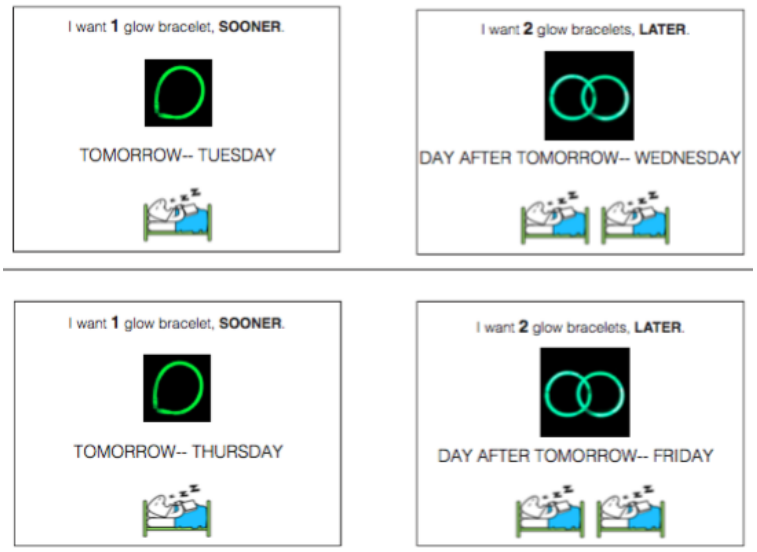


Figure B1: Poster 1 and response sheets: Monday cohort received top panel and Wednesday cohort received bottom panel

Doing the Practice Book

Before we start the activity, you are each going to fill out a practice book to make sure you understand the rules. Please remember to try your best! Do not look at other people's work and no talking! Only turn the pages when I tell you to. I will read out loud and you should follow along. You should circle your answers. (Pass out the practice book and pencils and read out loud.) Okay let's start. Okay please open the practice book. Remember its okay if you don't know the answer. Just try your best.

- This is Katie. Katie played the same game that you are going to play. She decided to circle THIS choice, 2 glow bracelets, LATER. LATER is the DAY AFTER TOMORROW, WEDNESDAY/ FRIDAY. (Say 'Ok, turn the page' after each page you read.)
  - The practice book asks, How many bracelets will Katie get? 1, 2 or 3? (Point to each answer.) Circle your answer now. Ok, turn the page.
  - The practice book asks, When will Katie get her bracelets? Today, tomorrow, or the day after tomorrow? (Point to each answer) Circle your answer now. Ok, turn the page.
- Now let's do another one. This is Sarah. Sarah played the same game that you are going to play. She decided to circle THIS choice, 1 glow bracelet, SOONER. SOONER is tomorrow. Ok, turn the page.
  - The practice book asks, How many bracelets will Sarah get? 1, 2, or 3? (Point to each answer.) Circle your answer now. Ok, turn the page.
  - The practice book asks, When will Sarah get her bracelets? Today, tomorrow, or the day after tomorrow? (Point to each answer.) Circle your answer now. Ok, turn the page.
  - The practice book asks, Who will get their bracelet sooner? Remember, this was Katie's choice (Point back to Katie picture). This was Sarah's choice (point back to Sarah

