5-2019

An Investigation into the Structure of Self-Control

Parker A. Dreves
East Tennessee State University

An Investigation into the Structure of Self-Control

A dissertation

presented to

the faculty of the Department of Psychology

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Arts in Psychology

---

by

Parker A. Dreves

May 2019

Ginette C. Blackhart, Ph.D., Chair

Diana Morelen, PhD

Eric Sellers, PhD

Matthew McBee, PhD

Keywords: self-control, impulse inhibition, delay of gratification, willpower, measurement

ABSRACT

An Investigation into the Structure of Self-Control

by

Parker A. Dreves

Self-control has been measured using a variety of methods including self-report measures, cognitive inhibition tasks, delay discounting and delay of gratification tasks, and persistence and willpower tasks. Although these are all theoretically linked to processes involved in self-control, recent evidence has shown that these diverse measurement techniques relate only minimally to one another. Assuming that self-control is a reflective construct, this would indicate that many of these tasks are poor indicators of self-control. The present research challenges the common assumption that self-control is a reflective construct and instead proposes that self-control is a formative construct. Conceptualizing self-control as a formative construct could reconcile some of the inconsistencies in the literature, in particular the fact that many indicators for self-control do not correlate highly. To examine the possibility of a formative model of self-control, this research examines 13 commonly used measures of self-control and investigates indicator intercorrelations, indicator relationships with the theoretical consequences of self-control, and performs a vanishing tetrad test (Bollen & Ting, 2000). Results show that in general, indicator intercorrelations are low and nonsignificant as well as indictor correlations with theorized construct consequences. The results of the vanishing tetrad test suggest a reflective interpretation of self-control, but concerns with uniformly low covariances between indicators limit the interpretation of this test.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Self-control is arguably one of the most important capacities that humans possess. Although animals may possess self-control over extremely short time spans, such as a dog sitting patiently in anticipation of a treat, humans are able to direct behavior toward rewards that are considerably more distant and may not be realized for months, years, or decades. This ultimately allows humans to engage in behaviors that promote positive long-term outcomes, such as better emotional stability, health, and school and work performance (Tangney, Baumeister, & Boone, 2004).

Fundamentally, self-control control dilemmas are characterized by a battle between long-term goals or values and the immediate temptations that thwart those goals or values (Fujita, 2011). Successful self-control is defined as engaging in behaviors that progress one toward a long-term goal or, conversely, abstaining from behaviors that thwart the realization of long-term goals. Although many consider the second part of this definition – abstaining – to be the defining feature of self-control, it should be noted that impulse inhibition is but one strategy among many for furthering long-term goals (Magen & Gross, 2010). Values are included in this definition to highlight the fact that not all self-control dilemmas involve temporally distant goals. For example, one might value animal welfare, honesty, or fairness, which are not necessarily long-term goals, but do provide individuals with reasons to control behavior. Like long-term goals, values weigh on decision-making processes.

There are two main types of self-control: inhibitory self-control and initiatory self-control (de Ridder, de Boer, Lugtig, Bakker, & van Hooft, 2011). Inhibitory self-control refers to

impulse inhibition or abstaining from behaviors that thwart long-term goals, such as resisting eating a tempting slice of cake that would thwart a weight loss goal. By contrast, initiatory self-control refers to enacting some behavior that promotes a long-term goal, such as going to the gym to further one's weight loss goal. In this example, both behaviors are considered self-control because they both promote the attainment of a long-term goal, but one involves inhibiting behaviors where the other involves initiating a behavior. Factor analysis suggests that inhibitory and initiatory self-control are separable constructs and each form of self-control relates differently to various goal-related behaviors. Inhibitory self-control is more predictive of the frequency of undesired behaviors, such as smoking cigarettes or drinking, whereas initiatory self-control is more predictive of desirable behaviors such as hours of study (de Ridder et al., 2011). This highlights the fact that self-control is a multidimensional construct not just limited to the inhibition of impulses.

Strategies for Exercising Self-Control

Self-control can be subdivided further into specific strategies for goal-pursuit (Fujita, 2011; Magen & Gross, 2010). These include situation selection, situation modification, cognitive change, distraction, impulse inhibition, or initiatory strategies such as forming implementation intentions and actively engaging in behaviors that lead to goal-progress. Situation selection refers to avoiding situations where temptations are likely. An example of situation selection would be a recovering alcoholic avoiding going to the bar altogether, or a dieter avoiding the snack food isle at the supermarket. Situation modification refers to changing something about a situation to make self-control more likely, such as bringing a friend with you to the gym to make sure you exercise. Cognitive change, also referred to a reconstrual (Fujita & Carnevale, 2012), refers to reframing a temptation to make it seem less appealing. As an example, a dieter may counter the

temptation of unhealthy food by thinking about the negative health effects of eating poorly, or how they might feel after overeating. Implementation intentions, colloquially known as planning, refer to forming plans for remediation if one does indulge (Gollwitzer, 1999). An example of an implementation intention would be "if I snack too much today, I will go to the gym." Finally, impulse inhibition (sometimes called response modulation; see Magen & Gross, 2010) refers to the effortful suppression of impulses (Baumeister, 2002; Diamond, 2013) and, in some models of self-control, is considered a last resort after other strategies for self-control have failed (Fujita, 2011).

The variety of strategies available for exercising self-control poses challenges for researchers seeking to measure this construct. First, it produces a lack of clarity in terms of which strategy is being assessed, since all of these strategies are grouped under the umbrella term "self-control." For example, the Stroop task, the go/no-go task, delay of gratification tasks, the cold pressor task, food taste tests, and persistence tasks have all been grouped under the label "self-control" (Duckworth & Kern, 2011; Hagger, Wood, Stiff, & Chatzisarantis, 2010). However, each of these tasks may allow or disallow the use of different self-control strategies. For example, distraction from the tempting stimulus reliably increases delay times on delay of gratification tasks (Mischel, Ebbesen, & Raskoff Zeiss, 1972), but distraction cannot be used on measures of cognitive inhibition such as the Stroop task. The fact that different goal pursuit strategies may be used on different tasks muddies the interpretation of what processes are at work for any given test of self-control. Second, isolating components of self-control decreases the ecological validity of research using these measures since in everyday life individuals use a variety of inhibitory and initiatory strategies for goal pursuit to meet situational demands.

Although all of these strategies are important subcomponents of self-control, considering any one of them to be an indicative test of global self-control may be an error.

Given all of the different ways that self-control may be exercised, it is perhaps unsurprising that self-control has been assessed using seemingly disparate measures. To be sure, measuring self-control has been met with difficulty and experts in the field still hotly debate which types of tasks do and do not utilize self-control resources (Baumeister & Vohs, 2016). As a result, self-control has been assessed using a wide variety of measures including self-report measures, attention and response inhibition tasks, delay of gratification and delay discounting tasks, and persistence and willpower tasks. The fact that each of these measures assess different subcomponents of self-control may, in part, explain the lack of comparable results among labs using different measures (Blázquez, Botella, & Suero 2017; Carter, Kofler, Forster, & McCullough, 2015; Lurquin & Miyake, 2017; Monterosso & Luo, 2013). The following section reviews each form of measurement, the rationale for using them, and which particular strategies of goal pursuit each measure may assess. The main focus of this paper will be behavioral measures of self-control that are frequently used in laboratory experiments, rather than self-report measures.

Approaches to Measuring Self-Control

In a discussion of measurement, an important distinction is the difference between trait and state self-control. State self-control refers to self-control at a specific time -- specifically at the time of measurement. By contrast, trait self-control refers to self-control in general, or an average across the person's life. This is relevant to a discussion of measurement for three reasons. First, whether a researcher is interested in state or trait self-control will determine the choice of measures. State self-control is most commonly assessed with behavioral measures,

such as in ego-depletion paradigms where self-control is measured following an exhausting task (Baumeister, Bratslavsky, Muraven, & Tice, 1998). Trait self-control is most commonly assessed via self-report measures that allow individuals to reflect upon their behavior across a variety of circumstances and times. The present research is primarily interested in examining measures of state self-control as they are the most dubious in the sense that many of them lack formal validation (Lurquin & Miyake, 2017).

Second, trait self-control does not necessarily predict state self-control. This is because there are many idiosyncratic situational factors that influence state self-control such as the proximity to temptations, intoxication, fatigue, and the presence of others (Hofmann, Baumeister, Förster, & Vohs, 2012). Third, and perhaps most importantly, in many experimental paradigms state self-control typically only refers to one specific strategy for exerting self-control, which is determined by the researcher's choice of measures. For example, a researcher using the go/no go task as a measure of self-control is equating state self-control with cognitive inhibition, whereas a researcher using the cold pressor task is equating state self-control with pain tolerance. Still other tasks may assess behavioral inhibition, attention, or preference for delayed rewards, but all of these are frequently grouped under the term self-control.

Constructs such as attention, inhibition, initiation of behavior, distraction, pain tolerance, and delay of reward are distinct and separable from one another, so using the blanket term "self-control" can be problematic when discussing interpretations of experiments using different measures. Researchers should be aware of these assumptions and conscious of the fact that self-control is often operationally defined in different ways within different experimental paradigms. In everyday life, people use a variety of strategies for self-control, including impulse inhibition, situation selection, distraction, cognitive change, and counteractive evaluations (Fujita, 2011;

Fishbach, Zhang, & Trope, 2010; Magen & Gross, 2010). Although each of these may be important subcomponents of self-control, considering any one of them in isolation will fail to fully encompass the construct.

In discussions of measurement, there are typically three categories that measures of state self-control fall into. These are executive function/impulse inhibition, delay discounting/delay of gratification, and persistence/willpower tasks. Each of these measures may assess different components of self-control. For example, impulse inhibition measures assess the ability to override an automatic response, whereas delay of gratification measures assess attention toward distal rewards. The assumptions for using each type of measurement are outlined below.

Executive function and response inhibition tests of self-control include tasks such as the Stroop task (Stroop, 1935), the go/no go task (Nosek & Banaji, 2001), the stop signal task (Verbruggen & Logan, 2008), and the flanker task (Eriksen & Eriksen, 1974). The rationale for using these as measures of self-control is that these tasks involve inhibiting an automatic or prepotent response, which many consider a defining feature of self-control (Baumeister, 2002; Diamond, 2013). Consider, for example, the Stroop task. In this task, participants are presented with the words of colors. On some of the trials the words are printed in the same color of ink as what the word says (e.g., GREEN printed in green ink). On other trials, the words are printed in a different color of ink than what the words says (e.g., GREEN printed in red ink). The participants are tasked with correctly naming the font color and must ignore what the word says. This elicits competition between language and visual systems, such that participant must inhibit the impulse to read the word and instead correctly name the color of ink in which the word is printed. This reliably produces an interference effect, where individuals take longer to correctly name colors on trials with colors that do not match the text (Stroop, 1935). The interference score is the

average time it takes a participant to override an automatic response and is thus considered a measure of impulse inhibition or self-control.

The go/no go is another frequently used measure of impulse inhibition. In the go/no go task, participants respond to a certain stimulus as fast as possible. As an example, a participant might be told to press the left or right arrows keys on a keyboard as fast as they can when they see a stimulus appear. However, on a certain number of trials (the no go trials), participants will receive some cue to inhibit their responses. For example, if the stimulus is outlined in red, they are to withhold their response. The reasoning behind this task is that on the majority of trials (go trials) the participant can quickly and automatically respond when they see the target stimulus. However, on a minority of trials (no go trials), the participants will have to inhibit their automatic impulse to respond. This, like the Stroop, is considered to be a measure of self-control because it assesses impulse inhibition.

Within the category of impulse inhibition, however, there are distinctions that are seldom acknowledged. Friedman and Miyake (2004) identified three subcomponents of cognitive inhibition including prepotent response inhibition, resistance to distractor interference, and resistance to proactive interference. Prepotent response inhibition refers to acts of overriding automatic responses, such as in the Stroop task and go/no go task. Resistance to distractor interference refers to the ability to ignore information that is irrelevant to the task at hand, such as in the Eriksen flanker task (Eriksen & Eriksen, 1974). Resistance to proactive interference refers to the ability to resist memory intrusions from previously cued information, such as in cued recall tasks. The fact that these are rarely distinguished from one another and are often all grouped under the term self-control or inhibition may, at least in part, explain the limited

convergent validity even among these executive function tasks (r = .15; Duckworth & Kern, 2011).

Delay of gratification tasks involve presenting individuals with a choice between a smaller immediate reward or a larger delayed reward. This paradigm was pioneered in the early 1960s when Walter Mischel and colleagues (1958; 1961) investigated individual differences in temporal discounting among children. In a seminal work, researchers offered children a choice between a one-cent candy and a ten-cent candy that was markedly higher in quality and presumably more desirable. If they chose the smaller candy, children were able to receive it right then and there, representing immediate gratification. If children wanted the larger and more expensive candy, they were told that they would have to wait one week to receive it. Although in this study the outcome was binary (e.g., smaller candy or larger candy), this paradigm underwent subsequent modifications and a continuous measure was eventually developed. Later, the task was modified such that during the delay period children were able to signal the experimenter if they wanted to eat the smaller and less desirable reward or in some cases had the food right in front of them and could eat it at any time (Mischel & Ebbesen, 1970). This thus introduced delay time as a continuous measure, with longer delay times being an indication of higher self-control.

Although such tasks have historically been used with children, these can be modified for adults, which usually involves using money as the focal reward and employing longer delay times (weeks or months). Generally, however, adjusting amount delay discounting tasks are used with adults. Adjusting amount procedures present individuals with either hypothetical or real choices between immediate and delayed rewards (e.g., would you rather have $7 right now or $10 in one week) with the goal of determining individuals' rates of delay discounting, or their indifference point (Frye, Galizio, Friedel, DeHart, Odum, 2016; Koffarnus & Bickel, 2014;

Richards, Zhang, Mitchell, & de Wit, 1999). As individuals are presented with increasingly large delay times, at some point the individual will begin to prefer the smaller immediate reward over the larger delayed reward due to the temporal distance of the delayed reward. The point at which the individual switches to preferring the smaller immediate reward over the larger delayed reward is known as the indifference point, since the immediate and delayed reward are equal in terms of their subjective value. The indifference point (V) is calculated by dividing the amount of the delayed reward (A) by the product of delay time (D) and a scaling factor (k) that represents the individual's rate of delay discounting ($V = A/(1 + Dk)$; see Odum, 2011). Individuals with lower rates of delay discounting (k) and higher indifference points are considered to have higher self-control. Often just k, or an individual's rate of delay discounting, is used as a measure of self-control (Kirby, Petry, & Bickel, 1999).

Finally, persistence and willpower tasks include diverse tasks such as the cold pressor, holding a handgrip, drinking bitter tasting drinks, persisting on impossible puzzles, completing difficult anagrams or math problems, and food "taste tests" (Baumeister et al., 1998; Hagger et al., 2010; Vohs et al., 2014). In the cold pressor task, participants are instructed to hold their hand in ice water (usually maintained between 35 and 40 degrees Fahrenheit) for as long as they can (Mitchell, MacDonald, & Brodie, 2004). This is intended to simulate physical pain, and the reasoning provided for using this as a measure of self-control is that it requires impulse inhibition to endure the unpleasant stimulus of the cold water. Similar reasoning is provided for the bitter tasting drink task and the handgrip task in that both require an individual to inhibit the impulse to give up on an unpleasant task. In the taste test task, participants are told that they are to judge the quality of a food item and that they may eat as much of the food item (usually cookies or candy) that they need to in order to accurately judge its quality (Baumeister et al.,

1998). The dependent variable is the amount of food consumed, which ostensibly reflects self-control because people with lower self-control will be less able to inhibit the impulse to continue eating. Finally, on the impossible puzzle task, participants are presented with an extremely difficult (or even impossible) puzzle to solve. The dependent variable is the amount of time they persist before giving up, with greater persistence being associated with higher self-control.

Although these tests are often framed in terms of impulse inhibition (e.g., suppressing the impulse to remove one's hand from ice water), there is evidence that cognitive inhibition (e.g., the Stroop task) is distinct from behavioral inhibition (e.g., suppressing a triggered motor response; see Engelhardt, Nigg, Carr, & Ferreira, 2008). Thus, persistence and willpower tasks blur the lines between behavioral inhibition and cognitive inhibition. Moreover, such tasks may confuse initiatory self-control and inhibitory self-control. For example, the food taste test may rely more on inhibitory self-control (resisting the food), whereas the puzzle task may rely more on initiatory self-control (actively trying to solve a puzzle). Such nuances are rarely acknowledged but may have implications for the interpretation and generalization of results. In subsequent analyses in this paper, persistence and willpower tasks will be split into two categories. These will be tasks that require inhibiting an action (e.g., not eating unhealthy foods, cold pressor) and tasks that require initiating an action (e.g., eating healthy foods, impossible puzzle, anagrams). For a summary of self-control measures, see Table 1.

Table 1.

A Summary of Measures of State Self-Control

Measure	***Examples***	***Factors Assessed***
Delay of Gratification/ Delay Discounting Tasks	Monetary or food related delay of gratification, adjusting amount	Behavioral inhibition, attention to rewards, rate of delay discounting
Response Inhibition	Stroop, Go/no go, Stop-signal	Cognitive inhibition, attention, reaction time
Resistance to Distractor Interference	Video-watching attention control, Eriksen flanker task	Cognitive inhibition, attention, reaction time
Inhibitory Persistence/Willpower Tasks	Cold Pressor, Taste Test, Handgrip	Behavior inhibition, resistance to unpleasant stimuli, pain tolerance
Initiatory Persistence/Willpower Tasks	Impossible Puzzle, Bitter Drink, Anagrams	Behavior initiation, achievement motivation

The Structure of Self-Control

The preceding discussion illustrates the fact that self-control is a broad construct and encompasses diverse behaviors. However, substantial inconsistencies in the self-control literature raise questions about the dimensionality of the self-control construct. A central concern of the present research is whether self-control is more appropriately conceptualized as a reflective construct or a formative construct. Reflective constructs are constructs that are proposed to exist independently of measurement and can be estimated though their effect on a variety of indicators. In other words, causality flows from the construct (self-control) to the indicators (see Figure 1). By contrast, a formative construct is a construct that is defined by a particular set of indicators but has itself no causal influence on its indicators. Simply put, in a formative model

the latent variable is the aggregate of several indicators (see Figure 2). A common example of a formative construct is socioeconomic status. SES does not cause education level or income but is instead defined as the collection of such indicators.

Borsboom et al. (2003) discuss the theoretical implications of conceptualizing of a construct as reflective or formative. Entity realism is the position that there are real constructs such as intelligence, dominance, or aggression that create variance in behavior. Reflective models require entity realism. To say that a construct is reflective is to say that it literally exists and has causal influence over indicators. By contrast, a formative model would imply a constructivist view of a psychological construct. Constructivism holds that there are not real constructs driving behaviors but rather that constructs are just ways of classifying behaviors and that theoretical constructs have no causal power over behaviors. Although the majority of psychological constructs are conceptualized as reflective constructs, there is no a priori reason why this should be the case (Borsboom, Mellenbergh, & van Heerden, 2003).

Self-control, like many other psychological concepts, is typically conceptualized as a reflective construct. This is apparent both in the way that researchers discuss self-control as well as the way researchers measure self-control. First, many researchers discuss self-control as a single construct that, if sufficiently developed, leads to better performance on a variety of tests such as executive function tasks, delay of gratification tasks, and persistence measures (Baumeister, Vohs, & Tice, 2007; Hagger et al., 2010). Second, the fact that researchers try to manipulate self-control, such as in ego-depletion paradigms, implies that they think of it as a reflective construct. For example, ego-depletion researchers presume that exhaustion in one sub-domain (e.g., resisting food) will produce changes in another sub-domain (e.g. cognitive inhibition) by virtue of them sharing self-control as a common cause (Baumeister et al., 1998).

On the other hand, with formative constructs, a change in one indicator would not necessarily be expected to produce changes in every other indicator. Third, the fact that researchers use the various indicators of self-control interchangeably is predicated on conceptualizing self-control as a reflective construct. In theory, indicators of a reflective construct can be used interchangeably, whereas this is not the case for formative constructs.

I suggest here that one possible explanation for the limited convergent validity among measures of self-control is that self-control is not a reflective construct and is instead a formative construct. In this view, "self-control" would just be a label for grouping goal-oriented behaviors. These behaviors would include things like impulse inhibition, distraction from tempting stimuli, goal initiation, or any of the strategies discussed previously. It should be noted that a constructivist approach to self-control leads to some interesting conclusions. Since formative constructs are not thought to determine measurements, but are rather just a summary of measurements, one is left with the proposition that changing the indicators literally changes what the construct is. In a formative model a factor estimate is merely an index statistic (van der Maas, Kan, & Borsboom, 2014). As an example, the Dow Jones Industrial Average is an index statistic that is defined by the aggregate of the thirty trade reports (indicators) that make it up. The Dow is not a reflective construct that causes the trade reports that make it up, but rather a summary of indicators. If one were to change the indicators that comprise the Dow Jones Industrial Average, the meaning of that construct would change accordingly. If self-control is conceptualized as a formative construct, a similar claim must be made. If the indicators that define self-control are changed, the meaning of the self-control construct is changed. Therefore, if self-control is a formative construct, there will be a need for researchers collectively decide which indicators comprise self-control. Just as sociologists can all agree on what indicators comprise SES,

economists can all agree on what indicators comprise the Dow, and climatologists can all agree on what indicators comprise the Air Quality Index, psychologists would need to be able to agree on what indicators comprise self-control. This is presumably an unpalatable position and very few people would want to claim that how self-control is defined changes based on a researcher's choice of measures.

