Behaviour Research and Therapy 101 (2018) 30–45

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Measuring self-efficacy, executive function, and temporal discounting in Kenya

Kristina Esopo[a,1], Daniel Mellow[b,1], Catherine Thomas[c,2], Hannah Uckat[d,2], Justin Abraham[e], Prachi Jain[f], Chaning Jang[b], Nicholas Otis[g], Michala Riis-Vestergaard[a], Amanda Starcev[b], Kate Orkin[b,d,3], Johannes Haushofer[a,b,h,*,3]

[a] Department of Psychology, Princeton University, USA
[b] Busara Center for Behavioral Economics, Nairobi, Kenya
[c] Department of Psychology, Stanford University, USA
[d] Department of Economics, University of Oxford, UK
[e] Department of Economics, University of California, San Diego, USA
[f] Department of Economics, Loyola Marymount University, USA
[g] Department of Public Health, University of California, Berkeley, USA
[h] Woodrow Wilson School for Public and International Affairs & Department of Economics, Princeton University, USA

ARTICLE INFO

Keywords:
Temporal discounting
Self-efficacy
Executive function
Measurement
Psychometrics

ABSTRACT

Developing countries have low adherence to medical regimens like water chlorination or antenatal and postnatal care, contributing to high infant and child mortality rates. We hypothesize that high levels of stress affect adherence through temporal discounting, self-efficacy, and executive control. Measurement of these constructs in developing countries requires adaptation of existing measures. In the current study, we adapt psychological scales and behavioral tasks, measuring each of these three constructs, for use among adults in Kenya. We translated and back-translated each measure to Kiswahili and conducted cognitive interviewing to establish cultural acceptability, refined existing behavioral tasks, and developed new ones. Then, in a laboratory session lasting 3 h, participants ($N = 511$) completed the adapted psychological inventories and behavioral tasks. We report the psychometric properties of these measures. We find relatively low reliability and poor correlational evidence between psychological scales and behavioral tasks measuring the same construct, highlighting the challenges of adapting measures across cultures, and suggesting that assays within the same domain may tap distinct underlying processes.

1. Introduction

The infant mortality rate, defined as the probability of dying before age 1, is 32 per 1000 live births worldwide, with 60% of these deaths occurring in the first 28 days of life (WHO, 2014). Despite improvement, the global maternal mortality rate remains at 210 per 100,000 live births (WHO, 2014). In developing countries such as Kenya, these figures are even higher, with an infant mortality rate of 55 per 1000 live births and a maternal mortality rate of 362 deaths per 100,000 live births (Kenya National Bureau of Statistics et al., 2014).

Two factors are thought to play an important role in accounting for these figures: contaminated drinking water (Carroli, Rooney, & Villar, 2001; McDonagh, 1996) and insufficient antenatal and postnatal care (ANC/PNC) takeup. Contaminated drinking water is the main cause of diarrheal diseases in developing countries and is estimated to cause 2100 child deaths every day worldwide (WHO, 2014). Chlorination of drinking water is a highly effective method to prevent diarrhea and is widely available at low cost in developing countries, such as Kenya. For instance, a bottle of WaterGuard to treat 1000l of water costs USD 0.26 PPP in Kenya, which is less than the typical household spends on sugar every day. Nevertheless, only about 5% of households chlorinate their drinking water (Kremer, Miguel, Mullainathan, Null, & Zwane, 2009). Similarly, the standard ANC/PNC regime for women in Kenya calls for two or more doses of SP/Fansidar, an anti-malarial medication;

* Corresponding author. Princeton University, Department of Psychology, 427 Peretsman Scully Hall, Princeton, NJ 08544, USA.
E-mail address: haushofer@princeton.edu (J. Haushofer).
[1] Contributed equally.
[2] Contributed equally.
[3] Contributed equally.

https://doi.org/10.1016/j.brat.2017.10.002
Received 9 March 2017; Received in revised form 3 October 2017; Accepted 4 October 2017
Available online 17 November 2017
0005-7967/

however, only 17% of pregnant women report adhering (Kenya National Bureau of Statistics et al., 2014). As a result, malaria infections are frequent and contribute to low birth weight and increased infant mortality.

Thus, adherence to water chlorination and ANC/PNC takeup is low in developing countries, despite significant benefits. Using a mechanisms-focused, experimental medicine approach to behavior change, we hypothesize that stressors related to poverty may contribute to low adherence by affecting three behaviors: *temporal discounting, self-efficacy, and executive control*. These targets are of interest because a) they are likely to be affected by stress, and b) they are likely to affect regimen adherence. The next section clarifies these links in more detail. In addition, these targets map neatly to models of decision-making in economics: in these models, the motives for behavior are fully characterized by preferences over outcomes, beliefs about oneself and the world, and the constraints one faces. Because temporal discounting is a preference, self-efficacy a belief measure, and executive control a constraint on decision-making, the three targets cover each of these three determinants of behavior in economic models.

In the broader context of the Science of Behavior Change (SOBC) framework, which aims to identify specific, malleable targets that are hypothesized to be relevant to behavior change, the goal of the current study is to identify, refine, and test the psychometric properties of a set of psychological scales and behavioral tasks that measure these targets. We subsequently use the results presented here to inform our selection of target measures in later phases of the project: testing the effects of stress on the three targets; developing interventions to engage the targets; and testing whether the engagement of these targets affects water chlorination and ANC/PNC regimen adherence.