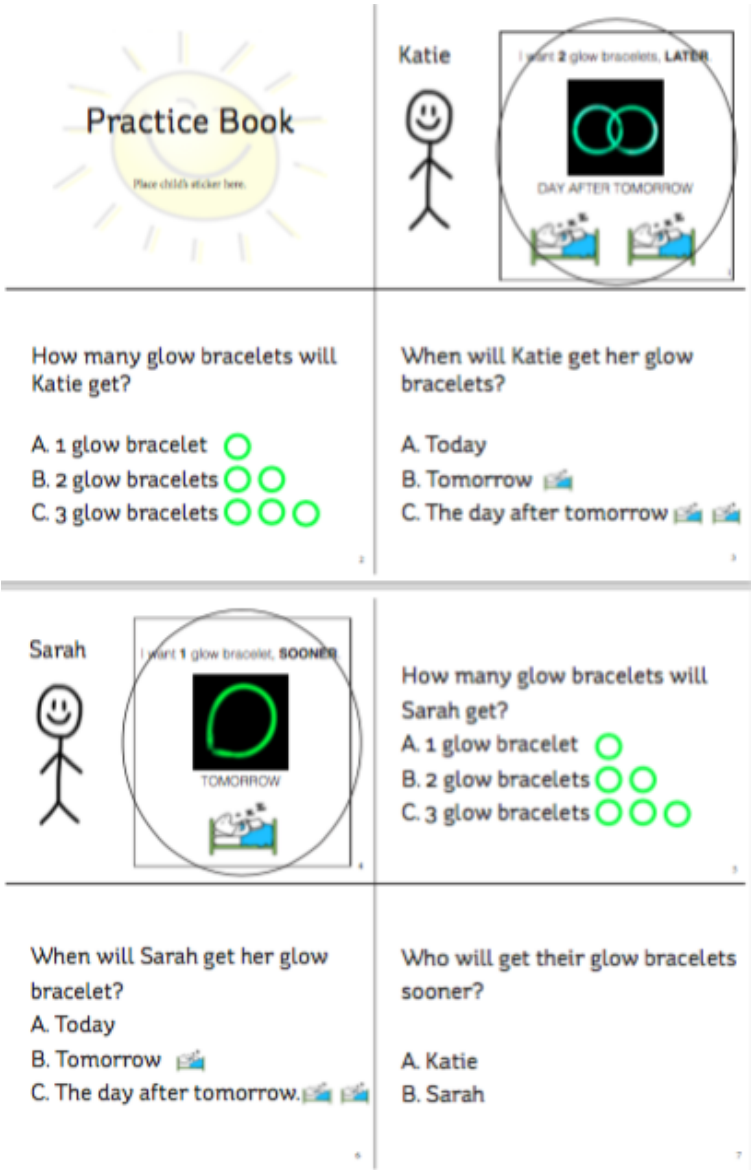


Figure B2: Practice Book Screenshots

Picture). So who will get their bracelet sooner, Katie, or Sarah? (Point to each answer). Circle your answer now.

Great job! Now I will collect your practice books and we can start the activity.

Doing the Activity

(go back to poster 1) Now remember,

- If you want 1 glow bracelet TOMORROW - that's SOONER, you should circle this choice (Point to the 1 glow bracelet sheet).
- If you want 2 glow bracelets the DAY AFTER TOMORROW - that's LATER, circle this choice (Point to the 2 glow bracelets sheet)

Here are your activity sheets. Remember to make your decision quietly, keep your eyes on your paper and do not talk to anyone. If you have a question, raise your hand and I will come to you. Don't start until I say GO. [pass out decision sheets]. Whatever you choose, we will give you the glow bracelet at the end of that day so that you can take it home with you right away. The choice you make is up to you. Remember you are NOT choosing for today so you will get glow bracelets tomorrow or the day after tomorrow.

When you are done, turn your paper over, and wait quietly. I will collect your sheet only after everyone is finished. Okay, GO!

Okay, I will collect the activity sheets now. Remember, your decision was private! Please do not talk about your decision to anyone. Thanks for doing the activity today. Let's go back to class.

B.3 Week 1 Day 2

Thanks for doing the activity. First, we are going to sit quietly. It is very important that we do not talk during the activity. If anyone talks, they will have to go back to class and will not get any prizes. Can we practice being quiet? [wait for kids to settle down]. Here are the glow bracelets that we are playing with. When you get the bracelets, they will glow for 1 day. But after 1 day they will stop glowing. (Walk around with tray of glow bracelets).

Yesterday you made a choice about 1 glow bracelet TOMORROW or 2 glow bracelets the DAY AFTER TOMORROW. Since you made that choice yesterday, that means the choices are for 1 glow bracelet TODAY or 2 glow bracelets TOMORROW. Here is the activity sheet you had (Show poster 1). Here were the rules:

- If you said you wanted 1 glow bracelet TOMORROW, that's SOONER, you circled this choice. Now that you slept for one night, that's turned into TODAY (Point to the 1 glow bracelet sheet - use dry eraser maker to cross out TOMORROW and write TODAY, and cross out the sleeping person).
- If you wanted 2 glow bracelets the DAY AFTER TOMORROW, that's LATER, you circled this choice. Now that you slept for one night, that's turned into TOMORROW (Point to the 1 glow bracelet sheet - demonstrate crossing out DAY AFTER TOMORROW and writing TOMORROW and cross out one sleeping person).

If you chose for TODAY (Point to poster 1), you will get your glow bracelet as soon as we finish the activity today. If you chose for TOMORROW (Point to poster 1), you will wait and get your glow bracelets tomorrow before you go home from school.

Okay, I will pass out your activity sheets from yesterday now, and you can see the choice that you made quietly. (Pass out sheets.) Please keep your eyes on your paper only and no talking.

Doing the Activity

(DO NOT COLLECT ACTIVITY SHEETS UNTIL THE END!) Now there is a surprise! You can change your decision if you want. You can either STAY WITH YOUR CHOICE from yesterday or CHANGE YOUR MIND. Let me show you the rules:

There is a new yellow activity sheet that counts as the final decision. You will only get glow bracelets if you follow the rules for the new yellow sheet. (Show yellow poster 2). Here are the rules:

- If you want 1 glow bracelet TODAY - that's SOONER, circle this choice. This shows ONE bracelet. It also shows you will get the bracelet as soon as we finish the activity TODAY.