There is another possibility that avoids the conclusion that self-control is defined by the indicators that researchers choose to measure it with. It could be the case that self-control is a hybrid construct containing elements of both reflective end formative models. In this view, self-control itself could be formative, but be made up of reflective constructs such as attentional control, inhibition of undesired behaviors, or initiation of desired behaviors (see Figure 2). This avoids the unpalatable conclusion that self-control is literally defined by a researcher's choice of measures by positing that self-control, although still formative, is a collection of sub-processes important for goal pursuit that are themselves reflective. This would mean that researchers would still need to decide on what the important sub-domains are (e.g., cognitive inhibition, behavior initiation, etc.), but that indicators would still be interchangeable within a sub-domain. This would somewhat relax the requirements for what indicators need to be used, since indicators of the same reflective construct are interchangeable. For example, if both the Stroop and the go/no go assess inhibition, they would be interchangeable

This discussion cuts to the heart of what self-control is. As evidenced by the plethora of measures that assess self-control (Duckworth & Kern, 2011; Hagger et al., 2010) and the numerous discussions about different strategies for exerting self-control (Fujita, 2011; Magen & Gross, 2010), it seems that self-control is a broad construct that encompasses a wide variety of behaviors. As argued previously, what unites these behaviors is that they are all ways of pursuing

long-term goals. Inhibiting impulses, initiating desirable behaviors, being selective about the situations one puts themselves in, reconstruing the meaning of a tempting stimulus, and distracting oneself when tempted are all strategies that people use to progress toward a long-term goal. This means, of course, that in a formative model of self-control there would be a considerable number and variety of indicators that should be included to comprise self-control. As noted, one way around this is to propose a hybrid model where self-control is a formative construct that is itself made up of reflective constructs. This avoids the conclusion that self-control is defined by a researcher's choice of measures and means that measures would be interchangeable within a given sub-domain (Figure 2). The goal of this paper is not to provide an exhaustive list of every sub-domain within self-control nor to provide an exhaustive list of all of the indicators that could be used to assess reflective sub-domains, but rather to examine relationships between some commonly used indicators and determine whether the data favor a reflective model of self-control or a formative model. If the data should favor the latter, it will be the task of future researchers to flesh out a complete taxonomy of self-control (e.g., what reflective constructs need to be included in a complete model of self-control) or develop a standardized set of indicators (much like SES or the Dow) that comprise self-control. The unspoken assumption that self-control is a reflective construct has, to my knowledge, gone unchallenged. Challenging this assumption constitutes the major contribution of this paper.

Coltman et al. (2008) provide detailed guidelines for determining whether a construct is best conceptualized as reflective or formative. The way in which each of these guidelines inform a conceptualization of self-control will be examined in turn. The first theoretical consideration presented by Colman et al. (2008) concerns the nature of the construct. This asks whether a construct exists or whether a construct is formed. In more precise terms, it asks whether a

construct is defined by its indicators (formative) or whether indicators are a reflection of a real underlying construct (reflective). The corresponding empirical consideration is to examine indicator intercorrelations. For a reflective construct, where all indicators share a common cause, significant intercorrelations would be expected among indicators. By contrast, indicators of a formative construct do not necessarily have any expected pattern of intercorrelations. This is because in a formative construct, indicators are not hypothesized to share a common cause, and thus will not have any expected pattern of correlations.

The second theoretical consideration concerns the direction of causality from the latent factor to the indicators. If variance in the latent construct theoretically creates variance in the indicators, then it is a reflective construct. If the indicators cause changes in the construct, then it is formative. In the context of self-control, the question would be whether changes in self-control produce changes in things like impulse inhibition, planning, and goal initiation or whether changes in these factors cause changes in self-control. This is also noted by Edwards and Bagozzi (2000) who suggest considering temporal precedence to determine whether a construct is reflective or formative. If a change in the construct is proposed to cause a change in its indicators, then it is reflective. This is the assumption in ego-depletion paradigms, where self-control is manipulated through an exhausting task and then measured using some indicator. However, one could take an alternative position and suppose that changes in things like impulse inhibition and persistence instead cause changes in self-control. This would be consistent with a formative model.

The next empirical consideration suggested by Coltman et al. (2008) is to examine indicator relationships with construct antecedents and consequences. Assume for a moment that self-control is a reflective construct, and that changes in self-control are expected to produce

changes in a variety of domains such as school or work performance, substance use, financial stability, and health behaviors. Since under a reflective model each indicator is a reflection of self-control, one would expect each individual indicator to be able to predict the same outcomes as the latent construct itself. In simpler terms, if self-control theoretically leads to a particular outcome, and a particular indicator assesses self-control, then one would expect that indicator to also predict that particular outcome. By contrast, if self-control were a formative construct, this would not need to be the case. It could be the case that certain indicators only predict certain types of outcomes and not others. In fact, this already appears to be the case based on the findings of de Ridder et al. (2011) who found that inhibitory and initiatory self-control predicted different behaviors.

The final theoretical consideration is the treatment of measurement error. In reflective models, factor analysis can be used to identify and extract out the measurement error of the construct. This is possible because in reflective models, observed scores are indicators of a latent factor, and thus measurement error in the latent factor can be determined by examining the variance in the factor not accounted for by the indicators. In formative models, it is not possible to account for measurement error in the latent factor. This is because each indicator is not caused by the same factor, but rather each indicator has its own set of factors that influence it and bring with them their own unique sources of variance. There are no constraints on the covariances among indicators in a formative model.

The fact that reflective models and formative models treat sources of error differently allows for a test of model specification called the vanishing tetrad test (Bollen & Ting, 2000). The tetrad test works by examining the differences between the products of pairs of covariances.

For any four indicators, there are six covariances and it is possible to form three sets, or tetrads, that may be tested. For example, for the indicators a, b, c, and d, the tetrads are as follows:

$$T_{abcd} = \sigma_{ab}\sigma_{cd} - \sigma_{ac}\sigma_{bd}$$

$$T_{acdb} = \sigma_{ac}\sigma_{db} - \sigma_{ad}\sigma_{cb}$$

$$T_{adbc} = \sigma_{ad}\sigma_{bc} - \sigma_{ab}\sigma_{dc}$$

In these equations, σ is equal to the population covariance of the two variables that are indexed below it. Vanishing tetrads, implied by reflective models, mean that $T_{abcd} = 0$. By contrast, tetrads not equal to zero would imply that each set of covariances have unique sources of variance and would thus be most consistent with a formative model.

Based on the theoretical and empirical considerations presented above, the following section will critically evaluate two different models of self-control. The first model (Figure 1) is a reflective model of self-control and is arguably the model currently preferred by researchers in the field. Under this model, self-control is a psychological construct that has causal influence over a diverse range of behaviors including impulse inhibition, planning, attention deployment, goal initiation, delay of gratification, and persistence. The second model (Figure 2) is a formative model of self-control.

Based on which model most closely resembles the true structure of self-control, different patterns in the data are expected. If self-control is a reflective construct, with all indicators sharing self-control as a common cause, one would expect to see substantial correlations both between different individual indicators as well as between sub-domains. By contrast, in the formative model one would not necessarily expect correlations between the different domains.

However, one would still expect correlations among indicators of the same sub-domain since these are still represented as reflective constructs.

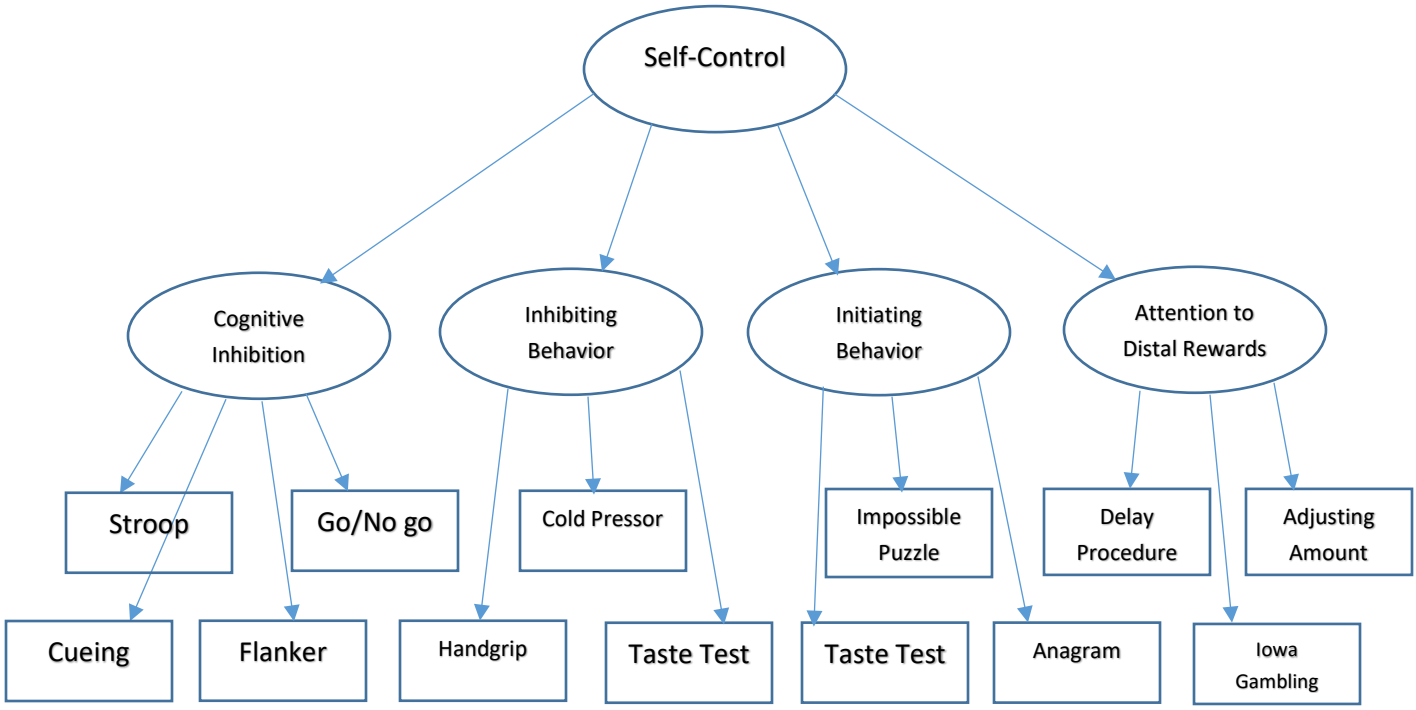

Figure 1. Reflective Model of Self-Control

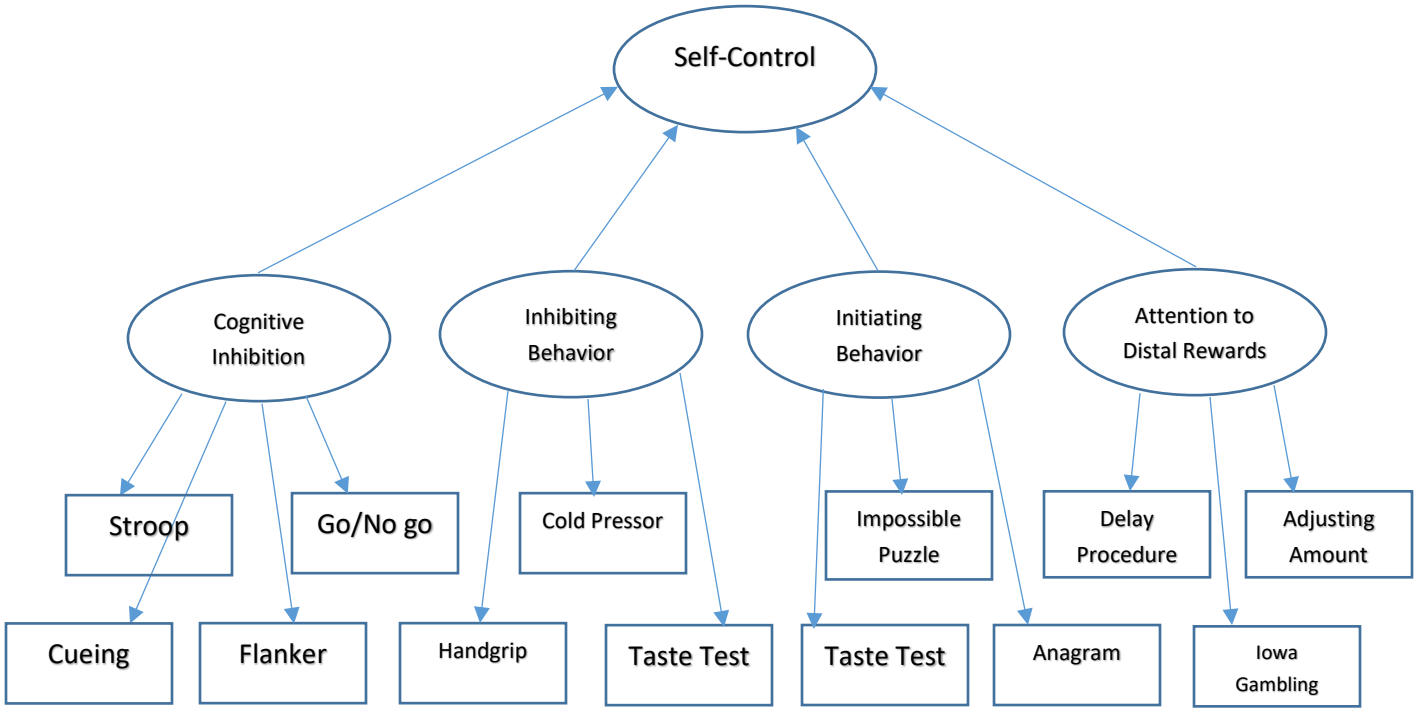

Figure 2. Hybrid Model of Self-Control

Previous Evidence for Convergent Validity

The first empirical consideration suggested by Coltman et al. (2008) is to examine indicator intercorrelations. This is essentially a question of convergent validity. If indicators do not correlate, there is no way for them to be meaningfully assessing the same reflective construct. Under a formative model, however, there would be no expected pattern of correlations between the different indicators. By and large, this is what the available data show. In a meta-analysis of self-control measurement, Duckworth and Kern (2011) examined relationships between self-report measures, impulse inhibition measures, and delay of gratification tasks. Low correlations were observed between each type of measurement ranging from $r = .10$ to $r = .15$. Although these correlations were significant, the small magnitude of these relationships casts

doubt on the comparability of such measures. In fact, the small magnitude of these relationships may well be interpreted as effectively meaningless due to what Paul Meehl (1990) called the "crud factor". The crud factor is a phenomenon in the social sciences such that, given sufficient sample size, everything correlates with everything else. This is due to the fact that in the social sciences, measured traits are affected by shared background characteristics such as upbringing, genetics, social and cultural environment, nutrition, and numerous other unmeasured variables. Therefore, the pattern of low correlations observed between behavioral measures of self-control cannot safely be interpreted as evidence of convergent validity, but rather noise or "crud" that is produced by shared background variables. Indeed, this pattern of low or non-existent correlations between various measures of self-control would be most consistent with a formative model of self-control, rather than a reflective model.

In addition to the meta-analysis conducted by Duckworth and Kern (2011), numerous other researchers have taken up the task of examining the convergent validity of self-control measures, with relatively similar findings. Reynolds et al. (2006) examined relationships between four behavioral measures of self-control, which included a stop signal task, the go/no-go task, a computerized delay discounting procedure, and the balloon analogue risk task (which involves earning money to blow up a balloon that could pop at any time and result in a loss of earnings). The researchers also administered three self-report measures of impulsivity, including the Barratt Impulsiveness Scale (Patton et al., 1995), the I7 (Eysenck et al., 1985), and the Multidimensional Personality Questionnaire (Patrick, Curtin, & Tellegen, 2002). Although correlations between self-report measures were high, correlations between self-report measures and behavioral measures were low and largely non-significant. Moreover, the various behavioral measures did not correlate with one another and loaded on two separate factors labeled by the

researchers as impulsive disinhibition (stop signal task and go/no go) and impulsive decision making (delay of gratification and risk taking), highlighting the danger of grouping these measures under the global term "self-control."

In a similar study, researchers administered four self-report measures of impulsivity and five behavioral measures. The four self-report measures included the Barratt Impulsiveness Scale (Patton et al., 1995), the I7 (Eysenck et al., 1985), the Dickman Impulsivity Inventory (Dickman, 1990), and the Wender Utah Rating Scale (Ward, Wender, & Reimherr, 1993). The five behavioral measures included a delay of gratification choice task, a delay discounting task in which participants made hypothetical choices between monetary rewards, a contingent delay discounting task in which participants had an opportunity to receive the rewards they selected, a response inhibition task that required participants to wait a set amount of time before responding to a stimulus, and a digit span recognition task. The results showed a pattern of low and non-significant correlations between self-report measures and behavioral measures, as well as no significant relationships between the various behavioral measures (Lane, Cherek, Rhodes, Pietras, & Tcheremissine, 2003). Like Reynolds et al. (2011), these researchers found that behavioral tasks loaded onto two factors that they labeled as response inhibition and delay of reward.

Numerous other studies have reported similar findings. White et al. (1994) examined relationships between measures, such as the Stroop task, the Trail Making Test, a card playing task, a circle tracing task, a delay of gratification task, and self-reported impulsivity, and found that performance on these measures was generally unrelated. In other studies, self-reported self-control has been unrelated to performance on the go/no go and delay of gratification tasks (Crean, de Wit, & Richards, 2000; Fine, Steinberg, Frick, & Cauffman, 2016). Taken together,

the results of these studies suggest that measures of self-control relate only minimally to one another, if at all, which would be most consistent with a formative model of self-control rather than a reflective model.

A final piece of evidence comes from research by Steimke et al. (2016). In this study, researchers administered a spatial attentional cueing task, the Stroop task, an unsolvable anagram task, and a delay of gratification task using chocolate. The cueing task required that participants identify a target letter amidst either aversive (disgusting), tempting (erotic), or neutral distractors. Using eye-tracking data, the researchers were able to examine the effect of aversion, temptation, and neutral distraction on task performance. The results of this research showed that the ability to endure aversion, resist temptations and ignore neutral distractions are independent of each other. This is especially relevant to the present discussion because it highlights the fact that self-control is not necessarily a singular construct and avoiding something is not the same as engaging with something. However, the researchers did find that although the unsolvable anagram task did not correlate with the cueing task, the Stroop and the delay of gratification task did correlate. Still, the fact that the different strategies are separable would be more indicative of a formative model.

To summarize, a growing body of research suggests that various behavioral measures of self-control only relate minimally to one another. Under a reflective model of self-control, where the different indicators are thought to reflect variance in a single underlying construct, the different measures would be theoretically interchangeable. This appears to be the dominant conceptualization of self-control, as evidenced by studies that use these different measures interchangeably and interpret the results indiscriminately as levels of self-control (Baumeister et al., 1998; Hagger et al., 2010; Vohs et al., 2014). The assumption that such diverse tasks may be used interchangeably as measures of self-control has gone largely unchallenged. However, the

available data suggests that this may not be the case, which casts doubt on the comparability of results across studies using different measures of self-control. The low correlations reported between the various measures, which can hardly be interpreted as more than "crud" (Meehl, 1990), would be most consistent with a formative model of self-control. Although each of these may be processes involved in the pursuit of long-term goals, it is likely that none of them on their own capture the entire construct of self-control. Thus, results using different measures may be incomparable and researchers should be wary of using such measures interchangeably as tests of global self-control.

Absent from the discussion thus far is the use of persistence and willpower measures to assess self-control. These measures, such as the cold pressor or handgrip task, are frequently used in ego-depletion paradigms. The rationale generally provided is that these tasks involve impulse inhibition. However, such measures differ from the computerized cognitive inhibition tasks in a few important ways. First, persistence and willpower tasks may assess behavioral inhibition rather than cognitive inhibition due to the fact that many of these tasks are behaviorally based (e.g., handgrip or cold pressor). This is an important distinction as research suggests that behavioral inhibition and cognitive inhibition are distinct constructs and should be considered separately in research (Engelhardt et al., 2008; Harnishfeger, 1995). Second, various strategies of self-control may be used in persistence and willpower tasks, such as distraction, planning, or cognitive change. By contrast, cognitive inhibition tasks limit participants to the use of only one strategy and thus may underestimate self-control. For example, thinking about something else may be an effective strategy for persisting on the cold pressor task, and in my own research I frequently observe participants trying to distract themselves during this task (e.g.,

foot tapping, talking, looking elsewhere in the room). On the other hand, distraction would not be an effective strategy for improving performance on the go/no go task or Stroop task.