In the following, we briefly discuss how stress may affect each target, and how the targets may affect adherence.

1.1. Target 1: Temporal discounting

Our first candidate mechanism through which stress may affect adherence is temporal discounting, defined as the loss of value that rewards undergo as they are delayed into the future. A recent meta-analytic review by Fields, Lange, Ramos, Thamotharan, and Rassu (2014) found that discounting was related to stress with a moderate to large effect size. Indeed, our own work under SOBC-1 has shown that stress can focus individuals on the present in economic choice: after administration of 20 mg of hydrocortisone, which raises cortisol levels, subjects in our study showed increased temporal discounting, i.e. they were less willing to forgo smaller immediate rewards in favor of larger future rewards (Riis-Vestergaard et al., in press). Note, however, that we found no evidence that a physical stressor (i.e. the cold pressor task) and a social stressor (i.e. the Trier Social Stress Test) affect temporal discounting (Haushofer et al., 2013, 2015), raising the possibility that different types of stress may have different effects on discounting.

It is easy to see that an effect of stress on discounting may also negatively affect adherence, which requires incurring an immediate cost (e.g. traveling to a clinic, the discomfort of taking medications and their side effects) for a greater but delayed benefit (e.g. a healthy child). On this view, stress may decrease the attractiveness of the delayed benefit, or increase the disutility from the immediate cost, and thus reduce adherence (Fields, Ramos, & Reynolds, 2015). Indeed, in developed countries, high temporal discounting has been shown to be negatively associated with adherence to recommended screening regimes for cholesterol, breast, cervical, and prostate cancer and use of dental care, flu shots, and physical exercise (Bradford, 2010). Discounting is also positively associated with adherence-related adverse health behaviors, such as binge eating (Davis, Patte, Curtis, & Reid, 2010) and addiction and substance abuse (Andrade & Petry, 2012; Bickel & Marsch, 2001; Bickel et al., 2007; MacKillop et al., 2011; Reynolds, 2006; Rogers, Moeller, Swann, & Clark, 2010). In Kenya, time preferences have been shown to predict mortality among HIV-infected adults receiving antiretroviral therapy (ART); those with higher discount rates also had lower ART adherence, although the association was not statistically significant (Thirumurthy et al., 2015).

It is important to note that there are several context-dependent reasons why people may engage in discounting behavior, possibly reflecting environmental constraints in contexts of poverty. For example, Becker and Mulligan (1997) show that economic conditions, such as poverty, and environmental influences, such as mortality and risk, can endogenously lead to behavior that looks like impatience. Similarly, Carvalho, Meier, and Wang (2016) show that low-income participants have higher discount rates when making choices about monetary rewards before payday. Credit market imperfections may also explain behavior that resembles high discounting rates in developing countries (Banerjee, 2001; Holden, Shiferaw, & Wik, 1998; Pagiola, 1996; Shiferaw & Holden, 2001). Temporal discounting may also be partly driven by beliefs about environmental constraints and the likelihood of positive distal outcomes (Laajaj, 2012). In line with these arguments, we hypothesize that stress is a contextual factor that increases temporal discounting in developing countries.

1.2. Target 2: Self-efficacy

The second candidate mechanism is self-efficacy, defined as the belief that one can perform well in specific situations (Bandura, 1982). We hypothesize that stress may affect adherence and other health behaviors by decreasing an individual's perceived sense of control (i.e. personal mastery), and consequently lead to low self-efficacy.

Self-efficacy is linked to stress as an important determinant of individual responses to stressors (Bandura, 1988); numerous studies in Western contexts have demonstrated that those with high self-efficacy can cope better with stressors and trauma than those with low self-efficacy (Benight & Bandura, 2004; Perkins & Jenkins, 1998; Shnek et al., 1997; Shorey, Chan, Chong, & He, 2015; Tan-Kristanto & Kiropoulos, 2015). Self-efficacy has also been shown to be strongly related to adherence to medical regimens and other health behaviors. In a recent comprehensive meta-analysis of 207 studies on adherence to ART drugs for HIV, Langebeek et al. (2014) find that self-efficacy is the single strongest predictor of adherence, with an effect size more than 50% larger than the next-best predictor (substance abuse). In line with this finding, interventions targeted at improving self-efficacy have been shown to increase adherence to exercise regimens (Azizan, Justine, & Kuan, 2013; Barkley & Fahrenwald, 2013; Seghers, Van Hoecke, Schotte, Opdenacker, & Boen, 2014). In a recent study from a developing country context, Ghosal, Jana, Mani, Mitra, and Roy (2013) show that a training program for building “agency” (closely related to self-efficacy) among sex workers in India strongly increased self-efficacy and raised the likelihood of having undergone a health checkup in the last month by nine percentage points. We therefore build on previous work showing that self-efficacy is an integral part of behavior change by asking if low levels of self-efficacy are part of the mechanism through which stress lowers adherence to chlorination and ANC/PNC services in Kenya.

1.3. Target 3: Executive control

The final mechanism we consider is executive control. Executive control is a broad term commonly referring to the maintenance and execution of high-level plans and goals, and involves planning, cognitive flexibility, inhibitory control, and working memory processes. We combine these different concepts because deficits in these abilities may affect adherence and other health behaviors through a similar mechanism, namely a failure to make a plan or follow through on it. For instance, when faced with the task of attending a doctor's appointment, an individual might simply forget about it, be distracted by other tasks and therefore fail to attend, or fail to make a plan to go.