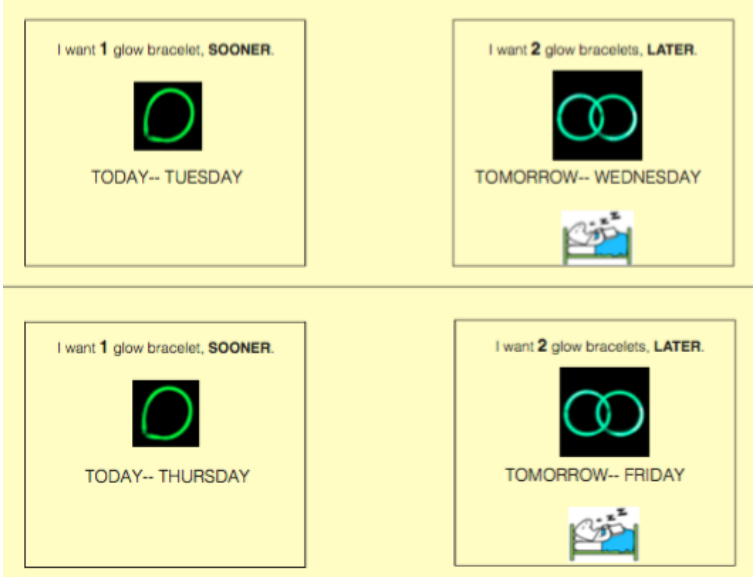


Figure B3: Poster 2 and response sheets: Monday cohort received top panel and Wednesday cohort received bottom panel.

Since, today is TUESDAY/ THURSDAY (Use yellow poster - point to the 1 glow bracelet sheet and each component).

- If you want 2 glow bracelets TOMORROW - that's LATER, circle this choice. This shows TWO bracelets. It also shows that you will have to sleep for ONE night before tomorrow comes. Since tomorrow is WEDNESDAY/ FRIDAY (Point to the 2 glow bracelets sheet and each component)

Remember, you are choosing to get 1 glow bracelet TODAY OR 2 glow bracelets TOMORROW - so you should only circle one choice. You will get your glow bracelet in an envelope on the day you decide to get it, and get a blank envelope on the other day. The choice you make is up to you. The choice you make on the new yellow activity sheet is the one that counts as the final decision.

(Pass out activity sheets/pencils. Staple yellow sheet on the child's white sheet.) I will pass out your new yellow activity sheets now. Don't start until I say GO. If you have a question, raise your hand and I will come to you. When you are done, turn your paper over, and wait quietly. I will collect your sheets only after everyone is finished. Okay, GO.

Okay, I will collect the activity sheets now and pass out your envelopes. Remember, your decision was private! Please do not talk about your decision to anyone. (As you collect the activity sheets, check to see what the child chose and give them the corresponding envelope for day 2 - either an envelope with 1 bracelet or an envelope with nothing)

Thanks for doing the activity today. Let's go back to class.

B.4 Week 2 Day 1

Thanks for doing the activity. First, we are going to sit quietly. It is very important that we do not talk during the activity. If anyone talks, they will have to go back to class and will not get any prizes. Can we practice being quiet? [wait for kids to settle down]

Here are the glow bracelets that we will play with. (Walk around with tray of glow bracelets).

In the activity, just like last week, you are going to decide WHEN you want to get glow bracelets. When you get the glow bracelets, they will glow for 1 day. But after 1 day they will stop glowing. Today you will decide whether you want 1 glow bracelet TOMORROW - that's sooner, OR 2 glow bracelets the DAY AFTER TOMORROW - that's later. You will get your glow bracelets in an envelope at the end of the day on the day you decide to get it, and get a blank envelope on the other day. The choice you make is up to you.

You will only get glow bracelets if you follow the rules for the activity sheet. Here is the activity sheet you will have (Show poster 1). Here are the rules:

- If you want 1 glow bracelet TOMORROW - that's SOONER, you should circle this choice. This shows ONE glow bracelet. It also shows that you will have to sleep for ONE night before tomorrow comes since tomorrow is TUESDAY/THURSDAY (Point to the 1 glow bracelet sheet as well as each component).
- If you want 2 glow bracelets the DAY AFTER TOMORROW - that's LATER, circle this choice. This shows TWO glow bracelets. It also shows that you will have to sleep for TWO nights before the day after tomorrow comes. Since the day after tomorrow is WEDNESDAY/ FRIDAY (Point to the 2 glow bracelets sheet and each component)

Remember, you are NOT choosing for today and will not get any glow bracelets today.

Doing the Activity I

Okay, I will pass out the activity sheet now. If you have a question, raise your hand and I will come to you. When you are done, turn your paper over, and wait quietly. We will collect your sheets only after everyone is finished. Okay, GO.