For these reasons, persistence and willpower tasks may more closely resemble real-life acts of self-control because they can involve both inhibition and initiation, allow the use of multiple strategies of goal-pursuit, and are closer in kind to the self-control challenges that people face on a daily basis. Thus, such measures may theoretically appreciate higher ecological validity than computerized inhibition tasks. However, despite some conceptual advantages, persistence and willpower tasks have not been validated as indicators of self-control. To my knowledge, only one study has directly examined the relationship between performance on such tasks and self-reported self-control. Schmeichel and Zell (2007) conducted two studies and found that participants who scored higher on the Self-Control Scale were more able to refrain from blinking during a two-minute period as well as persisted longer on the cold pressor task. Although these findings show that persistence and willpower tasks may assess important components of self-control, more extensive research is needed to draw sound conclusions (Lurquin & Miyake, 2017). This will be another major contribution of the current work.

Previous Evidence for Predictive Validity

When differentiating between a reflective and a formative construct, Coltman et al. (2008) also suggest examining indicator relationships with the theoretical antecedents and consequences of a construct. This essentially concerns predictive validity. If self-control were a reflective construct, one would expect the indicators to predict theoretically implicated outcomes with similar efficacy due to them all sharing a common cause. If self-control were a formative construct there would not necessarily be any expected pattern in terms of what indicators predict what outcomes. In more concrete terms, under a reflective model of self-control one would

expect impulse inhibition, delay of gratification, and persistence indicators to predict similar consequences (e.g. GPA, drug use, health behaviors, spending) by virtue of sharing self-control as a common cause. If self-control were a formative construct, one might instead expect certain domains to predict certain behaviors and other domains predict different behaviors. For example, perhaps impulse inhibition would be a better predictor of drug use and delay of gratification would be a better predictor of saving money. The following section will examine each type of measure in terms of predictive validity and consider whether the data support a reflective or a formative model of self-control.

Regarding the predictive validity of cognitive inhibition tasks, there is mixed evidence. Sharma, Markon, and Clark (2014) conducted a meta-analysis examining relationships between self-report measures of impulsivity, laboratory tests of impulsivity, and daily life behaviors. The behavioral measures included in the analysis were the Stroop Task, the go/ no go task, the stop signal task, the Iowa Gambling Task, and the Wisconsin Card Sort Task. The daily life outcomes assessed included measures of drug, alcohol and tobacco use as well as aggression, gambling, and sex. The meta-analysis revealed that, on the whole, behavioral measures were weakly or insignificantly related to daily life behaviors. One exception was the Stroop Task, which correlated with alcohol use, aggression, and gambling. Moreover, behavioral measures of impulsivity loaded on different factors labelled inattention, inhibitory dyscontrol, impulsive decision making, and set shifting.

Other researchers have reported similar findings. White et al. (1994), discussed previously, found that scores on cognitive tasks were either weakly or insignificantly related to parent and teacher ratings of impulsivity as well as delinquency at ages ten and thirteen. In one study of adolescent criminal behavior, the go/no go did not reliably predict self-reported self-

control or long-term self-reported offending, but go/no-go times were significantly correlated with short-term self-reported offending (Fine, Steinberg, Frick, & Cauffman, 2016). By contrast, Crean et al. (2000) found no evidence that performance on the go/no-go could distinguish between low-risk and high-risk psychiatric outpatients. Overall, evidence of the predictive validity of cognitive inhibition tasks is sparse.

With regard to delay of gratification tasks, evidence of predictive validity has also been mixed. On one hand, there are researchers who have not had success using delay of gratification tasks to meaningfully predict behavioral outcomes. In the meta-analysis presented above (Sharma et al., 2014), delay tasks were only successful in predicting tobacco use, gambling, and sex. White et al. (1994) found that a delay of gratification task was unsuccessful in predicting delinquency, although it was related to teacher-rated impulsivity and patience/persistence. Similarly, Lane et al. (2003) included two delay discounting tasks in their study and found that it did not relate significantly to any of the other indicators of impulsivity except the Attention/Impulsivity subscale of the Wender Utah Rating Scale (Ward et al., 1993). This is, however, a pattern of correlations one might expect under a formative model of self-control. Some of these tasks may predict some outcomes better than other tasks due to the specific domain tapped by the task and the specific demands of the outcomes of interest.

Showing support for the predictive validity of delay of gratification tasks, Mischel and colleagues (1988; 1989) conducted several follow-up studies of the children included in their original delay of gratification experiments (Mischel & Ebbesen, 1970; Mischel, Ebbesen, & Zeiss, 1972). Children who were better at delay of gratification tasks at ages four and five displayed more positive life outcomes as adolescents, such as higher academic, verbal, and social competence and were more rational and attentive. For example, children with longer delay times

were later found to have higher SAT scores, lower likelihood of obesity, better social skills, and fewer risk taking behaviors, such as gambling and drug use (Mischel, Shoda, & Peake, 1988; Shoda, Mischel, & Peake, 1990; Mischel et al., 1989; Reynolds, 2006). Even as adults, participants who performed well on childhood delay of gratification tests had higher educational attainment, higher sense of self-worth, better responses to stress, and less drug use (Ayduk et al., 2000; Mischel et al., 2011). Reynolds et al. (2004) have also conducted studies examining the predictive validity of delay of gratification tasks. In one study, the researchers were able to show that rates of delay discounting distinguished between smokers and non-smokers, with smokers being significantly higher in delay discounting (Reynolds et al., 2004). In yet another study, longer delay times were associated with better achievement test scores, higher GPA, lower BMI, and less risky behavior (Duckworth et al., 2013).

Other researchers have found similar associations between the ability to delay gratification and outcomes theoretically associated with self-control (Reyna & Wilhelms, 2017; Reynolds, 2006; Watson & Milfont, 2017), suggesting that delay of gratification tasks do have reasonable predictive validity. This could be because in delay of gratification tasks, individuals can sometimes use other strategies besides impulse inhibition, such as distraction from the tempting stimulus, cognitive change, or situation modification, to aid in obtaining the delayed reward. For example, presenting children with toys or another way to distract themselves during the delay period has been shown to increase delay times (Mischel et al., 1972; Peake, Hebl, & Mischel, 2002; Sethi et al., 2000). Indeed, anyone who has observed children trying to resist eating a marshmallow has probably noticed the use of the self-control strategies the children employ (e.g., looking elsewhere in the room, singing/talking to themselves, rocking back and forth in the chair). Individuals who are especially adept at using such strategies may perform

better on delay of gratification tasks, but these differences would be undetectable if using measures of cognitive inhibition.

The studies presented above show that the different measures of self-control inconsistently predict the theoretical consequences of the construct, which would be most consistent with a formative model. Long-term goals may be pursued through a variety of means, such as impulse inhibition, situation selection, situation modification, distraction, forming implementation intentions, goal initiation, and cognitive change. Some of these may be more effective for certain types of goal pursuit than others, and the notion that self-control is a single construct that determines behavior in all of these domains may be inaccurate.

Although all of these strategies may be ways that individuals pursue long-term goals, being adept at using just one or a few of these strategies may suffice. For example, an individual who is adept at avoiding tempting situations and surrounds themselves with other similarly responsible individuals will not have to use impulse inhibition as much as someone who places themselves in situations where temptations are frequent. In this way, self-control would not drive behavior in a variety of domains, but rather behavior one or a few domains would drive self-control. It may be that self-control is additive in the sense that if an individual is sufficiently skilled in one or a few of these strategies, they are labelled as having "high self-control". An analogy can again be taken from economics. The Dow Jones Industrial Average is comprised of trade reports of thirty large U.S. companies. This means that a minority of those companies can dip, but so long as the majority are doing well the Dow will indicate economic growth. In the context of self-control, an individual could be poor as using a few strategies (say, poor impulse inhibition), but as long as they are sufficiently skilled at a some of the strategies of goal pursuit (say, situation selection and goal initiation), then they will appear as having high self-control.

In establishing the construct validity of a psychological test, Cronbach and Meehl (1955) discuss the nomological network. A nomological network is an interlocking system of constructs and observables that are theoretically linked to one another. In order to establish that a measure has construct validity, a nomological network must be developed for the measure. Cronbach and Meehl hold that a rigorous chain of induction is needed to determine the degree to which a construct behaves the way that it should in a nomological network. Given that self-control is not directly observable and has no definite criterion ("gold standard") by which to judge it, self-control must be inferred based on relationships between variables in a nomological network. It should be noted, however, that the logic of establishing nomological validity in this way only works for reflective constructs. For formative constructs, where there is not necessarily any expected pattern of correlations between indicators, the logic of establishing nomological validity (which is based on examining expected patterns of correlations between constructs and observables) breaks down. As Cronbach and Meehl (1955; p. 287) state "if the obtained correlation departs from the expectation, however, there is no way to know whether the fault lies in test A, test B, or the formulation of the construct." The goal of the present research is to examine the possibility that, in the case of self-control, the problem lies with the formulation of the construct. I propose that self-control is not a reflective construct but rather a formative construct. As such, standard procedures used in the psychological sciences to establish construct validity (e.g., examining the nomological network, factor analysis) have been and will continue to be fruitless endeavors.

Current Study

The primary goal of the present study is to determine whether self-control is more appropriately conceptualized as a reflective construct or a formative construct. To do so, three

smaller goals will be set. First, as suggested by Coltman et al. (2008), indicator intercorrelations will be examined. Under a reflective model, high indicator intercorrelations would be expected. Under a formative model, high indicator intercorrelations would only be expected within each reflective sub-domain (e.g., correlations between measures of cognitive inhibition), but not across sub-domains. Second, this research will examine indicator relationships with the theoretical consequences of self-control. If self-control is a reflective construct, all indicators should correlate with the theoretically implicated consequences of self-control such as GPA, health behaviors, spending, and drug use. If self-control is a formative construct, indicators may relate to outcomes differentially. Third, the vanishing tetrad test will be applied to sub-domain factor scores. In order to do this, structural equation modeling will be used to produce a factor covariance matrix (with each factor corresponding to one of the sub-domains identified in Figures 1 and 2). The resulting covariance matrix will be submitted to a vanishing tetrad test. If the resulting tetrads are not significantly different from zero, this will be evidence of a reflective model. If the tetrads are significantly different from zero, this will be evidence of a formative model. The vanishing tetrad test will not be applied to every possible tetrad of indicators for two reasons. First, the number of sets of tetrads for models with n observed indicators is $n!/(n\text{-}4)!4!$ (Bollen & Ting, 2000). This means that for thirteen tests of self-control, there would be 715 sets of tetrads, or 2145 implied vanishing tetrads. Second, given that some measures of self-control share reflective sub-domains (e.g., the Stroop, go/no go, and flanker all being indicators of cognitive inhibition), some covariance tetrads (within a sub-domain) would be expected to vanish, while others (across sub-domains) would not be expected to vanish. This would muddy the interpretation of the results since some tetrads would vanish and others would not. Due to

these reasons, I have decided to reduce the number of indicators and use the factor covariance matrix to perform the vanishing tetrad test.

CHAPTER 2

METHODS

Power Analysis

In order to estimate the number of participants needed to detect relationships between measures, correlations from previous studies of convergent validity were examined. Based on research by Duckworth and Kern (2010), Saunders et al. (2017), Reynolds et al. (2006), Lane et al. (2003), Schmeichel and Zell (2007) and White et al. (1994), correlations were estimated. Table 2 summarizes previous work on convergent validity. Based on the results of previous studies, correlations between different measures range from small and non-significant up to as high as .35. Based on this level of inconsistency, I chose a conservative estimate of effect size and calculated the sample size needed to detect a significant correlation of $r = .20$ between any two given measures. The necessary sample size needed to detect correlations as low as .20 at 80% power is $n = 192$. Therefore, target enrollment is 200 participants. Power analysis was conducted using the R package "pwr" (Champely, 2018).

Participants

Participants were 197 undergraduate students recruited from a public university in the southeastern United States. However, data from 6 participants was discarded due to errors in data collection or non-compliance with study instructions, leaving a final sample of 191. The majority of participants were female (61.3% female, 32.5% male, .5% transgender, and 5.7% declined to identify). The majority of participants were also Caucasian (80.1% Caucasian, 6.8% Black, 4.7% Hispanic/Latino, 8.4% other) and had a mean age of 19.7 (SD = 3.7). Research participation was

incentivized by offering 3 research credits for participation, which could be exchanged for extra credit in a psychology course.

Table 2.

Summary of Effect Sizes Used to Estimate Effect Size for Power Analysis

	Self-report	**Impulse inhibition/ Executive function**	**Delay of gratification/ Discounting**
Self-report	*r* = .50 (Duckworth & Kern, 2011) *r* = -.03 - .81 (Lane et al., 2003)		
Impulse inhibition/ Executive function	*r* = .10 (Duckworth & Kern, 2011) *r* = -2.0 - .254 (Reynolds, Ortengren, Richards, & de Wit, 2006) *r* = -.012 - .013 (Saunders et al., 2017)	*r* = .15 (Duckworth & Kern, 2011) *r* = .14 (Lane et al., 2003)	
Delay of gratification/ Discounting	*r* = .15 (Duckworth & Kern, 2011) *r* = -.230 - .170 (Reynolds, Ortengren, Richards, & de Wit, 2006) *r* = .08 (White et al., 1994)	*r* = .11 (Duckworth & Kern, 2011) *r* = -.18 - .23 (Lane et al., 2003) *r* = -.03 - .22 (White et al., 1994)	*r* = .21 (Duckworth & Kern, 2011) *r* = .25 (Lane et al., 2003)
Persistence/ Willpower	*r* = .28 (Schmeichel, & Zell, 2007) *r* = -.34 (Schmeichel, & Zell, 2007)	*r* = .09 (White et al., 1994)	*r* = .17 (White et al., 1994)

Procedure

Participants first completed the behavioral measures of self-control in a randomized order. These included the Stroop task, the go/no go task, the flanker task, the cueing task, the

handgrip task, the cold pressor, the taste test, the Iowa Gambling Task, and impossible puzzle tasks. Next, participants completed the survey portion of the study, which included the Self-Control Scale (α = .800), the Barratt Impulsiveness Scale (α = .841), the Achievement motives Scale (α = .835 for the motive to achieve subscale; α = .807 for the motive to avoid failure subscale), the AUDIT (α = .787), the Financial Stress Questionnaire (α = .921), the Weight Concerns Scale (α = .768), the Monetary Choice Questionnaire (α = .929), an impulse spending questionnaire (α = .887), and self-report measures of other outcomes such as GPA, health behaviors, and drug use. Finally, delay of gratification was assessed at the end of the study. This had to be assessed last due to the nature of the focal reward being used, as explained below. See Table 3 for descriptive statistics and reliability coefficients for all self-report variables.

Academic and Health Behaviors

First, academic performance was assessed via GPA. To ensure accuracy, participants were asked to log into their university account and view their GPA digitally, rather than estimate it. Next, participants completed surveys related to food, exercise, spending, and substance use behaviors. To assess food-related health behaviors, items were drawn from the Health-Related Behavior Questionnaire developed by the Schools Health Education Unit (Balding, 2008). This scale assesses the frequency with which certain healthy and unhealthy foods are consumed. Substance use was assessed using items from the Alcohol, Smoking, and Substance Involvement Screening Test (Ali et al., 2002). To assess exercise related behaviors, the very short Leisure Time Exercise Questionnaire (Godin & Shephard, 1985) was used. To assess impulse spending behavior, items were drawn from a survey developed by Güre (2012). For the list of health- and drug-related behavior questions, see Appendix E. Due to the fact that alcohol is more accessible than other drugs and people are expected to encounter alcohol-related temptations more

frequently than other drugs, I also chose to include the Alcohol Use Disorders Identification Test (AUDIT). This is a 10-item questionnaire that assesses the frequency of alcohol use as well as some of the thoughts and emotions surrounding alcohol use (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993). For the AUDIT, see Appendix H.

Table 3.

Descriptive Statistics for Self-Report Variables

Measure	*Mean*	*SD*	*Min*	*Max*	*Reliability*
Self-Control Scale	3.26	.60	1.69	4.77	α = .800
Barratt Impulsiveness	2.91	.37	1.53	3.87	α = .841
Weight Concerns	2.64	1.06	1.00	5.20	α = .768
Impulse Shopping	1.30	.38	1.00	5.00	α = .887
Financial Stress	3.80	.83	1.67	5.00	α = .921
Motive to Achieve	4.21	.59	2.40	5.00	α = .835
Motive to Avoid Fail	3.67	.85	1.00	5.00	α = .807
AUDIT	1.30	.83	1.00	2.70	α = .787
Exercise	2.49	1.04	1.00	5.00	α = .599
Healthy Food	4.69	1.04	1.60	7.00	α = .750
Unhealthy Food	3.98	1.04	1.50	6.50	α = .621

Self-Reported Self-Control

Two different self-report measures of self-control were administered. These included the Brief Self-Control Scale (Tangney et al., 2004) and the Barratt Impulsiveness Scale (Patton et al.,

1995). To view each of these scales, see Appendices B and C. The Brief Self-Control Scale is a 13-item inventory that assesses dimensions of self-control such as restraint and impulsivity. For a factor analysis of this scale, see Maloney et al. (2012). Items on this scale are ranked on a five-point Likert scale ranging from *not at all like me* to *very much like me*. Participants reported on how often they display behaviors, such as saying inappropriate things or indulging in temptations, and indicated whether they have trouble breaking bad habits or working toward long-term goals (e.g., "Pleasure and fun sometimes keep me from getting work done"). The Self-Control Scale has demonstrated acceptable reliability in past research (α = .83 - .85; see Tangney et al., 2004), as well as in the present research (α = .800). This scale has also demonstrated evidence of predictive validity in previous investigations by being able to predict higher grade point average, fewer problems regulating eating, and lower rates of anxiety and depression (Lindner, Nagy, & Retelsdorf, 2015; Tangney et al., 2004).

The Barratt Impulsiveness Scale (Patton et al., 1995) is a 30-item scale that assesses different dimensions of impulsiveness. First-order factors assessed include attention, cognitive instability, motor impulsiveness, perseverance, self-control, and cognitive complexity. Of particular interest in the current study is the self-control subscale, which is composed of 6 items. Responses are scored on a 4-point Likert scale ranging from *rarely/never* to *almost always*. Examples items include "I plan tasks carefully," "I buy things on impulse," and "I am restless at the theater or lectures." Since this scale measures impulsiveness, higher scores indicate lower self-control. Examination of this scale has shown acceptable internal consistency in multiple samples (α = .79 - .83; see Patton et al., 1995), as well as in the present sample (α = .841 for the full scale; α = .758 for the self-control subscale). Subsequent analyses were performed on both the full scale and the self-control subscale on its own.

Attentional Control

Four cognitive inhibition tasks were used in the present study. All of these tasks were programmed and administered using PsyToolkit (Stoet, 2017), a digital platform for developing psychological tests. The four inhibition tasks were the Stroop task (Stroop, 1935), the go/no-go task (Nosek & Banaji, 2001), and the flanker task (Eriksen & Eriksen, 1974) and a spatial cueing task (Posner, 1980). In principle, these tasks require impulse inhibition in order to override an incorrect automatic response and provide a correct response. For example, the Stroop task elicits competition between visual and language systems by presenting people with the words of colors that are printed in different colors and asking them to name the color of the font instead of what the word says. This requires individuals to override their initial impulse to read the word and instead correctly name the color of the font. Similarly, on a certain number of trials on the go/no go task (e.g., the no go trials), individuals must inhibit the impulse to respond when the no go stimulus is present.

The Stroop Task was administered through a web browser using code developed by PsyToolkit. After having the instructions explained to them, participants completed 150 computerized trials of the Stroop task, which took approximately 5 minutes. On this task, approximately 50% of the trials were congruent trials and 50% were incongruent trials. Responses were made using a keyboard with participants pressing “r” for red, “g” for green, “b” for blue, and “y” for yellow. At the end of 150 trials, the researcher recorded reaction time (ms) on congruent trials, reaction time on incongruent trials, and the Stroop interference score. The interference score, or the time it takes a participant to override the impulse to read the word, was calculated by subtracting average reaction time on congruent trials from average reaction time on

incongruent trials. This was calculated for only the trials on which participants responded correctly. In the current sample, the average Stroop effect was 113.73ms (SD = 64.81ms).

Participants also completed 100 trials of a modified go/no go task. This task was split into two portions. The first portion was training, where participants responded to a stimulus using the right and left keys. For this portion, an arrow appeared in the middle of the screen either pointing left or pointing right. If it was pointing left the participant was to hit the left key within 500 milliseconds and if the arrow was pointing right the participant was to press the right key within 500 milliseconds. There were not any no go trials in the training. To move on from the training, participants had to either get 20 correct in a row or complete 50 total. For the second part, participants were given additional instructions. On a certain portion of the trials (30%), the arrow had a red ring around it (e.g., the no go stimulus) which indicated that they should not respond. This meant that when the red circle was present, the participants had to inhibit their dominant response and not press one of the arrow keys. There were 100 trials (30 no go/ 70 go). The proportion of no go trials on which the participant did respond was recorded as a measure of impulsivity. This is preferred to just using the overall percent correct because trials on which the participants were supposed to respond but did not respond (or responded incorrectly) do not constitute as impulsivity. Only trials on which the participant was not supposed to respond but did respond (e.g., errors of commission) constitute impulsive behavior. This decision was made based on previous research showing that errors of omission relate to inattention, whereas errors of commission relate significantly to symptom counts of impulsivity (Bezdjian, Baker, Lozano, & Raine, 2009). Since this is a study about impulsivity, errors of commission were the primary outcome of interest. In the current study, the average number of errors of commission was 3.71 (SD = 2.91; min = 0, max = 15).