Several lines of inquiry have provided evidence suggesting that stress affects executive control. Early studies showed that stress impairs performance in cognitive control tasks, such as the Stroop task (Hartley & Adams, 1974), attentional selectivity and control (Hockey, 1970; Liston, McEwen, & Casey, 2009; Minor, Jackson, & Maier, 1984), cognitive flexibility (Alexander, Hillier, Smith, Tivarus, & Beversdorf, 2007), and working memory (Luethi, Meier, & Sandi, 2008). More recently, Mullainathan and Shafir (2013) have introduced the term "scarcity" to refer to a cognitive form of stress, induced in contexts of limited resources, that produces characteristic behaviors such as persistent tradeoff thinking. It has emerged in this fledgling literature that stress induced in this fashion amongst the poor can impair fluid intelligence on a Raven's Matrices test, as well as performance in a Stroop-like executive control task (Mani, Mullainathan, Shafir, & Zhao, 2013; Mullainathan & Shafir, 2013). A final group of studies has shown that stress impairs goal-directed decision-making relative to habit-based decision-making; in other words, both humans and animals fall back on default strategies under stress (Dias-Ferreira et al., 2009; Schwabe, Tegenthoff, Höffken, & Wolf, 2010; Schwabe & Wolf, 2009, 2010). Together, these findings suggest that stress may affect adherence by way of impaired executive control.

In line with the view that executive control may undermine adherence, a number of studies have shown that supporting processes related to executive control, such as memory, planning, and task monitoring, can improve adherence (Brock, Brock, & Thiedke, 2011). The best evidence comes from the effectiveness of reminders. For instance, Karlan, McConnell, Mullainathan, and Zinman (2010) find positive effects of reminders on savings rates, Duflo, Kremer, and Robinson (2011) on fertilizer use, and Banerjee, Duflo, Glennerster, and Kothari (2010) on vaccination take-up. Ahuja, Kremer, and Zwane (2010) and Kremer et al. (2009) show that providing a chlorine dispenser at the source where Kenyan villagers fetch their water dramatically increases chlorination rates due to the visual reminder that the dispenser provides. On the other hand, in a recent Cochrane Review (Nieuwlaat et al., 2014), only one of the five studies that tested the effect of mobile text message reminders on medical regimen adherence found a significant increase in adherence (Lester et al., 2010), while the remainder found no improvement (Boker, Feetham, Armstrong, Purcell, & Jacobe, 2012; Hou, Hurwitz, Kavanagh, Fortin, & Goldberg, 2010; Simoni et al., 2009; Zolfaghari, Mousavifar, Pedram, & Haghani, 2012). Notably, however, those studies reporting no improvement were conducted in high-income countries, whereas Lester et al. (2010) find a positive impact of reminders on ART adherence in Kenya. These results suggest that failures to remember may have more severe consequences in lower-income contexts, where fewer tasks have built-in reminders, thus making exogenous reminders more effective. Indeed, Rodrigues et al. (2012) find that forgetfulness was the most common reason for non-adherence to ART in South India, but also show that forgetfulness decreased rapidly as mobile phone reminders were introduced. Several similar studies in Kenya on mobile text message or alarm reminders, targeting deficits in memory, have also found improvements on health workers' adherence to malaria treatment guidelines (Zurovac et al., 2011), ART adherence (Pop-Eleches et al., 2011), and multi-vitamin medication adherence (Frick, Lavreys, Mandaliya, & Kreiss, 2001).

1.4. Measuring the targets

The present paper describes and tests a battery of questionnaires and behavioral tasks that measure these constructs among urban poor populations in Nairobi, Kenya. Since many questionnaires and behavioral tasks assessing temporal discounting, self-efficacy, and executive control were developed in English-speaking, Western, industrialized countries, we used cross-cultural validation procedures to adapt the scales to the Kenyan context (Eremenco, Cella, & Arnold, 2005; Ferraz, 1997). For the five psychological scales, we use the standard psychometric criteria of reliability, validity, and acceptability. For behavioral tasks, we analyze test-retest reliability and hypothesized relationships to other scales and tasks, as well as participant comprehension and logical consistency of answers. Finally, we examine correlations between scales and tasks believed to be tapping the same construct. All analyses presented were described in a pre-analysis plan (Haushofer & Orkin, 2017), except for the exploratory factor analysis, which was suggested to us after the analysis plan had been finalized. For brevity, we omit pre-specified analyses of Convex Time Budget and Multiple Price List tasks, as well as item validation and reduction; results are available upon request. Results from the Risk task could not be analyzed due to a data collection error.

We combine psychological self-report scales with incentivized behavioral tasks because each approach has strengths and weaknesses. Self-report questionnaires are often cheaper and easier to collect than behavioral tasks, which are usually incentivized and computer-administered. On the other hand, self-reports are more likely to be subject to social desirability bias or experimenter demand effects, acquiescence bias (agreeing with all questions in a measure), and nay-saying (denying all or not endorsing any statements in a measure) (Furnham, 1986). Further, bias can arise as a result of question phrasing, ordering, and response options presented (Kalton & Schuman, 1982). Such biases are thought to be reduced in incentivized behavioral tasks.

In addition, combining psychological self-report scales with incentivized behavioral tasks that aim to measure the same construct enables us to examine correlations between the two modes of elicitation, which can inform which measures to use in the field. If the correlation is high, survey measures might be preferable as they are cheaper and easier to implement in the field; on the other hand, low correlations between measures might indicate poor cross-cultural adaptation.