Doing the Activity II

(WAIT UNTIL EVERYONE IS DONE, BUT DO NOT COLLECT ACTIVITY SHEETS YET!) Now there is one more thing. Remember, last week you got a surprise: a new yellow activity sheet so you could make your final decision.

Today, you can decide whether you want your final decision to be TODAY - on the white activity sheet that you just filled out OR TOMORROW on the yellow activity sheet. Let me show you the rules. Here is the activity sheet you will have: (Show poster 3).

- If you want your final decision to be on the WHITE sheet that you JUST filled out TODAY, you should circle the white box.
- If you want your final decision to be on the YELLOW sheet that you will fill out TOMORROW, you should circle the yellow box.

Only kids who circle the yellow box will be able to make a new decision tomorrow. Kids who circle the white box will not be able to make a new decision tomorrow, though they will still get a yellow sheet.

Okay, I will pass out the activity sheet now. [pass out sheet.] If you have a question, raise your hand and I will come to you. When you are done, turn your paper over, and wait quietly. We will collect your sheets only after everyone is finished. Okay, GO.

Okay, I will collect the activity sheets now. Remember, your decision was private! Please do not talk about your decision to anyone.

Thanks for doing the activity today. Let's go back to class.

B.5 Week 2 Day 2

Thanks for doing the activity. First, we are going to sit quietly. It is very important that we do not talk during the activity. If anyone talks, they will have to go back to class and will not get any prizes. Can we practice being quiet? [wait for kids to settle down]

Here are the glow bracelets that we are playing with. When you get the bracelets, they will glow for 1 day. But after 1 day they will stop glowing. [walk around with tray of glow bracelets].

Yesterday you made a choice about 1 glow bracelet TOMORROW or 2 glow bracelets the DAY AFTER TOMORROW. Since you made that choice yesterday, that means the choices are for 1 glow bracelet TODAY or 2 glow bracelets TOMORROW.

Here is the activity sheet you had [show poster of decision sheet]. Here were the rules:

- If you said you wanted 1 glow bracelet TOMORROW, that's SOONER, you circled this choice. Now that you slept for one night, that's turned into TODAY (point to the 1 glow bracelet sheet -cross out TOMORROW and write TODAY, and cross out the sleeping person).
- If you wanted 2 glow bracelets the DAY AFTER TOMORROW, that's LATER, you circled this choice. Now that you slept for one night, that's turned into TOMORROW (point to the 1 glow bracelet sheet -cross out DAY AFTER TOMORROW and write TOMORROW and cross out one sleeping person).

Remember, you made one more choice. You decided whether to make your final decision on the white activity sheet yesterday or make your final decision on the yellow activity sheet today.

- If you decided make your final decision on the WHITE activity sheet YESTERDAY, you will get a blank yellow sheet and sit quietly.
- If you decided to make your final decision on the YELLOW activity sheet TODAY, you will get a new yellow sheet that will count as the final decision.

Doing the Activity

I will pass out your activity sheets now, and you can see the choices you made quietly. Please keep your eyes on your paper only and no talking. Don't start until I say GO. [pass out sheets and staple yellow w/ child's sticker on top of white sheet].

If you made your final decision yesterday on the white activity sheet, please sit quietly. You will get glow bracelets based on the decision you made yesterday on your white decision sheet.

If you decided to make your final decision TODAY on the YELLOW ACTIVITY SHEET, you will get a new yellow sheet to fill out. You will only get glow bracelets if you follow the rules for the new yellow sheet. Here is the activity sheet you will have [show poster 2]. Here are the rules:

- If you want 1 glow bracelet TODAY - that's SOONER, circle this choice. This shows ONE bracelet. It also shows you will get the bracelet as soon as we finish the activity TODAY. Today is TUESDAY/THURSDAY. (use yellow poster - point to the 1 glow bracelet sheet and each component).
- If you want 2 glow bracelets TOMORROW - that's LATER, circle this choice. This shows TWO bracelets. It also shows that you will have to sleep for ONE night before tomorrow comes. Tomorrow is WEDNESDAY/FRIDAY (point to the 2 glow bracelets sheet and each component)

Remember, you are choosing to get 1 glow bracelet TODAY OR 2 glow bracelets TOMORROW - so you should only circle one choice. You will get your glow bracelet in an envelope on the day you decide to get it, and get a blank envelope on the other day. The choice you make is up to you.

If you have a question, raise your hand and I will answer you privately. When you are done, turn your paper over, and wait quietly. We will collect your sheets only after everyone is finished. Okay, GO.

Okay, I will collect the activity sheets now. Remember, your decision was private! Please do not talk about your decision to anyone. (collect activity sheets and distribute envelopes)

Thanks for doing the activity today. Let's go back to class.