The flanker task was originally developed to study the effect of irrelevant or distracting stimuli on responses to a target stimulus (Eriksen & Eriksen, 1974). In this task, participants were tasked with discriminating between two sets of stimuli. In the present study, participants saw one of four letters flash in the middle of the screen. If the letter was an X or a C, they were to press the number 1. If the letter was a V or a B, they were to press the number 0. On these trials, irrelevant distractor letters were presented alongside the target letter. The participant therefore had to ignore the distractors in order to correctly respond to the target letter. On some of the trials, the distractors corresponded to the target stimuli (e.g., X flanked by X or C). On other trials, the distractors corresponded to a response contrary to the target letter (e.g., X flanked by V or B). Much like the Stroop task, participants reliably respond faster and more accurately when the target letter is flanked by a letter that corresponds to the same response, rather than letters that corresponds to the opposite response. The difference in reaction time between the congruent and incongruent trials is the flanker effect and was used as a measure of impulse inhibition in the present study. This is a measure of impulse inhibition because it represents the amount of time it took the participants to override interference from the incorrect distractors and correctly respond. Participants completed 150 flanker trials, which took approximately 5 minutes. The average flanker effect was 25.97ms (SD = 39.37ms)

The Posner cueing task (Posner, 1980) was developed to test the effect of attentional cueing on reaction time to a target stimulus. In this task, participants were shown two boxes. In one of the boxes, a “GO” signal appeared. If the signal appeared on the left, the participants was to press the left key. If it appeared on the right, the participant was to press the right key. On the majority of trials, a small “X” flashed on the side of the screen that the “GO” signal was about to appear. This conditioned the participants to cue their attention to the side of the screen on which

the "X" had just flashed. However, on a minority of trials (invalid cues; roughly 30%), the cue appeared on the opposite side of the screen than where the "GO" signal would appear. To correctly respond on the invalid cue trials, the participants had to override the cued information and alter their response. As a result, participants took reliably longer to respond on invalid cue trials than on the valid cue trials. The difference between the average response time on valid and invalid trials represents the amount of time it takes the participant to override the incorrect cue and may thus be considered a measure of impulse inhibition. Versions of this task have been used to assess self-control in the past (Bartholdy, Cheng, Schmidt, Campbell, & O'Daly, 2016; Steimke et al., 2016). In the present study, the average cueing effect was 71.24ms (SD = 55.13). In other words, participants took on average 71ms longer to respond when their attention had been cued to the invalid position.

Attention to Distal Rewards

To assess individual's willingness to delay immediate gratification in favor of larger rewards the Monetary Choice Questionnaire was used (Kirby et al., 1999). This is a 27-item questionnaire wherein participants must choose between hypothetical rewards delayed over varying amounts of time. An example item is "would you prefer to have $55 today or $75 in 61 days?." Delay discounting was examined by calculating *k*. *k* ranges from .00016 to .25, and lower values represent lower rates of delay discounting, or higher self-control. This was calculated based on the procedure outlined by of Kirby et al. (1999). *k* can be estimated based on a participant's pattern of responses to the questions on the Monetary Choice Questionnaire. For example, question 23 asks "would you prefer $41 today, or $75 in 20 days?." A participant with a discount rate (*k*) of .041 would be indifferent between these two rewards, so if a participant chooses the delayed reward one could infer that their *k* is less than .041. Question 21 asks

"would you prefer $34 today, or $50 in 30 days?," and a participant who is indifferent between these two rewards would have a *k* of .016. If they selected the immediate reward, it can be inferred that they have *k* greater than .016. This would then narrow the estimate of *k* to between .016 and .041. Next, Kirby et al. (1999) suggest taking the geometric mean of these numbers to avoid underweighting the smaller of the two discounting rates. In this example, the estimated *k* would be equal to .0256. Although these can be calculated by hand, the popularity of this scale has led to the creation of programs and spreadsheets that will estimate *k* values automatically. Gray et al. (2016) provide R and SPSS syntax for estimating *k*, which was utilized in the current study. In the present sample, the average *k* was .02319 (SD = .03605). This indicates that, on average, participants in this sample had moderate rates of delay discounting.

In addition to the delay discounting task, participants also completed a delay of gratification task. This task was designed to resemble the experiments carried out by Mischel and colleagues (1958; 1961; 1970; 1972; 1988; 1989) wherein participants were presented with an option between a less desirable but immediate reward or a more desirable but delayed reward. However, because the sample in the current study was comprised of college students, the focal reward used was not food but instead research credits. The decision to not use food was made because adults vary widely in the intensity of their food and weight related goals, meaning food would not elicit a goal conflict of equal strength among participants. By contrast, it is a comparatively safer assumption that participants enrolled in our study are motivated to obtain research credits. This is because at the university in which this research is being conducted, students are able to exchange research credits for extra credit in courses and this constitutes the primary way that research participation is incentivized. Since participants were not offered any incentive besides research credits for engaging in this study, it is a safe assumption that all

participants enrolled in the current study were, by virtue of signing up in the first place, motivated to obtain research credits. Since it is also a safe assumption that college students value their time, presenting them with a choice between leaving immediately (but earning less credits) or staying longer (and earing more credits) represents a direct conflict between an immediately gratifying option with a smaller reward or a delayed option with a larger reward.

Therefore, for the delay task, participants were presented with three options at the end of the study. The first option (immediate gratification option) was to be awarded the three research credits they had earned and be allowed to leave immediately. The second option (intermediate gratification) was to stay in the study for an additional half-hour completing extra surveys but receive an additional half research credit (3.5 total) in return. The third option (delayed gratification) was to stay in the study for an additional hour of surveys but receive an additional full research credit (4 total) in return. This directly assessed participants' willingness to sacrifice an immediate desire (leaving early) in favor of a longer-term goal (extra credit in a course). To ensure that participants' decisions were not affected by scheduling conflicts, all participants were told to allow at least three hours for the study. In the current sample, 54.5% chose to leave immediately, 11% chose to stay for an extra half credit, and 34.5% chose to stay for the additional full credit.

Participants also competed the Iowa Gambling Task. This task was first developed to assess impulsivity in people with damage to the prefrontal cortex (Bechara, Damasio, Damasio, & Anderson, 1994), but has subsequently been used to assess self-control (Sharma, Markon, & Clark, 2014). In this task, participants begin with a virtual loan of $2000. They are then able to draw cards from one of four decks. Two of the decks are high-risk high-reward in that they have a potential to pay out $100 but may also have a penalty fee of $250. Another two decks are low-

risk low-reward and will only pay $50 but has a possible penalty fee of only $50. The task is calibrated such that, in the long run, it is far more advantageous to draw from the low-risk deck. Drawing from exclusively the high-risk deck will result in a net loss and drawing from the low-risk deck will result in a net gain. This task assesses impulsive behavior through people decisions to draw from the high-risk deck or the low risk deck. Participants completed 100 trials and the number of times they picked from the high-risk deck was counted. Higher scores indicate higher impulsiveness. In the present study, the average number of times participants picked from the high-risk deck was 31.35 (SD = 17.5, min = 0, max = 87).

Inhibition of Behavior

Three willpower tasks that have been used in past research were used in this study. These were the taste test, the cold pressor, and the handgrip. The taste test was a measure of caloric intake and has been used as a dependent measure of self-control in many studies of ego-depletion (Hagger et al., 2010). This was performed identically to how it was described in Hagger et al. (2013), since this is representative of the way this has been carried out in most ego-depletion paradigms. For this task, participants were presented with a cover story about market researchers being interested in college students' perceptions of various foods. Participants were presented with two different types of candies and asked to rank them on a variety of dimensions such as taste, texture, and appearance. For the stimuli participants received with the taste test task, see Appendix D. Participants were told that they may eat as much as they needed to in order to accurately judge the quality of the candy. The two candies were Skittles™ and M&Ms™. 50 grams of each were weighed out and presented to the participant. When the participant signaled that they were done evaluating the foods, the experimenter weighed the remaining candies and

recorded the amount consumed in grams. In this sample, the average amount of candies consumed was 11.04 grams (SD = 10.3, min = 0, max = 50).

The second persistence task was the cold pressor task. For this task, participants were told that we would be testing their pain tolerance and were instructed to hold their hand in a bowl of ice water for as long as they could. The water was maintained between 35 and 40 degrees Fahrenheit. The cold pressor is frequently used in psychological studies to simulate pain (Peckerman et al., 1998) and therefore measures a participant's willingness to tolerate an unpleasant stimulus. In theory, this task requires impulse inhibition because individuals must inhibit the impulse to remove their hand from the ice water. To ensure participant safety, anyone who passed the 3-minute mark was instructed to remove their hand from the water. The time (in seconds) was recorded by the experimenter, with longer times indicating higher self-control. The average amount of time that participants persisted on the cold pressor was 97.94 seconds (SD = 66.1, min = 3, max = 180).

This is identical to the protocol used by Schmeichel and Zell (2007), who did find that persistence on the cold pressor was significantly related to self-reported self-control. The only difference between the protocol used by Schmeichel and Zell (2007) and this protocol is that they maintained the water at 34 degrees Fahrenheit, whereas I maintained the water between 35 degrees and 40 degrees Fahrenheit as advised by Mitchell et al. (2004). Of note, this decision may have resulted in longer persistence times, since the average time reported by Schmeichel and Zell (2004) was only 44 seconds.

The final task in this category was the handgrip task. This task has also been used in ego-depletion paradigms (Hagger et al., 2010; Muraven, Tice, & Baumeister, 1998). On this task, participants were given a handgrip and told squeeze it for as long as possible. To account for

individual differences in maximum grip strength, participants first completed a baseline measure of grip strength using a dynamometer. The handgrip was then calibrated to match their maximum grip strength. Thus, this task should have been equally difficult for everyone regardless of their maximum grip strength. Emulating prior research, a coin was placed between the two grips. When the participant loosened their grip enough that the coin fell out, the experimenter stopped the timer. The average amount of time that participants persisted on the handgrip was 17.98 seconds (SD = 15.77, min = 3, max = 87).

Initiation of Behavior

The impossible puzzle task was used to measure persistence in the face of failure. For this task, participants were given an extremely difficult mathematics puzzle and told they had up to 30 minutes to work on it. However, if they found the puzzle too difficult, they could notify the experimenter at any time to move on to the next portion of the study. The dependent variable was the amount of time they persisted on the puzzle. Participants read the following prompt: "The King of a small country invites 1000 senators to his annual party. As a tradition, each senator brings the King a bottle of wine. Soon after, the Queen discovers that one of the senators is trying to assassinate the King by giving him a bottle of poisoned wine. Unfortunately, they do not know which senator, nor which bottle of wine, is poisoned, and the poison is completely indiscernible. However, the King has 10 prisoners he plans to execute. He decides to use them as taste testers to determine which bottle of wine contains the poison. The poison, when taken, has no effect on the prisoner until exactly 24 hours later when the infected prisoner suddenly dies. The King needs to determine which bottle of wine is poisoned by tomorrow so that the festivities can continue as planned. Hence, he only has time for one round of testing. How can the King administer the wine to the prisoners to ensure that 24 hours from now he is guaranteed to have

found the poisoned wine bottle?" As evidence of the difficulty of this puzzle, only one participant was able to solve it and it took them the entire 30 minutes. The average amount of time spent on the puzzle was 12.22 minutes (SD = 6.98, min = 1.3, max = 30).

The second measure of behavioral initiation was the anagram task, which has often been used in ego depletion paradigms. Emulating prior methodology, participants were presented with a list of anagrams. Participants were told that each of the anagrams had at least three solutions (e.g., agntleir can be rearranged as triangle, integral, or relating). However, this was false information, and many of the anagrams only had one solution. Since both solvable anagrams (Baumeister et al., 1998) and unsolvable anagrams (Muraven, Shmueli, & Burkley, 2006) have been used as dependent measures of self-control, I decided to use a combination of both. Persistence was assessed by the amount of time participants persisted at finding all of the "solutions" before giving up. Participants were allowed a maximum of 30 minutes but were told that they could notify the experimenter at any time if they would like to move on. The average amount of time participants spent on the anagrams was 18.72 minutes (SD = 8.4, min = 2, max = 30).

The last task used to assess initiatory self-control was part of the taste test previously mentioned. In addition to testing and rating candies, participants also rated several healthy but commonly disliked foods. These included raw cauliflower and raw broccoli. Whereas inhibition is required to resist eating tasty but unhealthy food, it theoretically requires initiatory self-control to eat healthy but unpleasant foods. Therefore, whereas the sweet foods assess behavioral inhibition, the healthy foods assess behavioral initiation. The experimenter placed 50 grams of broccoli and 50 grams of cauliflower in front of the participants as part of the taste test task. The vegetables were weighed afterwards, and the amount consumed was recorded in grams. The

average amount of vegetables consumed was 9.7 grams (SD = 10.75, min = 0, max = 50). See Table 4 for descriptive statistics for the behavioral measures used.

Table 4.

Descriptive Statistics for Behavioral Measures

Measure	*Mean*	*SD*	*Min*	*Max*
Stroop	*113.73 (ms)*	*64.81 (ms)*	*-107.00 (ms)*	*298.00 (ms)*
Flanker	*25.97 (ms)*	*39.37 (ms)*	*-112.00 (ms)*	*130.00 (ms)*
Cueing	*71.23 (ms)*	*55.13 (ms)*	*-287.00 (ms)*	*291.00 (ms)*
Go/no go	*3.71*	*2.94*	*0*	*15*
Monetary Choice (k)	*.02319*	*.03605*	*.00015*	*.24942*
Handgrip	*17.98 (s)*	*15.77 (s)*	*3 (s)*	*87 (s)*
Cold Pressor	*97.94 (s)*	*66.10 (s)*	*3 (s)*	*180 (s)*
Vegetables Consumed	*9.74 (g)*	*10.75 (g)*	*0 (g)*	*50 (g)*
Candies Consumed	*11.04 (g)*	*10.33 (g)*	*0 (g)*	*50 (g)*
Iowa Gambling Task	*31.35*	*17.50*	*0*	*87*
Anagram Persistence	*18.72 (m)*	*8.40 (m)*	*2 (m)*	*30 (m)*
Math Persistence	*12.23 (m)*	*6.98 (m)*	*1.6 (m)*	*30 (m)*

Control Variables

Due to the fact that individuals come from diverse backgrounds, have diverse interests, and are motivated to achieve different goals, it is important to account for this in an analysis of self-control measures. For example, the food taste test task will have a different meaning for

someone high in dietary restraint or with a weight loss goal as compared to someone who is indifferent about how many calories they consume. Therefore, variables that may confound the interpretation of self-control measures were measured and controlled for. The tasks that could have potentially been confounded by participant motives included the delay of gratification task, the food taste test, the impossible puzzle, and the adjusting amount procedure.

The delay of gratification task could be confounded by participant motives for extra credit. Participants who are already doing well in their courses will have less of a need for extra Sona credits and would thus be less motivated to stay. By contrast, students who are in dire need of extra credit may be more likely to stay. To account for this, a single item was added that assessed student's perceptions of how well they are doing in their classes (e.g., how worried are you that you won't make the grade you want in your psychology classes?) This was ranked on a 5-point Likert scale. The average score on this item was 2.21 (SD = 1.2), indicating only moderate worry about grades, on average.

The food taste test could have been confounded by participant's specific motives surrounding caloric intake and body weight. For someone with a goal of losing or maintaining weight, resisting food is an excellent measure of self-control as it activates a conflict between short-term goals (e.g., eat the tempting sweets) and long-term goals (e.g., losing weight). As defined in this paper, self-control is any act that furthers long-term goals at the expense of short-term goals. However, for someone without any calorie or weight related goals, indulging in the candies does not constitute a failure of self-control. Therefore, it is important to be able to differentiate between the participants with weight-related goals and those without. To this end, the Weight Concerns Scale (Killen et al., 1994) was administered. This scale has shown acceptable reliability in previous investigations ($\alpha = .77$; see Dias, da Silva, Maroco, & Campos,

2015) and showed good reliability in the current sample (α = .768). See appendix I for the full Weight Concern Scale.

The impossible puzzle could be confounded by achievement motivation. Students high in achievement motivation may tend to persist longer as compared to students low in achievement motivation. Importantly, the impossible puzzle task may only constitute a measure of self-control for those students high in achievement motivation. Absent achievement motivation, there would be no conflict of short-term and long-term goals, as the motive to finish the puzzle would be non-existent. To measure achievement motivation, the Achievement Motives Scale (Lang & Fries, 2006) was used. This scale contains two subscales which are the motive to achieve (α = .835) and the motive to avoid failure (α = .807). See appendix J for the achievement motives scale.

Financial stress may confound the results of the adjusting amount procedure. Although these are hypothetical choices, I find it likely that individuals experiencing significant financial stress would be more likely to choose the smaller immediate rewards than the larger delayed rewards. In order to account for this, the Financial Stress Questionnaire was used (Maumary-Gremaud, 1997). This is a 9-item scale that assesses how much stress people experience as it relates to having enough money for food, clothes, and leisure. This scale showed good reliability (α = .921). See appendix K for the scale.

CHAPTER 3

ANALYSIS AND RESULTS

The primary goal of the present research was to determine whether self-control is more appropriately conceptualized as a reflective construct or a formative construct. In order to do this, the theoretical and empirical considerations noted by Coltman et al. (2008) were applied to measures of self-control. These considerations included an examination of indicator intercorrelations, examining indicator relationships with the theoretical antecedents and consequences of the construct, and the vanishing tetrad test.

Data Cleaning and Normalization

Of note, there were several extreme scores present in the data. For example, although 90% of participants scored between 4.6 and 52.8 on the handgrip, a small number of participants exceeded 80 seconds. Therefore, in order to normalize the data, all scores more than 3 interquartile ranges from the median were capped at 3 IQR above the median. Finally, to ensure that all variables were scaled similarly and to avoid large discrepancies in covariances, all variables were converted to Z-scores. The distributions for each variable, after the normalization of outliers and conversion to Z-scores, are presented in Figure 3.

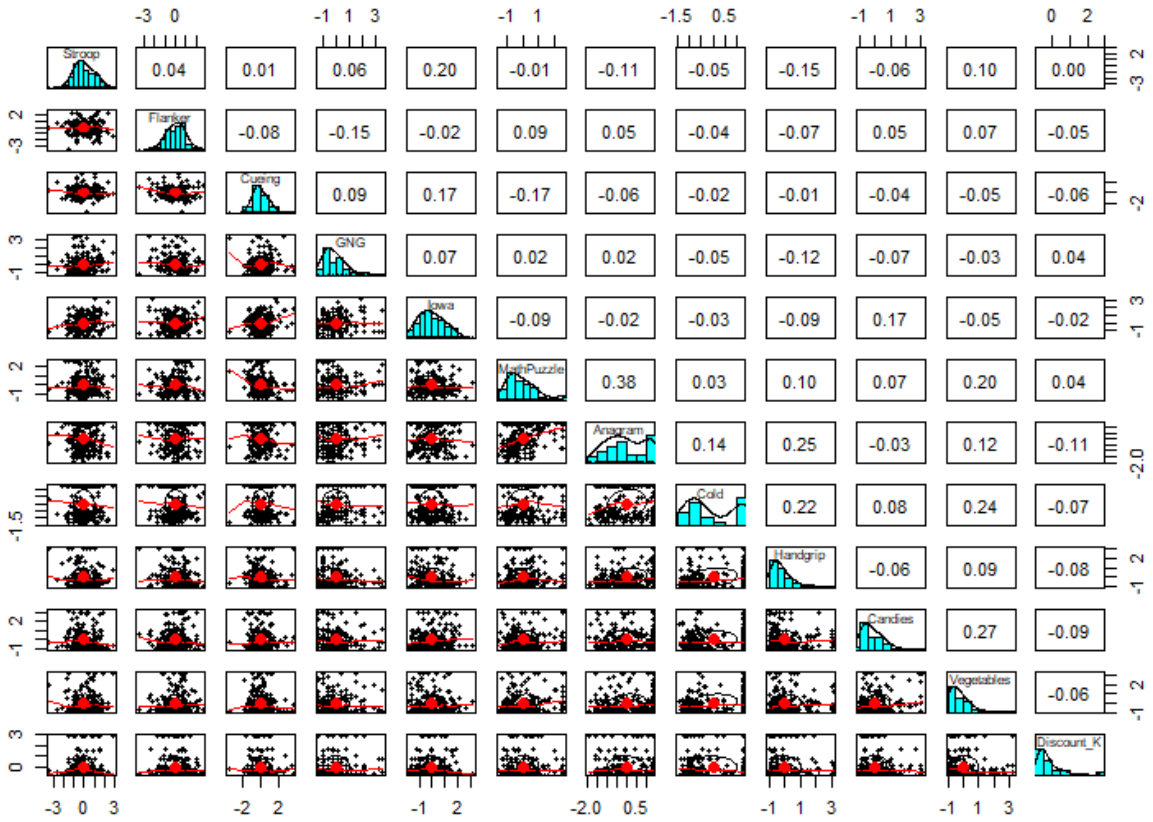

Figure 3. Pair Plots After Normalization

Indicator Intercorrelations

The first consideration outlined by Coltman et al. (2008) concerns indicator intercorrelations. This is essentially a question of convergent validity. If all indicators were influenced by a shared common cause, relatively high correlations would be expected between indicators. Of course, due to measurement error and other unique sources of variance, it is unlikely that any two indicators will correlate perfectly. On the other hand, due to the crud factor (Meehl, 1990), it is also unlikely that any indicators will have no correlation with one another. Unfortunately, there are no hard and fast rules on what ranges of correlations would be expected between indicators under a reflective model versus a formative model. To determine a reasonable

range of values, I turned to the literature on intelligence -- a construct that is often considered to be reflective. A review of the literature on *g* shows that correlations between different indicators of intelligence range roughly from $r = .50$ to $r = .80$ (Canivez, Konold, Collins, & Wilson, 2009; Borghese & Gronau, 2005; Erford & Pauletta, 2005; Spearman, 1904). Based on these considerations, correlations between indicators of $r < .40$ were identified a priori to be considered evidence against a reflective structure.