2. Methods

2.1. Participants and study procedure

The study was conducted at the Busara Center for Behavioral Economics in Nairobi, Kenya. Between October 2016 and January 2017, we recruited 511 (47% women, 53% men) adult residents of Kibera – a large informal urban settlement in Nairobi, Kenya, located 3 km from the lab — for a session lasting approximately 3 h. To be eligible, a prospective participant needed to have signed up for the Busara participant database, be at least 18 years of age, and have access to a phone and an MPesa mobile money account (a widely used mobile payment system provided by the mobile phone operator "Safaricom" through which participants are paid for their participation in the study). In this sample, the median age was 29 years (range: 18–40); 15.6% were unemployed while 84.4% were employed or self-employed; 66.1% completed secondary level education and 27.8% education beyond secondary level; and average income reported was KES 6918 (approximately USD 69, i.e. USD 1 was equal to ~ KES 100 at the time of the study) per month.

Using the inclusion criteria above, participants were drawn from Busara's participant pool, which is broadly representative of Nairobi and Kenya (Haushofer et al., 2014). Each session included up to 20 participants. Tasks and scales were presented on touchscreen computers using the zTree experimental interface (Fischbacher, 2007), on which the participants were briefly trained before the session and as needed before some tasks. Each session ended with a demographic questionnaire. The order of constructs (i.e. discounting, executive control, self-efficacy) was randomized at the session level, but within each construct, the task(s) always preceded the questionnaire(s). When two questionnaires related to the same construct (e.g. for self-efficacy and discounting), the order of the questionnaires was randomized at the session level. Participants received a KES 250 show-up fee, which is slightly above the average daily wage earned in this context, for their participation in the study. Performance in the tasks was incentivized;

the average total payment earned as a result of participation in the study was KES 790.

The last 95 participants of the study were invited to a second session in which they completed the same scales and tasks one week later. These data form the basis of our test-retest reliability measures described below.

2.2. Psychological scales and behavioral tasks

2.2.1. Selection

Each of the psychological scales and behavioral tasks, measuring our three constructs of executive control, self-efficacy, and temporal discounting, was selected through a literature review, which compared the psychometric properties reported for the most commonly cited measures in each domain, respectively. The literature review was conducted using Google Scholar by directly searching the terms of interest (e.g. “temporal discounting”, “self-efficacy”, “executive control”) alongside associated or interchangeable terms (e.g. “time preference”, “personal mastery”, “executive function”, etc.) and combining these with one term related to psychometric assessment (e.g. “measure”, “scale”, “validity”, “reliability”, “psychometric properties”, etc.). Full text, original articles and reviews published in English were included so long as they reported on at least one psychometric property of the instrument (e.g. construct validity, internal reliability, test-retest reliability, predictive validity in behavior change research, etc.). Articles had to concern the development or evaluation of the measurement properties of self-reported questionnaires or behavioral tasks, assessing our three constructs across a variety of samples. For temporal discounting, 34 studies evaluating 19 instruments met our criteria. For self-efficacy, 27 studies evaluating 10 instruments met our criteria. For executive control, 31 studies evaluating 10 instruments met our criteria. Most studies concerned instruments with known validity in Western populations. At least one self-report questionnaire and one behavioral task for each outcome measure with the greatest consensus for use among experts, as determined by the number of citations and assessment of psychometric properties, were chosen to be subsequently validated in Kenya in the current study. The full list of measures considered and their relative strengths on the criteria for inclusion are available in Appendix Table A6. We report more specific information about the psychometric properties of the chosen instruments in the following section.

2.2.2. Translation

Translation and adaptation of scales was achieved as follows. First, the text of both task instructions and scale items were translated and back-translated to Kiswahili, the lingua franca in Nairobi. Then, in the pilot stage of the study, we conducted “cognitive interviewing” with 16 participants, representative of the target population in Nairobi, to assess cultural acceptability and reduce the risk of response bias as a result of question phrasing, ordering, and response options presented. Specifically, following the manual developed by Willis (1999), cognitive interviewing consisted of one-on-one interviews, in which a field officer read each scale item and asked the participant several comprehension questions to explain how they arrived at their answer and any problems encountered when answering, including expressions that may have been difficult to understand or considered offensive.

To check for acquiescence bias influencing responses, at least two items with “reverse” coding were either included in the generic versions of each scale, or added by us if no reverse items were included in the original scale. In the latter case, we proceeded as follows: the final item was reversed and became the new first item, while the original first item was reversed and added as the last item. All scales were scored as the sum of values assigned to each answer choice, with adjustment for reversed items.

In concert with the psychological scales, we use five incentivized behavioral tasks to measure our constructs of interest. All prompts and answer choices appeared on the screen in both English and Kiswahili; in addition, to address potential issues of illiteracy, which can be common amongst this population, the instructions and each question were read aloud in Kiswahili by the enumerator. We describe each measure and discuss the rationale behind choosing each measure in the section that follows below. The appendix contains a comprehensive list of all items included for each of the psychological scales we adapted, as well as examples of a participant's screen from each of the behavioral tasks.