Table 5 shows intercorrelations between behavioral indicators of self-control. As can be seen in the table, the correlations between the various measures are generally low or non-significant. However, there are a few significant correlations between indicators. The Stroop task correlated significantly with the Iowa gambling task ($r = .196$, $p = .007$) such that higher interference scores related to more risky draws on the Iowa gambling task. The flanker correlated negatively with the go/no go ($r = -.148$, $p = .041$) such that a larger flanker effect related to fewer errors of commission of the go/no go – a finding that is somewhat unexpected. The cueing task correlated positively with risky draws on the Iowa gambling task ($r = .173$, $p = .016$) such that greater interference on the cueing task predicted more risky draws on the Iowa gambling task. Additionally, the cueing task correlated negatively with persistence on the math puzzle ($r = -.169$, $p = .020$) such that greater interference on the cueing task predicted lower persistence on the math puzzle. There was also a positive correlation between the Iowa gambling task and the number of candies consumed during the taste test ($r = .170$, $p = .019$) such that a higher number of risky draws on the Iowa gambling task predicted a greater consumption of candy. Persistence on the math puzzle also correlated positively with the number of vegetables consumed during the taste test ($r = .197$, $p = .006$) such that greater persistence on the math puzzle predicted eating more vegetables. Persistence on the anagrams correlated positively with both persistence on the

handgrip (r = .251, p < .001) and participants choice on the delay of gratification task (r = .188, p = .009), such that participants who persisted longer on the anagrams were more likely to stay for extra credits and hold the handgrip longer. Persistence on the cold pressor correlated positively with both persistence on the handgrip (r = .220, p = .002) as well as the number of vegetables consumed during the taste test (r = .245, p = .001). Finally, persistence on the handgrip predicted participants choice on the delay of gratification task (r = .305, p < .001) such that participants who held the handgrip longer were more likely to stay for additional credits. Controlling for possible confounds (Table 6) had little effect on indicator intercorrelations, except that the correlation between the Stroop and vegetables consumed (r = .145, p = .048) and the correlation between go/no go errors and the delay task (r = -.144, p = .048) became significant.

Although there were some significant correlations between the different measures, none of the observed correlations exceed the aforementioned r > .40 threshold. As such, it seems unlikely that these measures are assessing the same construct. In fact, the magnitude of these correlations hardly exceeds the "crud factor", mentioned earlier (Meehl, 1990). Moreover, many of the significant correlations were between measures with similar methods such as the two taste test outcomes or persistence on the two puzzles. Correlations of this nature are likely not due to a common underlying factor but instead method effects, which are defined as characteristics of a measurement process that produces variance in scores beyond what is attributable to the construct of interest (Maul, 2013). In this case, variance shared between the two taste tests is likely not attributable to self-control, but rather other factors relevant to the consumption of food such as appetite or weight concerns. One suggestion for future research, based on the presence of probable method effects, is to closely examine the multitrait-multimethod matrix (Campbell & Fiske, 1959) for evidence of method effects generating variance among measures of self-control.

Indicator Relationships with Construct Consequences

The second analysis examined indicator relationships with the theoretical consequences of self-control. As discussed earlier, if self-control is hypothesized to lead to a particular set of outcomes, and a particular set of indicators are measuring self-control, then those indicators should also be able to predict the outcomes associated with self-control. In order to do this, correlations between the different measures of self-control and five life outcomes were examined. The five outcomes were GPA, substance use, diet, exercise, and impulse shopping. Since these life outcomes are theoretically more distant from self-control than the indicators themselves and there is a large number of potential intervening variables between self-control and life outcomes, lower correlations are expected than with the indicator intercorrelations. I again turned to the literature on intelligence to determine an appropriate range of values for indicator relationships with the theoretical consequences of a reflective construct. A review of the literature shows that measures of full scale intelligence have correlations with life outcomes such as school or work performance ranging roughly from $r = .25$ to $r = .50$ (Gygi, Hagmann-von Arx, Schweizer, & Grob, 2017; Sternberg, Grigorenko, & Bundy, 2001). However, correlations between individual indicators of intelligence (e.g., measures of object assembly, picture completion, geometric design, block design, arithmetic, and vocabulary) and life outcomes such as academic performance range widely, generally falling between $r = .15$ and $r = .50$ (Kaplan, 1996). Based on these considerations, if self-control is a reflective construct, I expect indicators of self-control to correlate with the theoretically associated outcomes of self-

Table 5.

Correlations Between Measures of Self-Control

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Stroop	--												
2. Flanker	.035	--											
3. Cueing	.011	-.079	--										
4. Go/no go	.062	-.148*	.090	--									
5. Iowa	.196*	-.016	.173*	.073	--								
6. Math Puzzle	-.005	.094	-.169*	.025	-.086	--							
7. Anagram	-.112	.046	-.059	.023	-.021	.380*	--						
8. Cold Pressor	-.047	-.043	-.016	-.049	-.025	.027	.142	--					
9. Handgrip	-.150*	-.068	-.008	-.123	-.087	.099	.251*	.220*	--				
10. Candies	-.056	.049	-.038	-.067	.170*	.066	-.034	.078	-.055	--			
11. Vegetables	.096	.071	-.050	-.028	-.053	.197*	.121	.245*	.088	.268*	--		
12. Discount *k*	.002	-.045	-.058	.040	-.015	.039	-.113	-.070	-.077	-.087	-.056	--	
13. Delay of Grat.	-.090	.016	-.010	-.106	.013	-.029	.188*	-.017	.305*	-.065	.008	-.068	--

*$p < .05$

Note: Correlations are bivariate correlations not controlling for additional variables

Note: After applying Bonferroni correction for multiple comparisons, only 5 correlations remained significant (anagram and math puzzle, p = .00000006; anagram and handgrip, p = .0003; vegetables consumed and cold pressor, p = .00063; delay task and handgrip, p = .00003; vegetables consumed and candies consumed, p = .00013)

Table 6.

Correlations Between Measures of Self-Control – Controlling for Confounds

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Stroop	--												
2. Flanker	.033	--											
3. Cueing	-.006	-.073	--										
4. Go/no go	.045	-.156*	.004	--									
5. Iowa	.176*	-.010	.162*	.047	--								
6. Math Puzzle	-.015	.096	-.173*	.048	-.088	--							
7. Anagram	-.077	.051	-.027	.074	.014	.386*	--						
8. Cold Pressor	.007	-.005	-.006	-.058	-.012	.030	.108	--					
9. Handgrip	-.160*	-.078	-.019	-.099	-.049	.100	.247*	.232*	--				
10. Candies	-.034	.055	-.012	-.057	.201*	.069	-.069	.067	-.048	--			
11. Vegetables	.145*	.077	-.021	-.006	-.038	.200*	.094	.224*	.073	.268*	--		
12. Discount *k*	.095	-.092	-.029	.076	-.018	.066	-.061	-.080	-.048	-.104	.022	--	
13. Delay of Grat.	-.108	.007	-.008	-.144*	.007	-.016	.221*	-.008	.300*	-.069	.022	-.068	--

*$p < .05$

Note: Correlations are partial correlations controlling for grade concerns, financial stress, weight concerns, and achievement motives

Note: After applying Bonferroni correction for multiple comparisons, only 3 correlations remained significant (anagram and math puzzle, p = .00000005; delay task and handgrip, p = .00003; vegetables consumed and candies consumed, p = .0002)

control with at least $r = .20$. I consider this a conservative estimate, because $r = .20$ would mean that a particular indicator is only explaining 4% of the variance in a given outcome. Based on these considerations, I set an a priori limit of $r < .20$ as evidence against a reflective model with regard to indicator relationships with construct consequences.

Table 7 displays correlations between the various indicators and the outcomes theoretically associated with self-control. Importantly, all correlations are presented after controlling for the possible confounds noted above including weight concerns, need for achievement, financial stress, and worry about grades. Overall, there was a severe lack of significant correlations between the indicators of self-control and the hypothesized life outcomes associated with self-control, after controlling for potential confounds. The cold pressor correlated significantly with exercise ($r = .260$, $p < .001$). Somewhat unexpectedly, GPA correlated positively with k ($r = .209$, $p = .004$), suggesting that individuals who are higher in delay discounting have higher GPAs. Additionally, the consumption of healthy foods also correlated negatively with the amount of candy consumed on the taste test ($r = -.176$, $p = .017$). Likewise, self-reported consumption of unhealthy foods correlated negatively with the amount of vegetables consumed during the taste test ($r = -.147$, $p = .046$).

Although there are a few significant correlations between these measures of self-control and life outcomes, the inconsistency of the relationships, as well as the small magnitude of the relationships, casts doubt on whether or not these are indeed assessing self-control. There are a few possible conclusions based on this data. First, it is possible that self-control is still a reflective construct, but that these are poor indicators of self-control and thus are unable to provide useful information regarding outcomes associated with self-control. Second, it is possible the self-control is better conceptualized as a formative construct and that certain

indicators will only have predictive validity within certain domains. For example, perhaps the cold pressor relates to a subdomain of self-control that is important for regulating exercise behaviors, but not other types of behaviors. Both interpretations are explored further in the discussion.

Table 7.

Correlations Between Indicators and Life Outcomes

Measure/Life Outcome	*GPA*	*AUDIT*	*Impulse shopping*	*Healthy Foods*	*Unhealthy Foods*	*Exercise*
Stroop	.128	-.078	-.122	.079	-.019	-.049
Flanker	-.021	-.120	-.015	.027	.038	-.075
Cueing	-.002	-.099	-.049	.008	.060	.062
Go/no go	.037	.079	.067	-.034	-.076	.002
Iowa	.074	.021	.120	-.029	.039	.014
Math Puzzle	.024	.060	.110	-.003	-.032	-.045
Anagram	-.120	-.010	-.042	.095	-.106	-.015
Cold Pressor	-.062	-.064	-.056	.074	-.091	.260*
Handgrip	-.026	.076	-.114	.130	-.046	.076
Candies	.006	.076	.080	-.176*	.010	.117
Vegetables	-.063	-.027	-.008	.078	-.147*	.111
Discount k	.209*	-.005	.087	.019	-.054	.104
Delay of Grat.	-.062	-.067	-.121	-.005	-.008	-.166

*p < .05

Note: Correlations presented are controlling for weight concerns, financial stress, achievement motives, and worry about grades.

Indicator Relationships with Self-Report Measures

An additional test was to examine correlations between the various behavioral indicators and self-report measures of self-control. In general, self-report measures of self-control have been shown to be reliable as well as have good predictive validity (Tangney et al., 2004). Therefore, if any of these behavioral tasks are good measures of self-control, they should be related to self-reported self-control. Correlations between the various behavioral measures and two commonly used self-report measures of self-control – the Self-Control Scale and the Barratt Impulsiveness Scale – are presented in Table 8. By and large, there were very few relationships between the behavioral measures and self-reported self-control. The Self-Control scale correlated weakly, but significantly, with the flanker task (r = .195, p = .007) as well as the amount of candies consumed during the taste test (r = -.173, p = .017) and the amount of vegetables consumed during the taste test (r = -.176, p = .015). In general, individuals scoring higher on the Self-Control Scale consumed less food overall, regardless of type. Additionally, the amount of candies consumed during the taste test was correlated positively with scores on the Barratt Impulsiveness Scale (r = .164, p = .023), such that participants scoring higher on the BIS (more impulsive) consumed more candies. Somewhat unexpectedly, persistence on the math puzzle correlated positively with scores on the BIS (r = .156, p = .031), suggesting that higher impulsiveness assessed by the BIS predicted decreased persistence. Overall, there was a severe lack of correlations between behavioral measures and self-report measures, suggesting major problems with the way that behavioral measures assess self-control.

Table 8.

Correlations Between Behavioral Measures and Self-Report Measures

Measure	*Self-Control Scale*	*Barratt Impulsiveness*
Stroop	.008	-.028
Flanker	.195*	-.132
Cueing	.029	.047
Go/no go	.033	-.027
Iowa	.018	.041
Math Puzzle	-.125	.156*
Anagram	.046	-.087
Cold Pressor	.060	-.142
Handgrip	-.017	-.113
Candies	-.173*	.164*
Vegetables	-.176*	.052
Discount k	-.043	.114
Delay of Grat.	.024	-.096

*p < .05

Note: when using only the self-control subscale of the BIS, candies remained significant (r = .163, p = .024), but math puzzle became nonsignificant (r = .130, p = .073). All other relationships were nonsignificant.

Vanishing Tetrad Test

The final proposed test to probe the structure of self-control was the vanishing tetrad test. As noted previously, the vanishing tetrad test will not be applied to every possible tetrad of indicators. This is because the number of sets of tetrads for models with *n* observed indicators is *n*!/(*n*-4)!4! (Bollen & Ting, 2000). This means that for thirteen tests of self-control, there would be 715 sets of tetrads, or 2145 implied vanishing tetrads. Therefore, I decided to first reduce the number of indicators by obtaining factor scores corresponding to the four self-control sub-domains outlined in the introduction. These sub-domains were attentional control, behavioral inhibition, behavioral initiation, and attention to distal rewards.

In order to examine whether or not the various measures were likely to share variance due to a common factor, I first conducted the Kaiser-Meyer-Olkin (KMO) Test, which is a measure of how suited the data are for factor analysis. A low value on this test (<.60) means that there is a low proportion of shared variance among variables, making the data unsuitable for dimension reduction techniques. By contrast, a high value (closer to 1) means that there is a large proportion of shared variance that may be due to underlying factors (Kaiser, 1974). Not only would a low KMO value suggest that dimension reduction techniques may be unsuitable but would also render the idea of these indicators being part of a reflective construct untenable. Using all thirteen indicators of self-control, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy yielded a value of .528. Kaiser (1974) classified values falling between .50 and .59 as "miserable", suggesting that these data are not suited well for factor analysis. Furthermore, the fact that the KMO value falls in the miserable range is itself further evidence against these indicators measuring a common construct. If these indicators did indeed share a common cause, one would expect high KMO values due to a high amount of shared variance.

As an additional test prior to obtaining factor scores, I conducted Bartlett's test of sphericity. This tests the hypothesis that the correlation matrix is an identity matrix. An identity matrix, or a correlation matrix in which the off-diagonals are exactly zero, would be further evidence that these indicators do not share a common cause. Therefore, a non-significant Bartlett's test of sphericity could be considered further evidence against a reflective model. Despite the relatively poor KMO value, Bartlett's test of sphericity was significant, $\chi^2(78) = 180.13$, $p < .001$. However, further inspection of the correlation matrix shows that the significant Bartlett's test may be due primarily to method effects. As discussed previously, persistence on the anagrams correlated with persistence on the math puzzle, since these are similar types of tasks. Likewise, the amount of vegetables and the amount of candies consumed correlated positively, since these are also similar types of tasks.

Next, I attempted to obtain factor scores for each sub-domain to be used in the vanishing tetrad test. This step was performed using the Lavaan SEM package for R (Rosseel, 2012), which can be used to define latent factors, obtain factor covariance matrices, and examine indicators of model fit. The four hypothesized self-control subdomains were defined according to the theory outlined in the introduction. Specifically, I defined four reflective factors, which included cognitive inhibition (Stroop, flanker, go/no go, cueing), inhibiting behavior (cold pressor, handgrip, candies), initiating behavior (math puzzle persistence, anagram persistence, vegetables), and attention to distal rewards (Monetary Choice Questionnaire, delay of gratification task, Iowa Gambling Task). Unfortunately, this model did not converge. The reason for this seemed to be that the pattern of low and non-existent correlations between the various indicators did not contain adequate information for the SEM package to be able to generate

meaningful factors. In other words, the relationships between the indicators were too small for the SEM package to be able to estimate the hypothesized latent factors.

To rule out possible issues with the data being the cause for the lack of model convergence, the data was further normalized. Two indicators, the cold pressor and the anagram task, showed strong ceiling effects in that on both tasks many participants persisted for the full amount of time. In addition, the taste test indicators, the delay discounting (*k*) indicator, the go/no go, and the handgrip were positively skewed. Therefore, these variables were log transformed to reduce skew and the analysis was rerun. The distribution of the data following the removal of outliers and log transformations can be found in Figure 4. Notably, because log transformations are nonlinear, many of the indicator intercorrelations in the log transformed data are reduced. Ultimately, further normalization of the data still did not allow model convergence and the estimation of the hypothesized latent factors.

Due to these issues with model convergence, I decided to simplify the model and include only the two primary sub-domains of self-control identified by de Ridder et al. (2011). These two primary subdomains were initiatory self-control and inhibitory self-control. Two new models were specified, again using the Lavaan SEM package for R. The first model consisted of only indicators that assess the inhibition of behaviors. This included the Stroop, cueing, go/no go, flanker, candy portion of the taste test, and the cold pressor. As outlined in the introduction, all of these behaviors tasks theoretically require the participants to override a dominant response or ignore task-irrelevant information.

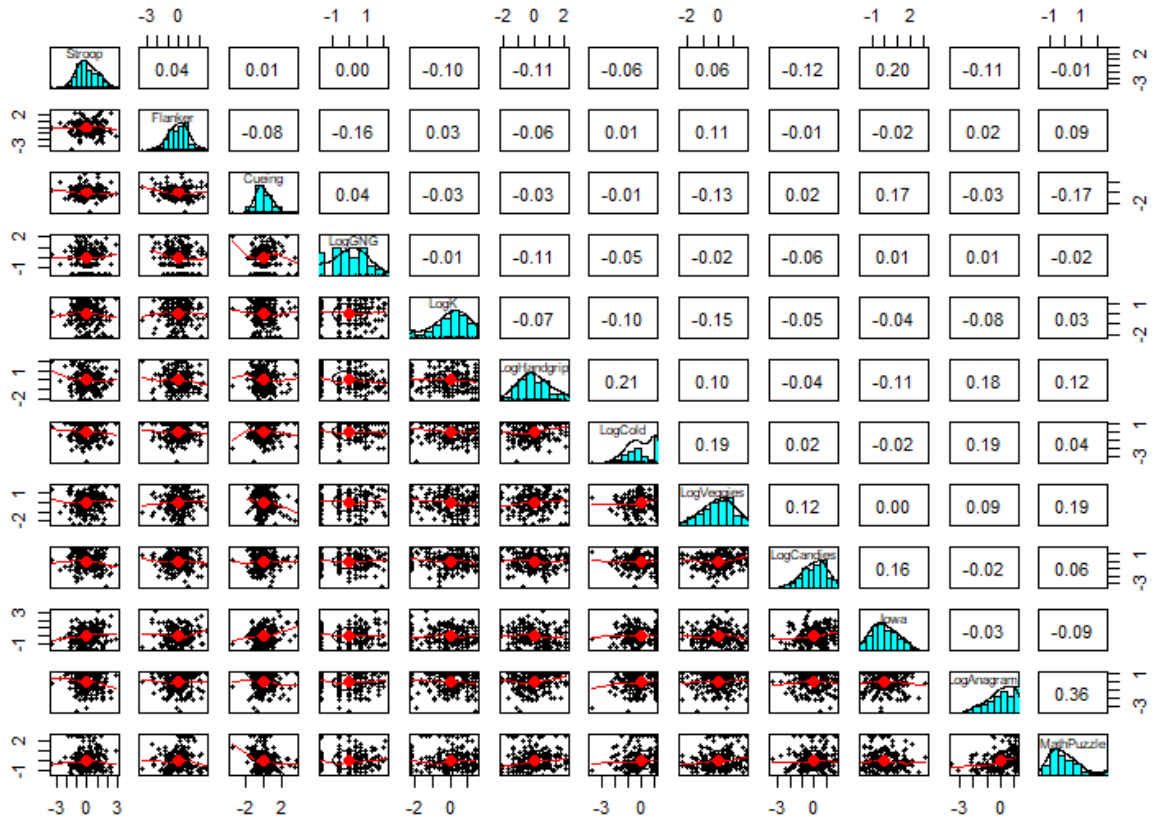

Figure 4. Pair Plots After Log Transformation

This model did converge and showed acceptable fit statistics, $\chi^2(9) = 3.3$, p = .951, CFI = 1.0, RMSEA = 0.000, SRMR = .028. Despite good indices of model fit, however, none of the indicators significantly loaded on the inhibition latent factor. Table 9 shows indicator factor loadings on the inhibitory factor. This further suggests that there is a very small amount of shared variance among these indictors, which would be inconsistent with a reflective model of self-control.

Table 9.

Indicator Loadings on Inhibitory Factor

Indicator	*Loading on Inhibitory Factor*	*Std. Error*
GNG	.489	.246
Stroop	.089	.129
Cueing	-.282	.161
Flanker	.208	.140
Candies Consumed	-.161	.134
Cold Pressor	-.072	.128

Finally, a vanishing tetrad test was performed on the six indicators in the inhibitory model using the R package "ConfirmatoryTetradAnalysis" (Hangcheng, 2019). This yielded a non-significant test statistic, $t = 2.43$, $p = .983$. In other words, the tetrad differences were not significantly different from zero. Although a nonsignificant test statistic would ordinarily be suggestive a reflective model, in this particular case there is another interpretation of this result. Recall that the vanishing tetrad test looks for differences between the products of pairs of covariances, where significant differences indicate data consistent with a formative model. The equations for any given set of four variables are:

$$T_{abcd} = \sigma_{ab}\sigma_{cd} - \sigma_{ac}\sigma_{bd}$$

$$T_{acdb} = \sigma_{ac}\sigma_{db} - \sigma_{ad}\sigma_{cb}$$

$$T_{adbc} = \sigma_{ad}\sigma_{bc} - \sigma_{ab}\sigma_{dc}$$

where σ is equal to the population covariance of the two variables that are indexed below it. Vanishing tetrads, implied by reflective models, mean that $T_{abcd} = 0$. By contrast, tetrads not equal to zero would imply that each set of covariances have unique sources of variance and would thus be most consistent with a formative model.

However, in this case, inspection of the covariance matrix shows several near-zero covariances. This means that for each equation calculating the differences between pairs of covariances, there are values being multiplied by zero. Since anything multiplied by zero is just zero, this will give the appearance that the tetrads vanish, when in fact this is just an illusion created from the exceedingly low covariances between indicators. In other words, the indicators relate to one another so poorly that the vanishing tetrad test is unable to provide reliable results. Table 10 shows the covariance matrix for the five inhibitory indicators submitted to the tetrad test. It can be seen that the covariances are all very nearly zero, thus explaining the non-significant result from the tetrad test.

Table 10.

Covariances for Inhibitory Indicators Submitted to the Tetrad Test

Variable	1	2	3	4	5	6
1. GNG	--					
2. Stroop	.034	--				
3. Flanker	-.143	-.023	--			
4. Cueing	.099	.016	-.067	--		
5. Sweets	.071	-.011	.048	-.033	--	
6. Cold Pressor	-.035	-.006	.020	-.015	.012	--

The second model consisted of only indicators that assess the initiation of behaviors, or tasks that require some active engagement on the part of the participant to complete. This model included the handgrip, persistence on the math puzzle, persistence on the anagrams, their choice of whether or not to complete extra work on the delay of gratification task, and the vegetable portion of the taste test. It should be noted that earlier, the handgrip had been classified as inhibitory self-control. This was a decision that had been based on previous discussion of the handgrip where it was framed as inhibiting an impulse (e.g., "Stamina counts as a measure of self-control because it involves resisting fatigue and overriding the urge to quit;" Baumeister et al., 2007, p. 352). However, conceptual reflection led to the conclusion that squeezing a handgrip is much more an act of initiating a behavior rather than inhibiting a behavior, so it was included in the behavioral initiation model.

This model had relatively poor fit statistics, $\chi^2(5) = 20.65$, p = .001, CFI = .757, RMSEA = 0.128, SRMR = .075. Both persistence on the math puzzle and persistence on the anagrams had loadings above .40 on the initiatory factor (.459 and .775, respectively). However, this is somewhat expected given that both of these are measures of persistence and are correlated with one another. Table 11 shows indicator factor loadings for the behavioral initiation factor.

The five variables in the behavioral initiation factor were then submitted to the vanishing tetrad test. This resulted in a non-significant test statistic, $t = 1.93$, p = .859. Like before, it should be noted that the non-significant vanishing tetrad test would ordinarily indicate support for a reflective model. However, inspection of the covariance matrix again shows many near-zero values, suggesting that the results of the vanishing tetrad test may be misleading. Table 12 shows the covariance matrix for the indicators in the behavioral initiation model.

Table 11.

Indicator Factor Loadings on Behavioral Initiation Factor

Indicator	*Loading on Initiatory Factor*	*Std. Error*
Handgrip	.347	.088
Math Puzzle	.459	.093
Anagram	.775	.122
Vegetables	.193	.088
Delay of Grat.	.250	.088

Table 12.

Covariances for Initiatory Indicators Submitted to the Tetrad Test

Variable	1	2	3	4	5
2. Handgrip	--				
3. Math Puzzle	.160	--			
4. Anagrams	.269	.355	--		
5. Vegetables	.067	.089	.150	--	
6. Delay of Grat.	.087	.115	.194	.049	--

CHAPTER 4

DISCUSSION

General Discussion

The goal of the current research was to apply theoretical and empirical considerations to indicators of self-control in order to determine whether indicators of self-control behave more like indicators of a reflective construct or a formative construct. Theoretical considerations included whether or not the construct is thought to exist independently of the measures used, whether causality is theorized to flow from the construct to the indicators, and whether the construct is manifested by the items. Empirical considerations included examining indicator intercorrelations, examining indicator relationships with theoretical antecedents and consequences of the construct, and the vanishing tetrad test (Coltman et al., 2008; Bollen & Ting, 2000).

With respect to the theoretical considerations, I considered whether self-control is expected to be causally influential on processes such as cognitive inhibition, behavioral inhibition, behavioral initiation, and delay of gratification or whether these tasks together constitute self-control. To be sure, both sides of this debate can be argued. In the introduction, I argued that these subprocesses could operate independently of one another, and that being sufficiently skilled in these domains is what constitutes high self-control. This view would be most consistent with a formative model. On the other hand, a large body of theorizing has conceptualized self-control as a single construct acting causally on these domains (Baumeister et al., 2007; Tangney et al., 2004). That is, self-control has been described as a construct that exists independently of the measures used, explains variance in a wide variety of indicators, and for which the indicators are interchangeable. Arguably, a reflective model is the dominant

conceptualization of self-control as evidenced by numerous research programs that use diverse indicators interchangeably and interpret the results indiscriminately as levels of self-control (Baumeister et al., 1998; Hagger et al., 2010; Vohs et al., 2014).

However, despite some theoretical considerations suggesting that self-control is a reflective construct, in the present study the indicators of self-control failed all of the empirical tests for being indicators of a reflective construct. This largely replicated previous research (Duckworth & Kern, 2010; Lane et al., 2003; Reynolds et al., 2006; Saunders et al., 2017; Schmeichel & Zell, 2007; White et al., 1994) suggesting that indicator intercorrelations are low and that indicators have little-to-no relationship with the theoretical consequences of the construct (Fine et al., 2016; Lane et al., 2003; Sharma et al., 2014; Ward et al., 1993; White et al. 1994). Assuming for the moment that the various cognitive and behavioral tests used in self-control research are indeed measuring important components of self-control (e.g., cognitive inhibition, behavioral inhibition, behavioral initiation), the consistent pattern of low and nonsignificant correlations between them makes the notion of a reflective construct untenable.

With regard to the current research, I observed a consistent pattern of low indicator intercorrelations ranging from r = -.169 up to r = .380. Moreover, two of the largest correlations (r = .268 between the amount vegetables consumed and the amount of candies consumed; r = .380 between persistence on the math puzzle and persistence on the anagrams) are almost certainly due to method effects rather than shared variance due to a common underlying construct. That is, these tasks were so similar in type that any correlation between them is likely due to the similarity of the tasks rather than variance shared due to an underlying latent construct.

Despite the generally low and non-significant correlations among the various measures, it should be noted that there were a few significant correlations that were not likely due to method effects. The two indicators that had the most consistent pattern of relationships with other measures were the number of impulsive draws on the Iowa Gambling Task and the handgrip task, with the Iowa Gambling Task correlating significantly with three other measures and handgrip persistence correlating significantly with four other measures. The number of impulsive draws on the Iowa Gambling Task correlated significantly with Stroop interference (r = .196, p = .007), cueing interference (r = .173, p = .016), and the number of candies consumed during the taste test (r = .170, p = .019). The amount of time that participants were able to persist at holding the handgrip correlated significantly with persistence on the anagrams (r = .251, p < .001), participants choice on the delay of gratification task (r = .305, p < .001), persistence on the cold pressor (r = .220, p = .002), and correlated negatively with Stroop interference scores (r = -.150, p = .038). Other significant correlations of interest include the correlation between persistence on the anagrams and participants choice on the delay of gratification task (r = .188, p = .009), persistence on the cold pressor and the number of vegetables consumed during the taste test (r = .245, p = .001), persistence on the math puzzle and vegetables consumed during the taste test (r = .197, p = .006), persistence on the math puzzle and cueing interference (r = -.169, p = .020), and a negative correlation between flanker scores and errors of commission on the go/no go (r = -.148, p = .041).

One interpretation of this pattern of significant correlations is that these measures share variance due to an underlying construct. Unlike the correlations between the two persistence tasks and the two taste test indicators, the correlations just mentioned are likely not due to method effects since these correlations are between measures of quite different types (e.g.,

heteromethod-monotrait; Campbell &Fiske, 1959). Indeed, the nature of these tasks vary widely from tests of reaction time and cognitive inhibition to persistence on puzzles to food consumption to physical endurance and pain tolerance. Further support for these measures sharing at least some variance due to a common latent factor is that all of the noted correlations are in the direction that would be expected based on theory, with the one exception being the negative correlation between the flanker and errors of commission on the go/no go. One possibility based on these findings is that these measures are indeed capturing variance associated with self-control, but that these are also very noisy measures with large sources of error variance resulting in low correlations between indicators. In fact, it is difficult to come up with an additional plausible explanation, besides a latent factor, that explains the shared variance between tasks as diverse as a gambling task, a taste test, handgrip persistence, a delay of gratification task, and tests of reaction time. Methodologically speaking, these tasks have little to do with one another. However, assuming for the moment that these measures are correlated due to a latent factor (e.g., self-control), the low magnitude of the correlations suggests that, at best, these are noisy measures with significant sources of error variance.

Therefore, based on these results it could be the case that these measures are assessing variance associated with self-control, albeit a very small amount. The majority of variance in individual's scores would be attributable to the unique demands of each task, rather than the latent construct they are intended to measure. For example, consider the correlation between math puzzle persistence and vegetables consumed. The correlation between these two variables was $r = .197$, or an r-squared = .039. In other words, persistence on the math puzzle and the amount of vegetable consumed share about 3.9% of their variance. Since these two tasks share very little in common methodologically, it is not unreasonable to assume that the shared variance

is not due to any method effects but is rather due to a shared latent construct. The rest of the variance that is not shared (e.g., 96.1%) would be variance that is unique to each task. For example, on the math puzzle, factors like need for cognition, the motive to achieve, enjoyment of math puzzles, or expertise in mathematics could be sources of variance that are unique to that task and are unrelated to the amount of vegetables consumed. Likewise, there are unique sources of variance in the vegetable taste test (e.g., appetite, weight concerns, taste preferences) that influence individual's scores and are not present in the math puzzle. In simpler terms, it might be that self-control accounts for a very small amount of the shared variance (about 3.9%) between these tasks, but idiosyncrasies in task demands (e.g., attention, appetite, pain tolerance, need for cognition, etc.) dominate the data and represent large sources of error variance. In this sense, these measures may in some small part assess self-control, but overall are far too noisy to yield meaningful relationships.

Despite the presence of some interesting indicator intercorrelations, the current research showed a pattern of low and nonsignificant correlations between the indicators of self-control and the theoretical consequences of self-control. As mentioned, self-control has been theorized to have a wide-reaching influence on diverse behaviors and outcomes including achievement, interpersonal success, adjustment, dietary restraint, substance use, and spending (Baumeister, 2002; Baumeister et al., 2007; Tangney et al., 2004). Therefore, assuming again for the moment that self-control is a reflective construct and that the various cognitive and behavioral tasks are valid measures of self-control, such measures would be expected to predict these life outcomes. By and large, this is not what was observed. Correlations between the indicators and life outcomes ranged from r = -.176 to r = .260, with only four significant correlations in total. Moreover, like with the indicator intercorrelations, two of these significant correlations (r = -.147

between vegetables consumed on the taste tests and self-reported consumption of healthy foods; r = -.176 between candies consumed on the taste test and self-reported consumption of healthy foods), are likely not due to the effects of a latent construct but due to the simple fact that the behavioral tests are assessing the same behavior about which the self-report measures are asking (e.g., both measure food consumption).

Like the indicator intercorrelations, the lack of relationships could be largely due to the fact that these appear to be noisy measures with large sources of error variance (e.g., variance that cannot be accounted for by their relationship with the latent construct). Even when using good measures of a construct, indicator correlations with the construct consequences would not necessarily be expected to be exceptionally high. This is because there are even more sources of potential variance. First, there is the variance associated with task idiosyncrasies, as discussed previously. Second, there is variance associated with the measurement of the outcome variable (e.g., health behaviors, impulse shopping, etc.). Finally, on the path from self-control to the proposed outcomes, there are additional factors at play that operate independently of self-control but also influence the outcome variable. For example, although self-control is proposed to influence student's GPA, there is a host of other variables, independent of self-control, that determine GPA, such as financial resources, the presence or absence of an illness, peer group, instructor quality, and many more. Put simply, even if you could measure self-control perfectly, this would not allow you to estimate theoretically implied outcomes with perfect accuracy. This fact, paired with the already noisy measurement methods, may be what explains the pattern of extremely low indicator correlations with the construct consequences. Self-control may play an important role in behaviors such as impulse shopping, substance use, and academic behaviors, but if self-control cannot be measured without large error variance and the consequences of the

construct are theoretically distant to begin with (e.g., outcomes that are manifested after the influence of a large number of intervening variables), correlations between the construct and it's theoretically implied consequences cannot possibly be expected to be high.

Next, I had planned to generate factor scores based on the four hypothesized subdomains of self-control (controlling attention, behavioral inhibition, behavioral initiation, attention to distal rewards) and submit these to the vanishing tetrad test as a way of testing the final empirical consideration of whether a construct is reflective or formative. I was, however, unable to do this because the hypothesized four factor model did not converge, suggesting that the four-factor model had extremely poor fit to the data. Examination of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy showed a value of .528, which Kaiser (1974) labeled as "miserable." This means that it is unlikely that there are patterns in the data that would be amenable to factor analysis.

I then examined a two-factor model including factors for inhibitory self-control and initiatory self-control, the two main types of self-control identified by de Ridder et al. (2011). The inhibitory factor was comprised of tasks that involve inhibiting an automatic response and included the Stroop, the Go/no-go, the cueing, the flanker, the candy portion of the taste test, and the cold pressor. Although this model had good fit statistics, none of the indicators loaded significantly on the inhibition factor, suggesting that these tasks have little to do with one another. This, in-and-of-itself, could be considered further evidence against a formative model. Next, I conducted the vanishing tetrad test on these six indicators and found a nonsignificant result. Although ordinarily a nonsignificant result on this test would favor a reflective interpretation, the pattern of near-zero covariances between indicators likely resulted in a false

negative. In other words, the covariance matrix contained so little information so as to render the vanishing tetrad test unreliable.

The initiatory factor was comprised of tasks that involve the initiation of behavior in order to complete a task. These indicators included the amount of vegetables consumed, persistence on the handgrip, persistence on the math puzzle, persistence on the anagrams, and the participant's choice of whether or not to stay in the study longer to earn additional credit. This model had relatively poorer fit indices, but persistence on the math puzzle and persistence on the anagrams did load significantly on the initiatory factor. Finally, I conducted the vanishing tetrad test on these five indicators and found a nonsignificant result. Like the tetrad test on the inhibitory indicators, however, it is likely that the covariance matrix contained so little information that it rendered the vanishing tetrad test unreliable.

Overall, the data do not support a reflective model of self-control. Although theory may suggest that self-control is reflective, the indicators for self-control fail all of the empirical tests that would suggest it is reflective. If self-control were a reflective construct, indicators of self-control would be expected to be interchangeable and highly correlated due to sharing a common cause. The fact that the data show a pattern of largely small and non-significant correlations between indictors suggests that it is highly unlikely that they are all measuring the same underlying construct.

There are three main interpretations of the data at this point. First, self-control could be reconceptualized as a formative construct. This would remove the assumption that indicators should be correlated, which is consistent with the data. It would also remove the assumption that all indicators should relate similarly to the consequences of the construct, since subdomains of self-control could operate independently of one another, with some being relevant for certain

behaviors and not others. For example, perhaps the cold pressor assesses a sub-domain of self-control that is relevant for explaining exercise behavior, but not shopping behavior.

The second interpretation is that self-control is a reflective construct, but that these measures are far too noisy and contain too large sources of additional variance making them unreliable measures of self-control. As discussed earlier, this may be suggested by a pattern of significant (albeit small) correlations between indicators. Moreover, all of these correlations were in the direction expected by theory (with the exception of one) and were between methodologically diverse measures. The fact that many of these correlations were between methodologically dissimilar indicators effectively rules out method effects as a plausible explanation for at least some of these correlations. However, one could counter the interpretation that these correlations are due to shared variance in the self-control construct and propose that they are instead due to shared variance in another construct that influences performance on these tasks, such as general intelligence. Theorists may want to consider this as an alternative possible explanation for the pattern of relationships among these indicators.

The third interpretation is that, regardless of whether self-control is reflective or formative construct, these measures are bad indicators of self-control. Under this interpretation, none of these measures are assessing any construct level variance, and any significant correlations are merely false positives or due to method effects. As such, one would not expect any of these indicators to tell us anything about the structure of the self-control construct. Under this interpretation, asking questions about the structure of self-control using these indicators may be akin to asking questions about the structure of socioeconomic status (SES) by using data on hair color and eye color. The latter is unrelated to the former and cannot answer any questions about the nature of the construct.

Unfortunately, the available data cannot definitively say which interpretation is correct. It can, however, provide some guidance. Let's assume for a moment that the first interpretation is correct and that self-control is best conceptualized as a formative construct. Even though we would not necessarily expect indicator intercorrelations to be high, we would still expect indicator correlations with global measures of the construct, such as self-report measures. Again, taking up the SES analogy from before, although one might not necessarily expect individual indicators to correlate with one another, one would expect indicators to correlate (at least somewhat) with overall measures of SES, since they are by definition part of what comprises this construct. Therefore, if self-control were a formative construct, one would expect individual indicators of self-control to correlate at least weakly with global self-control, since they are part of how the construct is defined. This is not what was observed. Indicator correlations with established self-report measures ranged from r = -.176 (SCS and vegetables consumed) to r = .195 (SCS and flanker score). These correlations were, on the whole, small and nonsignificant.

The fact that the indicators do not relate to global self-control is evidence that, regardless of whether self-control is reflective or formative, these are poor indicators of self-control. Unfortunately, if this is the case, then the available data is not in a position to answer the question of whether or not self-control is reflective or formative. Until behavioral measures are developed that relate significantly to global measures of self-control, this question probably cannot be answered. As Cronbach and Meehl stated in 1955 (p. 287), "if the obtained correlation departs from the expectation, however, there is no way to know whether the fault lies in test A, test B, or the formulation of the construct." Initially, I had hypothesized that the problem was with the formulation of the construct. However, based on an examination of the data and taking into account the startling lack of convergent, predictive, and concurrent validity of these

measures, I now find it likely that the problem lies in the tests. Furthermore, questions about the formulation of the construct will likely be unanswerable until the development of behavioral tests that relate to global self-control in a meaningful way.

Limitations

There are a few potential limitations of the current research. The first limitation concerns ecological validity. It is possible that many of the laboratory test of self-control used in the present research do not resemble real-life acts of self-control and thus do not assess the construct accurately. For example, on the adjusting amount task, hypothetical rewards were used instead of real rewards. This could be relevant because research suggests that participants tend to respond to risk differently on hypothetical tasks as opposed to real tasks, with participants being more risk-sensitive on real tasks (Xu, Fang, & Rao, 2013). Similarly, on tasks like the food taste test, the environment and context of a laboratory does not match that of naturalistic food related temptations, which more often occur in restaurants or grocery stores. For tasks like the cold pressor, participants may not have seen a good reason to be sufficiently motivated to endure the pain of holding their hand in ice water. Usually, when people force themselves to endure pain, it is in the service of a long-term goal, such as becoming stronger or more athletic, which was not the case in the present study. Although the low ecological validity of the laboratory setting may have had an effect on the accuracy of participant's scores, this is not a problem unique to this study. Ideal conditions or not, self-control researchers are going to have to conduct tests in the laboratory at some point, and it is important to assess the accuracy of these measures in this context.