2.2.3. Scales and tasks measuring executive control

Executive control refers to “a set of inter-related higher-order cognitive abilities involved in self-regulatory functions” (Roth, Lance, Isquith, Fischer, & Giancola, 2013), such as insight, judgement, working memory, or planning (Baddeley, Della Sala, Robbins, & Baddeley, 1996; Royall et al., 2002; Van der Linden, Frese, & Meijman, 2003). To measure aspects of executive control most relevant to adherence behavior, including inhibitory control, memory, planning, and task monitoring, we adapt one scale and two behavioral tasks.

- The psychological scale is the *Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A)*, a 75-item questionnaire using nine non-overlapping theoretically and empirically derived clinical subscales that measure various aspects of executive function (Roth, Isquith, & Gioia, 2005). We chose the BRIEF-A because of its well-established psychometric properties in a sample of 1050 adults in the United States: Cronbach's alpha coefficients ranging from 0.93 to 0.96, 1-month test-retest reliabilities ranging from r = 0.93 to 0.94 for the three major indices, and evidence of convergent and discriminant validity (Roth et al., 2005, 2013). The full questionnaire yields an overall score (Global Executive Composite) comprised of two index scores, Behavioral Regulation and Metacognition. We include 32 questions from the original BRIEF-A, comprising four subscales of the Metacognition index, which we hypothesize are most relevant to adherence behavior in this context: *1. Initiate*, which reflects an individual's ability to begin a task or activity and to independently generate ideas, responses, or problem-solving strategies; *2. Working Memory*, which measures the capacity to hold information in mind for the purpose of completing a task, encoding information, or generating goals, plans, and sequential steps to achieving goals; *3. Plan/Organize*, which measures an individual's ability to manage current and future-oriented task demands; and *4. Task Monitor*, which reflects the ability to keep track of one's problem-solving success or failure and to identify and correct mistakes during behaviors. Participants are presented with statements which exhibit examples of poor executive function, such as “I have trouble starting anything on my own” or “I don't plan early for future activities.” The response choices range from “never a problem” (scored as 0) to “always a problem” (6). In addition to the 32 items, we add two items with repeated content and reversed wording to check for acquiescence bias.

The two behavioral tasks are as follows:

- To measure inhibitory control, the ability to control one's attention, behavior, thoughts, and/or emotions to override automatic responses and selectively attend to one stimulus over another (Diamond, 2013), we adapt a *spatial version of the Stroop task*, using congruent and incongruent directional signals (arrows) rather than words (Wühr, 2007). We chose this task, rather than its numbers or letters analog, because it does not require literacy. In addition, Busara had previously piloted it for cultural acceptability and understanding with the target sample. On each screen, participants see a colored arrow that points either left or right and respond by pressing a box on the left or right side of the screen. Importantly, when the arrow is red, participants are required to select the side of the screen towards which it points (“congruent” trials); if the arrow

is blue, they are required to select the opposite side of the screen ("incongruent" trials). The sequence of arrows was randomized. Participants earned KES 25 for each correct response, but lost KES 3 for every second they took to complete the task (although the total payment for this task could not go below zero). We record correct and incorrect responses and reaction times by trial type. Significantly longer response times and lower frequency of correct responses to incongruent stimuli are interpreted as evidence of a Stroop Effect. For purposes of establishing construct validity and inter-construct relationships, we define overall performance on the Stroop task to be equal to the ratio of number of correct responses to total time in seconds. An example of the participant's screen is provided by Fig. A.1.

- Since successful adherence to health regimens requires the ability to successfully make a plan (Stilley, Bender, Dunbar-Jacob, Sereika, & Ryan, 2010), we also implemented a version of the *Tower of London task* (TOL; also known as the Stockings of Cambridge task when implemented electronically), which is designed to measure a participant's ability to plan ahead in sequential strategies (Phillips, Wynn, McPherson, & Gilhooly, 2001; Shallice, 1982). In our computerized version of the Tower of London task, participants see a screen with two parts: on the left side is the word "start" with a picture of three "pegs" and various shapes positioned on the pegs; on the right side is the word "goal" with a similar picture of three "pegs" and the same shapes positioned differently on the pegs. To complete the task, participants must reposition the shapes underneath the "start" on the left to match the "goal" position on the right. They are instructed to complete each round in as few moves as possible, with the minimum number of moves shown as a number on the screen. In addition to a practice round, participants attempt four rounds of increasing complexity, beginning with one shape requiring only one move, and concluding with three shapes in a pattern that necessitates at least four moves. For each trial, we record the number of moves, the time until the participant's first move, the overall time to completion, and whether the problem is solved correctly. In all rounds, participants are limited to a maximum of 20 moves; if this occurs, the round ends and the participant is required to contact a staff member to ensure she understands the task before continuing to the next round. Therefore, the distribution of scores is censored at both ends. Performance on the Tower of London task, for the purpose of establishing construct validity and reliability, is computed as the total number of moves used across the four rounds, the number of rounds completed correctly, and standardized average time to complete rounds. An example of the participant's screen is shown in Fig. A.2.