Another concern was that performance may be inconsistent across tests due to ego-depletion effects (Baumeister et al., 1998). The ego-depletion effect refers to a phenomenon

wherein an initial act of self-control can "deplete" self-control resources, leading to poorer performance on subsequent tasks requiring self-control. Although the effect is somewhat controversial (Carter et al., 2015), this was a concern in the present study due to the sheer quantity of self-control measures being administered. If the ego-depletion effect is a real phenomenon, then one would have expected to see substantial depletion effects by the end of this study. Moreover, there is some evidence that ego-depletion effects may become more severe after repeated engagement with depleting tasks (Vohs, Baumeister, & Schmeichel, 2012). To protect against this, the order of the behavioral measures of self-control was randomized. The randomization of the order of the tasks should have prevented ego-depletion effects from corrupting the data. However, due to the nature of the delay of gratification task (offering research credits for extra time in the study), it had to be the last measure administered. Therefore, if depletion effects were present, they may have influenced participants' decisions on this last task.

Another limitation unique to the delay of gratification task could have been that the time of semester that participants completed the study influenced participants' decisions. Early on in the semester, students have not yet earned any Sona credits and my thus be willing to stay and earn additional credits. As the semester progresses, students begin to meet the Sona requirements for their courses and will thus be less likely to desire additional Sona credits. Thus, students participating in the study toward the end of the semester may have been less likely to select the option for extra credits simply because they already had enough credits. On the other hand, it could be argued that students participating toward the end of the semester who had not yet fulfilled Sona requirements would be more desperate for Sona credits and would thus be more likely to stay for the extra credits. In any case, the time of semester is a potential confound with

regard to participants choice to stay for additional credits. In order to account for this, data collection was ceased two weeks prior to the end of the semester, although some of the effects of this may still have been present.

A final limitation of the data is that a few of the tests, in particular the cold pressor task and the anagram task, showed ceiling effects. On the anagram task, 20.9% of participants persisted for the entire thirty minutes. On the cold pressor task, 35.6% of participants persisted for the maximum of 180 seconds. These ceiling effects on these variables severely limits their variability and the inferences that can be drawn from them. Unfortunately, this is not an issue that can be addressed post-hoc, and it should be noted that the ceiling effects on these variables may limit the inferences drawn from them.

Implications for Future Research

The results of this research generate at least one clear direction for future research. First and foremost, this research shows that the research community does not currently have any validated "gold standard" behavioral measures of self-control. Therefore, at a bare minimum, researchers should seek to develop behavioral tests of self-control that display reliable relationships with validated self-report measures of global self-control. As noted earlier, this is a necessary precondition to being able to ask questions about the ontological nature of the construct. Regardless of whether self-control is reflective or formative, indicators should be able to correlate, at least somewhat, with global self-control.

Additionally, when developing future measures, researchers should pay special attention to the multitrait-multimethod matrix described by Campbell and Fiske (1959) to ensure that the construct indicators are behaving in a way consistent with theory. The multitrait-multimethod matrix is a method for establishing construct validity, which examines both convergent and

divergent validity in determining whether a supposed measure of a construct relates the way that it theoretically should to other measures. A proposed measure of a construct should correlate highly with measures of theoretically related constructs and bear less relationship with measures of theoretically distinct constructs. Based on the available data, currently used measures of self-control do not behave the way that measures of self-control would be theoretically expected to behave.

Additionally, examination of the multitrait-multimethod matrix allows for the assessment of method effects, which may be present among measures of self-control. This is because in examining the multitrait-multimethod matrix, it is expected that correlations between measures of the same trait using different methods (e.g., heteromethod-monotrait) should be higher than correlations between measures of different traits using the same method (e.g., monomethod-heterotrait) (Campbell & Fiske, 1959). If the latter is equal to or higher than the former, this suggests that characteristics of the methodology, rather than the construct, are explaining variance. In other words, traits should be empirically distinguishable even when assessed via the same method.

An example of what this might look like is shown in Table 13. To arrange a multitrait-multimethod matrix, two measures of the same trait using different methods are needed, as well as two measures of a different trait using the same two methods. In the example below, two traits (self-control and working memory) are being measured with two different methods (self-report and measures of cognitive performance). Characteristics of this matrix to which researchers should pay attention are convergent validity (e.g., tests of the same construct should correlate highly), discriminant validity (e.g., measures of one construct should not correlate highly with measures of a different construct), and considerations of multitrait-multimethod variance that

allow for the assessment of method specific variance (Campbell & Fiske, 1959). As stated previously, correlations between measures of the same trait using different methods (e.g., heteromethod-monotrait) should be higher than correlations between measures of different traits using the same method (e.g., monomethod-heterotrait). If, in this example, the correlation between the N-back and the Stroop (monomethod-heterotrait) were found to be higher than the correlation between the Stroop and the Self-Control Scale (heteromethod-monotrait), this could be evidence for strong method effects present in the Stroop task.

Table 13.

Example Multitrait-Multimethod Matrix for an Indicator of Self-Control

Test	Self-Control Scale	Behavioral Measure Being Validated (example: Stroop)	Working Memory Questionnaire	N-back Task
Self-Control Scale	--			
Behavioral Measure Being Validated (example: Stroop)	Heteromethod-Monotrait (highest expected)	--		
Working Memory Questionnaire	Monomethod-Heterotrait (low, less than monotrait)	Heteromethod-Heterotrait (lowest of all expected)	--	
N-back Task	Heteromethod-Heterotrait (lowest of all expected)	Monomethod-Heterotrait (low, less than monotrait)	Heteromethod-Monotrait (highest expected)	--

Second, I would strongly suggest that researchers consider taking into account individuals' specific goal conflicts when measuring self-control behaviorally. The temptation to eat a slice of chocolate cake is not equally strong across all individuals, nor is the desire to lose weight equally strong across all individuals. The handgrip task might have a different meaning

for a male who is trying to impress a female experimenter than for a female working with a female experimenter. The impossible puzzle task might have a different meaning for someone who bases their self-concept around their intelligence and ingenuity than someone who bases their self-concept on their appearance or physical ability. The point is that goal conflicts are different for different people, and it is unlikely that any one task will adequately assess self-control in all individuals.

When devising new tests of self-control, it will be of crucial importance to keep in mind what self-control fundamentally is. As defined in this paper and by Fujita et al. (2011), self-control is "the process of advancing distal rather than proximal motivations when the two compete" (p. 352). Therefore, what any good measure of self-control should be assessing is an individual's ability to pursue distal motivations, particularly in situations where more proximal temptations are competing with these. This aspect of motivation and goal-pursuit is something that is sorely lacking from current behavioral measures of self-control. For instance, rewards do not increase the longer an individual persists on the cold pressor or an impossible puzzle. In a best-case scenario, currently available methods could be modified to induce goal conflicts and assess the ways in which individuals resolve such conflicts. For example, imagine offering participants $10 if they were able to solve the impossible puzzle. What one would likely find is that suddenly, a lot more participants would have the "self-control" to persist the entire time.

Taking all of this into consideration, it may be that the best way to assess self-control behaviorally is to track people's behavior on a day-to-day basis using an experience sampling methodology (Bolger & Laurenceau, 2013). By doing this, researchers would be better equipped to understand the goals of participants, the types of goal conflicts they encounter, and the ways in which they resolve such goal conflicts. This would have several advantages over laboratory tests

of self-control, such as increased ecological validity, the ability to account for individual differences in goal conflicts, and would allow for the analysis of variables not present in laboratory settings that may influence behavioral outcomes.

Ultimately, the currently available measures of self-control do not appear to be assessing important aspects of the construct. This is evidenced by the poor convergent, predictive, and concurrent validity found in this study as well as numerous others. In order for self-control research to get back on its feet, researchers will need to develop behavioral measures of self-control that, at a minimum, can relate reliably to measures of global self-control. Until this first goal is accomplished, questions regarding the antecedents and consequences of state self-control and phenomena such as ego-depletion will have to go unanswered.

Conclusion

In sum, the data show a pattern of relatively low and nonsignificant correlations between the indicators of self-control. One possible interpretation of this finding is that self-control is a formative construct comprised of many unrelated subdomains that are all relevant for goal-related behaviors. In such a circumstance, high indicator intercorrelations would not necessarily be expected. However, based on a pattern of (albeit small) significant indicator intercorrelations, I find one of two additional interpretations to be more likely.

First, it could be that self-control is a reflective construct and, simply put, none of these measures are capturing any of the variance associated with the construct. In other words, these are bad measures of self-control and any observed correlations are due merely to method effects (Maul, 2013). In fact, the presence of at least two correlations (correlations between the two taste tests and correlations between the two puzzle persistence measures) are almost certainly due to method effects rather than a shared variance due to a latent construct.

A second and more optimistic interpretation is that self-control is a reflective construct and that some of these measures are in fact assessing variance shared due to this construct, as evidenced by a pattern of small but significant correlations between several indicators. Moreover, these relationships were (with the exception of one) in the direction that would be expected based on theory. That being said, the correlations between these indicators are small and the variance in these measures is most likely dominated by sources of variance that are unique to each task, rather than due to the construct they are intended to measure. This severely limits the inferences about self-control that can be drawn from these measures. At the present moment, it seems that measures of self-control are far too noisy in the sense that responses are dominated by error variance and method effects due to task idiosyncrasies. For example, although a small percentage of the variance in taste test outcomes may be attributable to self-control, such outcomes are ultimately better assessments of factors like taste preference or appetite, and thus such outcomes only provide minimal information on self-control. Thus, moving forward with self-control research, it is suggested that researchers first develop behavioral measures of self-control that are able to minimize the amount of variance due to task idiosyncrasies and method effects, while maximizing the amount of variance captured in the self-control construct.

REFERENCES

Ali, R., Awwad, E., Babor, T., Bradley, F., Butau, T., Farrell, M., & ... Vendetti, J. (2002). The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, reliability and feasibility. *Addiction, 97*(9), 1183-1194. doi:10.1046/j.1360-0443.2002.00185.x

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., & Rodriguez, M. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. *Journal Of Personality And Social Psychology*, *79*(5), 776-792. doi:10.1037/0022-3514.79.5.776

Balding, J. (2008). Collecting good data: The primary health related behaviour questionnaire (Version 11.2 Manual). Exeter: Schools Health Education Unit.

Bartholdy, S., Cheng, J., Schmidt, U., Campbell, I. C., & O'Daly, O. G. (2016). Task-based and questionnaire measures of inhibitory control are differentially affected by acute food restriction and by motivationally salient food stimuli in healthy adults. *Frontiers In Psychology*, *7.* doi: 10.3389/fpsyg.2016.01303

Baumeister, R. F. (2002). Yielding to temptation: Self-control failure, impulsive purchasing, and consumer behavior. *Journal Of Consumer Research, 28*(4), 670-676. doi:10.1086/338209

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource?. *Journal Of Personality And Social Psychology, 74*(5), 1252-1265. doi:10.1037/0022-3514.74.5.1252

Baumeister, R. F., & Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives On Psychological Science, 11*(4), 574-575. doi:10.1177/1745691616652878

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions In Psychological Science, 16*(6), 351-355. doi:10.1111/j.1467-8721.2007.00534.x

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1-3), 7-15. doi:10.1016/0010-0277(94)90018-3

Bezdjian, S., Baker, L. A., Lozano, D. I., & Raine, A. (2009). Assessing inattention and impulsivity in children during the Go/NoGo task. *British Journal Of Developmental Psychology*, *27*(2), 365-383. doi:10.1348/026151008X314919

Blázquez, D., Botella, J., & Suero, M. (2017). The debate on the ego-depletion effect: Evidence from meta-analysis with the p-uniform method. *Frontiers In Psychology, 8.* doi:10.3389/fpsyg.2017.00197

Bolger, N., Laurenceau, J.P. (2013). Intensive longitudinal methods: An introduction to diary and experience sampling research. New York, N.Y.: Guilford Press

Bollen, K. A., & Ting, K. (2000). A tetrad test for causal indicators. *Psychological Methods, 5*(1), 3-22. doi:10.1037/1082-989X.5.1.3

Borghese, P., & Gronau, R. C. (2005). Convergent and Discriminant Validity of the Universal Nonverbal Intelligence Test with Limited English Proficient Mexican-American Elementary Students. *Journal Of Psychoeducational Assessment*, *23*(2), 128-139. doi:10.1177/073428290502300202

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219. doi:10.1037/0033-295X.110.2.203

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal Of Experimental Psychology: General*, *144*(4), 796-815. doi:10.1037/xge0000083

Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler Abbreviated Scale of Intelligence and Wide Range Intelligence Test: Convergent and structural validity. *School Psychology Quarterly*, *24*(4), 252-265. doi:10.1037/a0018030

Champely, S. (2018). Pwr: basic functions for power analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250-1262.

Crean, J. P., de Wit, H., & Richards, J. B. (2000). Reward discounting as a measure of impulsive behavior in a psychiatric outpatient population. *Experimental And Clinical Psychopharmacology*, *8*(2), 155-162. doi:10.1037/1064-1297.8.2.155

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. doi:10.1037/h0040957

de Ridder, D. D., de Boer, B. J., Lugtig, P., Bakker, A. B., & van Hooft, E. J. (2011). Not doing bad things is not equivalent to doing the right thing: Distinguishing between inhibitory

and initiatory self-control. *Personality And Individual Differences, 50*(7), 1006-1011. doi:10.1016/j.paid.2011.01.015

de Ridder, D. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality And Social Psychology Review*, *16*(1), 76-99. doi:10.1177/1088868311418749

Diamond, A. (2013). Executive functions. *Annual Review Of Psychology*, *64,* 135-168. doi:10.1146/annurev-psych-113011-143750

Dias, J. R., da Silva, W. R., Maroco, J., & Campos, J. B. (2015). Validation of the Weight Concerns Scale applied to Brazilian university students. *Body Image*, *14*72-76. doi:10.1016/j.bodyim.2015.03.012

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal Of Personality And Social Psychology*, *58*(1), 95-102. doi:10.1037/0022-3514.58.1.95

Duckworth, A., & Kern, M. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal Of Research In Personality, 45*(3), 259-268. doi:10.1016/j.jrp.2011.02.004

Duckworth, A. L., Tsukayama, E., & Kirby, T. A. (2013). Is it really self-control? Examining the predictive power of the delay of gratification task. *Personality And Social Psychology Bulletin, 39*(7), 843-855. doi:10.1177/0146167213482589

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174. doi:10.1037/1082-989X.5.2.155

Engelhardt, P. E., Nigg, J. T., Carr, L. A., & Ferreira, F. (2008). Cognitive inhibition and working memory in attention-deficit/hyperactivity disorder. *Journal Of Abnormal Psychology*, *117*(3), 591-605. doi:10.1037/a0012593

Erford, B. T., & Pauletta, D. (2005). Psychometric Analysis of Young Children's Responses to the Slosson Intelligence Test-Primary (SIT-P). *Measurement And Evaluation In Counseling And Development*, *38*(3), 130-140.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143-149. doi:10.3758/BF03203267

Eysenck, S. B., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality And Individual Differences*, *6*(5), 613-619. doi:10.1016/0191-8869(85)90011-X

Fine, A., Steinberg, L., Frick, P. J., & Cauffman, E. (2016). Self-control assessments and implications for predicting adolescent offending. *Journal Of Youth And Adolescence*, *45*(4), 701-712. doi:10.1007/s10964-016-0425-2

Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal Of Experimental Psychology: General*, *133*(1), 101-135. doi:10.1037/0096-3445.133.1.101

Frye, C. C., Galizio, A., Friedel, J. E., DeHart, W. B., Odum, A. L. (2016). Measuring delay discounting in humans using an adjusting amount task. *Journal of Visual Experiments* (107), e53584. doi:10.3791/53584.

Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Personality And Social Psychology Review, 15*(4), 352-366. doi:10.1177/1088868311411165

Fujita, K., & Carnevale, J. J. (2012). Transcending temptation through abstraction: The role of construal level in self-control. *Current Directions In Psychological Science*, *21*(4), 248-252. doi:10.1177/0963721412449169

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci*, *10*(3), 141-146. doi:10.2466/03.27.PMS.120v19x7

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493-503. doi:10.1037/0003-066X.54.7.493

Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal Of Research In Crime And Delinquency, 30*(1), 5-29. doi:10.1177/0022427893030001002

Gray, J. C., Amlung, M. T., Palmer, A. A., & MacKillop, J. (2016). Syntax for calculation of discounting indices from the Monetary Choice Questionnaire and Probability Discounting Questionnaire. *Journal Of The Experimental Analysis Of Behavior*, *106*(2), 156-163. doi:10.1002/jeab.221

Güre, İ. (2012). *Understanding consumers' impulse buying behavior* (Doctoral dissertation).

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19,* 149-161. doi: 10.1007/BF02289162

Gygi, J. T., Hagmann-von Arx, P., Schweizer, F., & Grob, A. (2017). The predictive validity of four intelligence tests for school grades: A small sample longitudinal study. *Frontiers In Psychology*, *8.* doi:10.3389/fpsyg.2017.00375

Hagger, M. S., Panetta, G., Leung, C. M., Wong, G. G., Wang, J. C., Chan, D. K., ... & Chatzisarantis, N. L. (2013). Chronic inhibition, self-control and eating behavior: test of a 'resource depletion' model. *PloS one*, *8*(10), e76888. doi:10.1371/journal.pone.0076888

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495-525. doi:10.1037/a0019486

Hangcheng, L. (2019). ConfirmatoryTetradAnalysis: Functions for Confirmatory Tetrad Analysis. R package version 0.0.0.9000.

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster, C. J. Brainerd, F. N. Dempster, C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 175-204). San Diego, CA, US: Academic Press. doi:10.1016/B978-012208930-5/50007-6

Hipp, J. R., Bauer, D. J., & Bollen, K. A. (2005). Conducting tetrad tests of model fit and contrasts of tetrad-nested models: A new SAS macro. *Structural Equation Modeling*, *12*(1), 76-93. doi:10.1207/s15328007sem1201_4

Hoerger, M., Quirk, S. W., & Weed, N. C. (2011). Development and validation of the Delaying Gratification Inventory. *Psychological Assessment*, *23*(3), 725-738. doi:10.1037/a0023286

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *journal of the American Statistical Association*, *70*(351a), 631-639. doi:1080/01621459.1975.10482485

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kaplan, C. (1996). Predictive validity of the WPPSI—R: A four year follow-up study. *Psychology in the Schools, 33*(3), 211–220. doi:10.1002/(SICI)1520-6807(199607)33:3<211::AID-PITS4>3.0.CO;2-S

Killen, J. D., Taylor, C. B., Hayward, C., Wilson, D. M., Haydel, K. F., Hammer, L. D., & ... Kraemer, H. (1994). Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: A three-year prospective analysis. *International Journal Of Eating Disorders*, *16*(3), 227-238. doi:10.1002/1098-108X(199411)16:3<227::AID-EAT2260160303>3.0.CO;2-L

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal Of Experimental Psychology: General*, *128*(1), 78-87. doi:10.1037/0096-3445.128.1.78

Kline, R. B. (2016). *Methodology in the social sciences. Principles and practice of structural equation modeling (4th ed).* New York, NY, US: Guilford Press.

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental And Clinical Psychopharmacology*, *22*(3), 222-228. doi:10.1037/a0035973

Lane, S. D., Cherek, D. R., Rhodes, H. M., Pietras, C. J., & Tcheremissine, O. V. (2003). Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addictive Disorders & Their Treatment*, *2*(2), 33-40. doi:10.1097/00132576-200302020-00001

Lang, J. B., & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in german-speaking samples. *European Journal Of Psychological Assessment*, *22*(3), 216-224. doi:10.1027/1015-5759.22.3.216

Lindner, C., Nagy, G., & Retelsdorf, J. (2015). The dimensionality of the Brief Self-Control Scale—An evaluation of unidimensional and multidimensional applications. *Personality and Individual Differences*, *86*, 465–473. https://doi.org/10.1016/j.paid.2015.07.006

Lurquin, J. H., & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers In Psychology*, *8*

Magen, E., & Gross, J. J. (2010). The cybernetic process model of self-control: Situation- and person-specific considerations. In R. H. Hoyle, R. H. Hoyle (Eds.), *Handbook of personality and self-regulation* (pp. 353-374). Wiley-Blackwell. doi:10.1002/9781444318111.ch16

Maloney, P. W., Grawitch, M. J., & Barber, L. K. (2012). The multi-factor structure of the Brief Self-Control Scale: Discriminant validity of restraint and impulsivity. *Journal Of Research In Personality*, *46*(1), 111-115. doi:10.1016/j.jrp.2011.10.001

Maul, A. (2013). Method effects and the meaning of measurement. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00169

Maumary-Gremaud, A. (1997). Financial Stress Measure (Fast Track Project Technical Report). Durham, NC: Duke University

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195-244. doi:10.2466/PR0.66.1.195-244

Mischel, W. (1958). Preference for delayed reinforcement: An experimental study of a cultural observation. *The Journal of Abnormal and Social Psychology*, *56*(1), 57-61. doi:10.1037/h0041895

Mischel, W. (1961). Preference for delayed reinforcement and social responsibility. *The Journal Of Abnormal And Social Psychology*, *62*(1), 1-7. doi:10.1037/h0048263

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., & ... Shoda, Y. (2011). 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive And Affective Neuroscience*, *6*(2), 252-256. doi:10.1093/scan/nsq081

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal Of Personality And Social Psychology*, *16*(2), 329-337. doi:10.1037/h0029815

Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal Of Personality And Social Psychology*, *21*(2), 204-218. doi:10.1037/h0032198

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal Of Personality And Social Psychology*, *54*(4), 687-696. doi:10.1037/0022-3514.54.4.687

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*(4907), 933-938. doi:10.1126/science.2658056

Mitchell, L. A., MacDonald, R. A., & Brodie, E. E. (2004). Temperature and the cold pressor test. *The Journal of Pain, 5*(4), 233-237. doi:10.1016/j.jpain.2004.03.004

Monterosso, J., & Luo, S. (2013). Willpower is not synonymous with 'executive function.'. *Behavioral And Brain Sciences*, *36*(6), 700-701. doi:10.1017/S0140525X1300112X

Muraven, M., Shmueli, D., & Burkley, E. (2006). Conserving self-control strength. *Journal Of Personality And Social Psychology*, *91*(3), 524-537. doi:10.1037/0022-3514.91.3.524

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal Of Personality And Social Psychology*, *74*(3), 774-789. doi:10.1037/0022-3514.74.3.774

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition*, *19*(6), 625-666. doi:10.1521/soco.19.6.625.20886

Odum, A. L. (2011). Delay discounting: I'm a K, you're a K. *Journal Of The Experimental Analysis Of Behavior*, *96*(3), 427-439. doi:10.1901/jeab.2011.96-423

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, *14*(2), 150-163. doi:10.1037/1040-3590.14.2.150

Patton, J., Stanford, M. S., & Barratt, E. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal Of Clinical Psychology, 51*(6), 768-774. doi:10.1002/1097-4679

Peake, P. K., Hebl, M., & Mischel, W. (2002). Strategic attention deployment for delay of gratification in working and waiting situations. *Developmental Psychology*, *38*(2), 313-326. doi:10.1037/0012-1649.38.2.313

Peckerman, A., Saab, P. G., Llabre, M. M., Hurwitz, B. E., McCabe, P. M., & Schneiderman, N. (1998). Cardiovascular and perceptual effects of reporting pain during the foot and

forehead cold pressor tests. *International Journal Of Behavioral Medicine, 5*(2), 106-117. doi:10.1207/s15327558ijbm0502_2

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal Of Experimental Psychology*, *32*(1), 3-25. doi:10.1080/00335558008248231

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality And Individual Differences*, *40*(2), 305-315. doi:10.1016/j.paid.2005.03.024

Reyna, V. F., & Wilhelms, E. A. (2017). The gist of delay of gratification: Understanding and predicting problem behaviors. *Journal Of Behavioral Decision Making*, *30*(2), 610-625. doi:10.1002/bdm.1977

Reynolds, B., Richards, J. B., Horn, K., & Karraker, K. (2004). Delay discounting and probability discounting as related to cigarette smoking status in adults. *Behavioural Processes*, *65*(1), 35-42. doi:10.1016/S0376-6357(03)00109-8

Richards, J. B., Zhang, L., Mitchell, S. H., & de Wit, H. (1999). Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal Of The Experimental Analysis Of Behavior, 71*(2), 121-143. doi:10.1901/jeab.1999.71-121

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption: II. *Addiction*, *88*(6), 791-804. doi:10.1111/j.1360-0443.1993.tb02093.x

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2017). Reported self-control does not meaningfully assess the ability to override impulses. Retrieved from osf.io/8etus. doi: 10.17605/osf.io/bxfsu

Schmeichel, B. J., & Zell, A. (2007). Trait self-control predicts performance on behavioral tests of self-control. *Journal Of Personality*, *75*(4), 743-755. doi:10.1111/j.1467-6494.2007.00455.x

Sethi, A., Mischel, W., Aber, J. L., Shoda, Y., & Rodriguez, M. L. (2000). The role of strategic attention deployment in development of self-regulation: Predicting preschoolers' delay of gratification from mother–toddler interactions. *Developmental Psychology*, *36*(6), 767-777. doi:10.1037/0012-1649.36.6.767

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of 'impulsive' behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*(2), 374-408. doi:10.1037/a0034418

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*(6), 978-986. doi:10.1037/0012-1649.26.6.978

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal Of Psychology*, *15*(2), 201-293. doi:10.2307/1412107

Steimke, R., Stelzel, C., Gaschler, R., Rothkirch, M., Ludwig, V. U., Paschke, L. M., & ... Walter, H. (2016). Decomposing self-control: Individual differences in goal pursuit despite interfering aversion, temptation, and distraction. *Frontiers In Psychology*, *7.* doi: 10.3389/fpsyg.2016.00382

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, *47*(1), 1-41. doi:10.1353/mpq.2001.0005

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching Of Psychology*, *44*(1), 24-31. doi:10.1177/0098628316677643

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal Of Experimental Psychology, 18*(6), 643-662. doi:10.1037/h0054651

Tangney, J., Baumeister, R., & Boone, A. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal Of Personality, 72*(2), 271-322. doi:10.1111/j.0022-3506.2004.00263.x

van der Maas, H. L., Kan, K. J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, *2*(1), 12-15. doi:10.3390/jintelligence2010012

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends In Cognitive Sciences, 12*(11), 418-424. doi:10.1016/j.tics.2008.07.005

Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. Journal of Experimental Social Psychology, 48(4), 943–947. https://doi.org/10.1016/j.jesp.2012.03.002

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2014). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Motivation Science*, *1*(S), 19-42. doi:10.1037/2333-8113.1.S.19

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *The American Journal Of Psychiatry*, *150*(6), 885-890.

Watson, S. J., & Milfont, T. L. (2017). A short-term longitudinal examination of the associations between self-control, delay of gratification and temporal considerations. *Personality And Individual Differences*, *106*57-60. doi:10.1016/j.paid.2016.10.023

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal Of Abnormal Psychology*, *103*(2), 192-205. doi:10.1037/0021-843X.103.2.192

Xu, S., Fang, Z., & Rao, H. (2013). Real or hypothetical monetary rewards modulates risk taking behavior. *Acta Psychologica Sinica, 45*(8), 874-886. doi:10.3724/SP.J.1041.2013.00874

APPENDICES

APPENDIX A

Demographic Questionnaire

Age: ____

Gender:
__Female
__Male
__Transgender
Other:____

What is your racial/ethnic identity?
__White/Caucasian
__Black, African-American, or African
__Hispanic or Latino/a
__Middle Eastern
__East Asian or South Asian
__Native Hawaiian or other Pacific Islander
__Native American or American Indian
__Caribbean
Other: _____

What is your sexual orientation?
__Heterosexual
__Gay
__Lesbian
__Bisexual
__Pansexual
__Asexual
Other:_____

Employment Status:
__Employed for wages
__Military
__Out of work and looking for work
__Out of work but not currently looking for work
__Self-employed
__Student
__Retired
__Unable to work

Grades:
In your college courses, what grades do you typically receive?

_Mostly As
_As and Bs
_Mostly Bs
_Bs and Cs
_Mostly Cs
_Cs and Ds
_Ds and Fs
_No Answer

What is your estimated GPA?
_3.5 - 4.0
_3.0 - 3.5
_2.5 - 3.0
_2.0 - 2.5
_1.5 - 2.0
_1.0 - 1.5
_0.5 - 1.0
_0.0 - 0.5

How worried are you that you won't make the grade you want in your psychology classes?

1- Not at all worried
2- A little worried
3- Somewhat worried
4- Quite worried
5- Extremely worried

APPENDIX B

Self-Control Scale

Tangney, J., Baumeister, R., & Boone, A. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal Of Personality, 72*(2), 271-322. doi:10.1111/j.0022-3506.2004.00263.x

Instructions: Please indicate how much each of the following statements reflects how you typically are.

1	2	3	4	5
Not at all				Very much

_____ 1. I am good at resisting temptation.
_____ 2. I have a hard time breaking bad habits. (RC)
_____ 3. I am lazy. (RC)
_____ 4. I say inappropriate things. (RC)
_____ 5. I do certain things that are bad for me, if they are fun. (RC)
_____ 6. I refuse things that are bad for me.
_____ 7. I wish I had more self-discipline. (RC)
_____ 8. People would say that I have iron self-discipline.
_____ 9. Pleasure and fun sometimes keep me from getting work done. (RC)
_____ 10. I have trouble concentrating. (RC)
_____ 11. I am able to work effectively toward long-term goals.
_____ 12. Sometimes I can't stop myself from doing something, even if I know it is wrong. (RC)
_____ 13. I often act without thinking through all the alternatives. (RC)

APPENDIX C

Barratt Impulsiveness Scale

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal Of Clinical Psychology, 51*(6), 768-774. doi:10.1002/1097-4679

People differ in the ways they act and think in different situations. This is a test to measure some of the ways in which you act and think. Read each statement and indicate how often you do each of the following. Do not spend too much time on any statement. Answer quickly and honestly.

1 = Rarely/Never, 2 = Occasionally, 3 = Often, 4 = Almost Always/Always

1 I plan tasks carefully.
2 I do things without thinking.
3 I make-up my mind quickly
4 I am happy-go-lucky.
5 I don't "pay attention."
6 I have "racing" thoughts.
7 I plan trips well ahead of time.
8 I am self controlled.
9 I concentrate easily.
10 I save regularly.
11 I "squirm" at plays or lectures.
12 I am a careful thinker.
13 I plan for job security.
14 I say things without thinking.
15 I like to think about complex problems.
16 I change jobs.
17 I act "on impulse."
18 I get easily bored when solving thought problems.
19 I act on the spur of the moment.
20 I am a steady thinker.
21 I change residences.
22 I buy things on impulse.
23 I can only think about one thing at a time.
24 I change hobbies.
25 I spend or charge more than I earn.
26 I often have extraneous thoughts when thinking.
27 I am more interested in the present than the future.
28 I am restless at the theater or lectures.
29 I like puzzles.
30 I am future oriented.

APPENDIX D

Taste Test Form

This portion of the study is designed to assess your preferences and opinions about different types of food. This information could be used by marketers to better appeal to their audience and advertise the positive qualities of their products. Please take a moment to taste each type of food presented and rank each type of food on the dimensions listed below. You may eat as much as you need to accurately answer the questions below. Ring the bell to notify the experimenter when you have ranked all of the foods.

1) Skittles™

   On a scale of 1 – 10, how much do you like the taste?

   Please describe the taste as best you can.

   On a scale of 1 – 10, how much do you like the texture?

   Please describe the texture.

   On a scale of 1 – 10, how appealing does this food look?

   Please describe the appearance.

2) M&Ms™

   On a scale of 1 – 10, how much do you like the taste?

   Please describe the taste as best you can.

   On a scale of 1 – 10, how much do you like the texture?

   Please describe the texture.

   On a scale of 1 – 10, how appealing does this food look?

   Please describe the appearance.

3) Broccoli

   On a scale of 1 – 10, how much do you like the taste?

   Please describe the taste as best you can.

   On a scale of 1 – 10, how much do you like the texture?

Please describe the texture.

On a scale of 1 – 10, how appealing does this food look?

Please describe the appearance.

4) Cauliflower

On a scale of 1 – 10, how much do you like the taste?

Please describe the taste as best you can.

On a scale of 1 – 10, how much do you like the texture?

Please describe the texture.

On a scale of 1 – 10, how appealing does this food look?

Please describe the appearance.

APPENDIX E

Health Behaviors Questionnaire

Questions are coded on a 7-point Likert Scale:
1 – Never
2 - Rarely/Once a month
3- A few times a month
4 - Once or twice a week
5 – Three to four times per week
6 - Most days
7 - Every day

1) How often do you eat plant protein (lentils, seeds, beans, etc.)?
2) How often do you drink water?
3) How often do you drink soft drinks/soda?
4) How often do you eat fresh fruit?
5) How often do you eat sweets (candy or ice cream)?
6) How often do you eat vegetables or salads?
7) How often do you drink energy drinks?
8) How often do you eat fast food (burgers or fried food)?
9) How often do you eat potato chips or similar snacks?
10) How often do you eat granola or nuts?

Drug use

1) How often do you drink alcohol?
2) How often do you use tobacco products?
3) How often do you use cannabis/marijuana?
4) How often do you use cocaine?
5) How often do you use amphetamine type stimulants (speed, ecstasy)?
6) How often do you use inhalants (nitrous oxide)?
7) How often do you use sedatives or sleeping pills (valium, Rohypnol, etc.)?
8) How often do you use hallucinogens recreationally (LSD, psilocybin, etc.)?
9) How often do you use opioids (heroine, morphine, methadone, codeine etc.)?

Leisure Time Exercise Questionnaire

In a 7-day period (one week), how many times on average do you do the following kinds of exercise for more than 15 minutes during you free time?

1) Strenuous exercise (heart beats rapidly): running, jogging, hockey, soccer, football, basketball, martial arts, swimming, strenuous cycling, etc.
2) Moderate exercise (not exhausting): fast-walking, baseball, tennis, easy cycling, volleyball, badminton, easy swimming, etc.
3) Mild exercise (minimal effort): yoga, archery, fishing, bowling, golf, easy walking, etc.

In a 7-day period, during your free time, how often do you exercise long enough to work up a sweat (heart beats rapidly)?

1) Often
2) Sometimes
3) Never

Impulse Buying Questionnaire

1) How often do you engage in impulse buying (i.e. purchasing something that is not on your shopping list)? □ Never □ Seldom □ Sometimes □ Often □ Always
2) Generally speaking, I would consider myself to be an impulsive shopper. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree
3) I often buy things spontaneously. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree
4) When I shop I tend to decide what I want to buy while I am looking around in a store. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree
5) I carefully plan most of my purchases. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree
6) When I go shopping, I buy things I had not intended to purchase. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree
7) Even when I see something I really like, I do not buy it unless it is a planned purchase. □ Strongly Disagree □ Disagree □ Neither □ Agree or □ Strongly Agree

APPENDIX F

Monetary-Choice Questionnaire

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal Of Experimental Psychology: General*, *128*(1), 78-87. doi:10.1037/0096-3445.128.1.78

For each of the next 27 choices, please indicate which reward you would prefer: the smaller reward today, or the larger reward in the specified number of days.

1. Would you prefer $54 today, or $55 in 117 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

2. Would you prefer $55 today, or $75 in 61 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

3. Would you prefer $19 today, or $25 in 53 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

4. Would you prefer $31 today, or $85 in 7 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

5. Would you prefer $14 today, or $25 in 19 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

6. Would you prefer $47 today, or $50 in 160 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

7. Would you prefer $15 today, or $35 in 13 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

8. Would you prefer $25 today, or $60 in 14 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

9. Would you prefer $78 today, or $80 in 162 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

10. Would you prefer $40 today, or $55 in 62 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

11. Would you prefer $11 today, or $30 in 7 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

12. Would you prefer $67 today, or $75 in 119 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

13. Would you prefer $34 today, or $35 in 186 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

14. Would you prefer $27 today, or $50 in 21 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

15. Would you prefer $69 today, or $85 in 91 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

16. Would you prefer $49 today, or $60 in 89 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

17. Would you prefer $80 today, or $85 in 157 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

18. Would you prefer $24 today, or $35 in 29 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

19. Would you prefer $33 today, or $80 in 14 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

20. Would you prefer $28 today, or $30 in 179 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

21. Would you prefer $34 today, or $50 in 30 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

22. Would you prefer $25 today, or $30 in 80 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

23. Would you prefer $41 today, or $75 in 20 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

24. Would you prefer $54 today, or $60 in 111 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

25. Would you prefer $54 today, or $80 in 30 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

26. Would you prefer $22 today, or $25 in 136 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

27. Would you prefer $20 today, or $55 in 7 days?

[ ] smaller reward today

[ ] larger reward in the specified number of days

APPENDIX G

Anagram Stimuli

On this page are 10 multiple solution anagrams. Multiple solution anagrams are anagrams with more than one solution. For example, TESATR can be reorganized as TASTER, TATERS, and TREATS. This is an anagram with three possible solutions. On this page, *each anagram has at least three solutions.* Please come up with as many solutions as you can. If you run out of ideas and want to move on to the next portion of the study, ring the bell to notify the experimenter.

1) AGNTLEIR (triangle)
2) TEHRA (earth)
3) LMAAIN (animal)
4) NRMAOO (maroon)
5) RSPDEES (depress)
6) MTISE (times)
7) ALIRNE (linear)
8) UASTDNHO (thousand)
9) YPMOAINL (Olympian)
10) WOELT (towel)

APPENDIX H

Alcohol Use Disorder Identification Test

1) How often do you have a drink containing alcohol? (0) Never (1) Monthly or less (2) 2 to 4 times a month (3) 2 to 3 times a week (4) 4 or more times a week.
2) How many drinks containing alcohol do you have on a typical day when you are drinking? (0) 1 or 2 (1) 3 or 4 (2) 5 or 6 (3) 7, 8, or 9 (4) 10 or more.
3) How often do you have six or more drinks on one occasion? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily.
4) How often during the last year have you found that you were not able to stop drinking once you had started? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily
5) How often during the last year have you failed to do what was normally expected from you because of drinking? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily.
6) How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily.
7) How often during the last year have you had a feeling of guilt or remorse after drinking? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily.
8) How often during the last year have you been unable to remember what happened the night before because you had been drinking? (0) Never (1) Less than monthly (2) Monthly (3) Weekly (4) Daily or almost daily.
9) Have you or someone else been injured as a result of your drinking? (0) No (2) Yes, but not in the last year (4) Yes, during the last year.
10) Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down? (0) No (2) Yes, but not in the last year (4) Yes, during the last year.

APPENDIX I

Weight Concerns Scale

1) How much more or less do you feel you worry about your weight and body shape than other students your age?

1- I worry a lot less than other students.
2- I worry a little less than other students.
3- I worry about the same as other students.
4- I worry a little more than other students.
5- I worry a lot more than other students.

2) How afraid are you of gaining 3 pounds?

1- Not afraid
2- Slightly afraid
3- Moderately afraid
4- Very afraid
5- Terrified

3) When was the last time you went on a diet?

1- I've never been on a diet.
2- I was on a diet about one year ago.
3- I was on a diet about 6 months ago.
4- I was on a diet about 3 months ago.
5- I was on a diet about 1 month ago.
6- I was on a diet less than 1 month ago.
7- I'm now on a diet.

4) Compared to other things in your life, how important is your weight to you?

1- My weight is not important compared to other things in my life.

2- My weight is a little more important than some other things.

3- My weight is more important than most, but not all, things in my life.

4- My weight is the most important thing in my life.

5. Do you ever feel fat?

1- Never

2- Rarely

3- Sometimes

4- Often

5- Always

APPENDIX J

Achievement Motives Scale

1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree

1) I like situations in which I can find out how capable I am.
2) When I am confronted with a problem, which I can possibly solve, I am enticed to start working on it immediately.
3) I enjoy situations in which I can make use of my abilities.
4) I am appealed by situations allowing me to test my abilities.
5) I am attracted by tasks in which I can test my abilities.
6) I am afraid of failing in somewhat difficult situations, when a lot depends on me.
7) I feel uneasy to do something if I am not sure of succeeding.
8) Even if nobody would notice my failure, I'm afraid of tasks which I'm not able to solve.
9) Even if nobody is watching, I feel quite anxious in new situations.
10) If I do not understand a problem immediately I start feeling anxious.

APPENDIX K

Financial Stress Questionnaire

1 = strongly agree, 2 = agree, 3 = neutral/mixed, 4 = disagree, 5 = strongly disagree

1) My family has enough money to afford the kind of home we would like to have.
2) We have enough money to afford the kind of clothing we should have.
3) We have enough money to afford the kind of furniture or household equipment we should have.
4) We have enough money to afford the kind of car we need.
5) We have enough money to afford the kind of food we should have.
6) My family has enough money to afford the kind of leisure and fun activities we want to participate in.
7) We have enough money to afford the kind of medical care we should have.
8) We are able to pay the bills that we need to pay.
9) We generally have enough money to “make ends meet”.

VITA

PARKER DREVES

Education:	M.A., Psychology, East Tennessee State University, Johnson City, Tennessee 2017
	B.S., Psychology, East Tennessee State University, Johnson City, Tennessee 2015
	Franklin High School, Franklin, Tennessee 2011
Professional Experience:	Instructor, East Tennessee State University; Johnson City, Tennessee, 2017 – 2018
	Graduate Assistant, East Tennessee State University; Johnson City, Tennessee, 2015 – 2019
	Student Worker, East Tennessee State University; Johnson City, Tennessee, 2011 - 2013
Honors and Awards:	Outstanding Thesis in the Social Sciences and Education Award, East Tennessee State University, 2018
	Excellence in Teaching Award, East Tennessee State University, 2017
	Accepted into Honors in Discipline program, East Tennessee State University, 2013
	Academic Performance Scholarship recipient, East Tennessee State University, 2011