2.2.4. Scales and tasks measuring perceived self-efficacy

We measure self-efficacy with two psychological scales, tapping self-efficacy and mastery, respectively:

- We adapt the uni-dimensional *Generalized Self-Efficacy (GSE) scale* (Schwarzer & Jerusalem, 2010), which was chosen based on prior multi-cultural validation studies and evidence of strong psychometric properties across cultures (Luszczynska, Scholz, & Schwarzer, 2005). Our version contains 12 items, 10 from the generic version and two which are repeated and reversed. Participants are asked to rate the truthfulness of statements such as "I can always manage to solve difficult problems if I try hard enough" on a scale from "never true" (0) to "always true" (5). The creators of the scale operationalize their construct definition as "the belief of an individual in his or her ability to respond to any sort of difficult situation and cope with unforeseen setbacks" (Luszczynska et al., 2005).
- The *Pearlin Mastery Scale (PMS)* is another canonical measure in self-efficacy research. We chose to include this measure based on its relationship with stress, depression, and other health outcomes (Marshall & Lang, 1990a; Mausbach et al., 2007). It is defined by its author as measuring *mastery*, or "the extent to which one regards one's life-chances as being under one's own control in contrast to being fatalistically ruled" (Pearlin & Schooler, 1978). Participants are asked to identify with seven statements about self-efficacy on a scale ranging from "strongly disagree" (0) to "strongly agree" (5). Of these, five statements are negative (e.g. "I have little control over the things that happen to me") and two positive (e.g. "what happens to me in the future mostly depends on me"). The latter are adjusted accordingly in analysis.

Given a gap in the literature to measure self-efficacy with a behavioral task (the closest task approximation we found was the "learned helplessness" experimental manipulation developed in Hiroto (1974)), we propose a novel behavioral self-efficacy task. The task is structured as follows:

- We operationalize self-efficacy as having "high" beliefs about one's ability to complete a task and being approximately correct about these beliefs. The rationale is as follows: first, core to the self-efficacy concept is the ability to achieve desired outcomes; hence, the measure should increase in actual performance. Second, self-efficacy is distinct from overconfidence (and underconfidence): a person who has "high" beliefs about their ability, but actually has low ability, is better characterized as overconfident than has having high self-efficacy. The converse argument applies for "low" beliefs.
- Self-efficacy represents a belief about performance on a particular task. Here, we use the "*slider task*" (Gill & Prowse, 2012). Participants are shown an on-screen "slider," a horizontal line which represents the integers from 0 to 100 or 0 to 20. (Note that we altered the task approximately midway through the study ($N_1 = 283$, $N_2 = 228$), so that the slider would have 20 possible integers instead of 100. This was done to raise average number of sliders completed and, therefore, eliminate floor effects, as summary statistics compiled halfway through the study indicated that 3–4% of participants were unable to complete any sliders.) They are then instructed to click the point on the line which corresponds to a randomly selected specific integer on the line (i.e. if the integer on the screen is 19, the participant must position the slider to the corresponding integer 19). The corresponding slider integer selected is then shown on the screen, and the participant can elect to move on or keep trying until they have made a match. An example of the participant's screen is provided in Fig. A.3.
- Once they select the correct number, or elect to move on, they are presented with another randomly chosen integer they need to match. This "slider matching" process has the advantage of simulating effort, which is purely mechanical and, therefore, should be not be related to age and education. After a 60-s practice round, participants proceed to a 3-min round during which they are paid KES 10 for each slider matched.
- After this incentivized round, participants are asked to estimate their performance in the first round (in terms of total sliders matched), with a correct guess worth KES 50, as well as how confident they are of their estimate (unincentivized). Participants do not receive feedback on their performance to avoid changes in self-efficacy due to feedback.
- Next, as the core element of the task, participants are asked to set a goal for how many sliders they want to match in the following 2-min round, as well as their confidence level concerning that goal. Participants are informed that the payment will increase by KES 20 times the number of sliders indicated by the goal only if they achieve the goal, and nothing otherwise. Thus, a participant who indicates that they believe they can complete x sliders and actually completes at least that many sliders is paid KES 20 times their goal, even if they complete more sliders.
- Lastly, the final round is played and payment calculated. We define our measure of self-efficacy as:

$$SX = \frac{y_a}{1 + |y_a - y|} \quad (1)$$

Here, the goal set is denoted by y_a, and actual performance by y. Note that this measure increases in the goal set, in performance, and in accuracy about the goal.

2.2.5. Scales and tasks measuring temporal discounting

We employ two inventories from the psychology literature that measure two distinct concepts related to temporal discounting, namely consideration of future consequences (i.e. future orientation) and deferred gratification:

- Consideration of future consequences, which quantifies the extent to which individuals consider potential future outcomes of their current behavior, is predictive of a number of health behaviors (Chapman, 2005), which makes it especially relevant to adherence. To measure this construct, we use the *Consideration of Future Consequences (CFC) scale*, a common, cross-culturally validated measure with attractive psychometric properties: Cronbach's α coefficients ranging from 0.80 to 0.86 across four different samples, with two-week and five-week test-retest reliabilities of r = 0.76 and 0.72, respectively (Strathman et al., 1994). Participants are asked to indicate how much the behavior described in a statement is characteristic of them, from “not at all like me” (0) to “very much like me” (5). There are nine statements representative of forward thinking (e.g. “I am ready to sacrifice my current happiness or wellbeing in order to achieve future results”) and five reverse statements (e.g. “I only act to satisfy immediate needs, thinking the future will take care of itself”), which are scored accordingly.
- As an alternative measure to CFC we use the *Deferment of Gratification Scale (DGS)*, which assesses the ability to resist the temptation of an immediate reward and instead wait for a larger, later reward (Carducci, 2009). The DGS comprises two factors relevant to adherence behavior: controlling impulses and planning and waiting. This scale is particularly relevant to test alongside our discounting tasks as it is specifically designed to target intertemporal economic behavior, originally to explain social mobility and lack thereof (Ray & Najman, 1986). Participants select between “strongly disagree” (0) to “strongly agree” (5) on 12 items. Six items are ’positive’ (such as, “I am good at saving my money”) and six reversed (for example, “I agree with the philosophy ’eat, drink and be merry, for tomorrow we may all be dead.”’).

Our behavioral tasks to measure temporal discounting include an internally developed effort task and two monetary tasks, Convex Time Budgets (Andreoni & Sprenger, 2012) and Multiple Price Lists (Andersen, Harrison, Lau, & Rutström, 2008). We focus here on the effort task in the interest of brevity. We chose this task for three reasons. First, a common criticism of monetary discounting tasks is that money is fungible and, therefore, these tasks may not capture time preferences over consumption in the presence of functioning credit markets (Augenblick, Niederle, & Sprenger, 2015). An effort task addresses this concern. Second, monetary tasks involve payments, while an effort task involves small behavioral costs; this makes an effort task a better model of adherence behavior, which also involves small behavioral costs (e.g. seeing a doctor). Third, using monetary incentives to elicit time preferences may be inappropriate in developing countries since individuals might discount money for situational reasons other than preferences; for example, a participant might demonstrate low discount rates in a monetary discounting task when they have enough food on the table, but show different preferences when food is scarce because they need the money immediately. Indeed, Carvalho et al. (2016) show that low-income participants are present-biased, giving stronger weight to payoffs that are closer to the present rather than the future when making choices in a monetary discounting task before payday, but do not show present bias in non-monetary real effort tasks under these circumstances. For these reasons, we implement *a task of choices over time-dated effort*. Participants had to choose between an earlier and later amount of effort, in the form of a specific number of phone calls to the Busara Center at particular hours in the evening. The participants could choose to make two phone calls (or related acceptable contact, including SMS or a “please call me” message) on the earlier date, or a number between 1 and 6 calls at the later date, depending on the decision. Respondents were told they would be paid a fixed KES 500 one month after the session, conditional on completion of the task. An example of a participant's screen is provided in Fig. A.4.

We model intertemporal choices using a quasi-hyperbolic discounting function following Laibson (1997), which allows for time-inconsistent preferences. From participants’ choices, we estimate two discounting parameters: First, the ”present-bias” parameter, β, which attaches special weight to immediate outcomes. $\beta = 1$ implies no present bias, whereas $\beta < 1$ implies present bias. The second parameter, δ, describes the rate at which future outcomes are devalued exponentially. For example, $\delta = 1$ means that a participant behaves as if they are indifferent between making 2 calls tomorrow and 2 the day after tomorrow, whereas $\delta = 0.5$ implies that the participant would be indifferent between 1 call tomorrow and 2 the day after tomorrow.

2.2.6. Demographic measures and additional data

Our demographic questionnaire asks participants for the following information: age, gender, employment status, weekly earnings, financial dependency on someone else, daily consumption, household composition, marital status, and perceived social standing in the community relative to others (Adler & Stewart, 2007).

3. Results

Table 1 summarizes the psychometric criteria for test-retest reliability, construct validity, internal consistency/reliability, and acceptability against which we tested each of the adapted psychological scales and behavioral tasks.

3.1. Internal consistency: Cronbach's alpha

The degree to which items are interrelated and measure a single underlying construct in each subscale, or in a total score, is assessed using Cronbach's alpha (Cronbach, 1951). Following Streiner (2003), we consider $\alpha \geq 0.70$ acceptable internal consistency, and $\alpha \geq 0.90$ to indicate item redundancy. Cronbach's α is listed for each scale in Column (1) of Table 2. We find high values of 0.94 and 0.80 for BRIEF and GSE, respectively, and moderate values between 0.45 and 0.55 for DGS, CFC, and PMS. Thus, our questionnaire measures of executive control and generalized self-efficacy have satisfactory internal consistency, while that for mastery and those for temporal discounting have somewhat weaker consistency.

3.2. Test-retest reliability

We assess test-retest reliability uniformly across tasks and scales. For every scale score and parameter estimate, we calculate Lin's Concordance Correlation Coefficient (Lawrence & Lin, 1989). We consider acceptable reliability of a scale or parameter as $\rho_c > 0.70$. Column (2) of Table 2 lists Lin's Concordance Correlation Coefficient of intertemporal, or test-retest, reliability for the subsample of 93 participants who completed identical scales and tasks one week later. None of the scales has $\rho_c > 0.70$, our pre-specified criterion. However, some of the scales get close, with BRIEF and GSE both at 0.62. No scale has a $\rho_c < 0.40$. Note that these results mirror those for internal consistency in that BRIEF and GSE appear to exhibit better psychometric properties than other scales adapted to this context.

Table 1
Validation strategies.

Instrument	Measure	Test-retest Reliability	Construct Validity	Internal Consistency/Reliability	Acceptability
BRIEF	$BRIEF_i = \sum_{j=1}^{34} x_{ij}$ where x_{ij} is the numeric response of subject i to item j	$\rho_c \geq 0.70$	(i) Convergent with Stroop and Tower of London. Discriminant otherwise. (ii) Three factors (Roth et al., 2013). (iii) $RMSEA \leq 0.05$; $CFI > 0.95$; $TLI \geq 0.90$	$0.70 \leq \alpha \leq 0.90$	MEF ≤ 0.80.
GSE	$GSE_i = \sum_{j=1}^{14} x_{ij}$	$\rho_c \geq 0.70$	(i) Convergent with *PMS* and *SE*. Discriminant otherwise. (ii) Single factor structure (Scherbaum, Cohen-Charash, & Kern, 2006). (iii) $RMSEA \leq 0.05$; $CFI > 0.95$; $TLI \geq 0.90$	$0.70 \leq \alpha \leq 0.90$	MEF ≤ 0.80.
PMS	$PMS_i = \sum_{j=1}^{7} x_{ij}$	$\rho_c \geq 0.70$	(i) Convergent with *GSE* and *SE*. Discriminant otherwise. (ii) Single factor structure (Marshall & Lang, 1990b). (iii) $RMSEA \leq 0.05$; $CFI > 0.95$; $TLI \geq 0.90$	$0.70 \leq \alpha \leq 0.90$	MEF ≤ 0.80.
DGS	$DGS_i = \sum_{j=1}^{12} x_{ij}$	$\rho_c \geq 0.70$	(i) Convergent with *CFC* and all time preference measures. Discriminant otherwise. (ii) Single factor structure. (iii) $RMSEA \leq 0.05$; $CFI > 0.95$; $TLI \geq 0.90$	$0.70 \leq \alpha \leq 0.90$	MEF ≤ 0.80.
CFC	$CFC_i = \sum_{j=1}^{12} x_{ij}$	$\rho_c \geq 0.70$	(i) Convergent with *DGS* and all time preference measures. Discriminant otherwise. (ii) Single factor structure (Strathman et al., 1994). (iii) $RMSEA \leq 0.05$; $CFI > 0.95$; $TLI \geq 0.90$	$0.70 \leq \alpha \leq 0.90$	MEF ≤ 0.80.
Stroop Task	Correct arrows per minute.	$\rho_c \geq 0.70$	Convergent with BRIEF and Tower of London. Discriminant otherwise.	≥ 0.90 of participants should correctly answer ≥ 0.90 of comprehension questions.	(i) Qualitative variation in the total score; (ii) Presence of Stroop effect.
Tower of London	Total moves.	$\rho_c \geq 0.70$	Convergent with BRIEF and Stroop. Discriminant otherwise.	Noncompletion of a round should not be related to education.	MEF ≤ 0.80
Self-Efficacy Task	Goal set for round 2 y_a ; round 1 performance y; self-efficacy measure $SX = \frac{y_a}{1+\lvert y_a - y\rvert}$	$\rho_c \geq 0.70$	Convergent with GSE and PMS. Discriminant otherwise (especially risk).	≥ 0.90 of participants should correctly answer ≥ 0.90 of comprehension questions.	Proportion significantly exceeding their goal should be low (qualitative).
Time Preferences	β_i^{Effort}, δ_i^{Effort}		Parameters convergent with DGS and CFC. Discriminant otherwise.	Question consistency checks.	Parameters in line with previous Busara study (Balakrishnan, Haushofer, & Jakiela, 2015).

Table 2
Metrics of scale reliability and validity.

	α	ρ_c	MEF	CFI	TLI	RMSEA	AIC
BRIEF	0.94	0.62	0.74	0.85	0.84	0.07	51932.00
CFC	0.48	0.44	0.51	0.35	0.22	0.10	22125.97
DGS	0.49	0.46	0.49	0.44	0.30	0.09	18708.28
GSE	0.80	0.62	0.52	0.88	0.85	0.08	20151.51
PMS	0.52	0.54	0.45	0.69	0.54	0.12	12446.19

Notes: Metrics of scale validity and reliability: Cronbach's alpha (column 1), Lin's concordance coefficient measuring test-retest reliability (column 2), maximum endorsement frequency measuring acceptability (column 3), comparative fit index (column 4), Tucker-Lewis index (column 5), root mean square error approximation (column 6), and Akaike information criterion from confirmatory factor analysis (column 7) of the five psychological scales for executive function (BRIEF), temporal discounting (DGS, CFC), and self-efficacy (GSE, PMS).

3.3. Acceptability: Maximum Endorsement Frequencies

Where applicable, we calculate Maximum Endorsement Frequencies (MEF) as the percentages of respondents selecting a particular option in items with a discrete number of options (Bowling, 2014, p.117). In our Likert-type scales, for instance, these are equal to the proportion of participants choosing the most common answer. For the Tower of London task, MEF is the proportion of participants who complete a given trial in a specific number of moves. We reject any item which has a MEF exceeding 80% of the sample (Bowling, 2014). Floor and ceiling effects are accounted for by this metric as special cases in which the lowest and highest response possibilities, respectively, violate the MEF criterion. As shown in Column (3) of Table 2, we find that this is not the case for any of our scales. This is also true for the Tower of London task, which has MEF = 0.79. Thus, we find adequate acceptability for each of the scales according to MEF.

3.4. Confirmatory factor analysis

We use confirmatory factor analysis to compare the empirical factor structure of each scale to the intended factor structure. For each scale, we test a model that has a latent factor representing each subscale (or the scale itself if the original scale is not partitioned). We assume that each item loads on only its respective subscale and none of the other latent factors. Latent factors are allowed to be correlated with each other, and we assume that there is no error covariance among the items.

We test models for each scale in accordance with models found in the original validation papers for these scales, as displayed in Table 1. Specifically, for BRIEF-A, we test a model for each of the five subscales within the metacognition factor, from which we sample the 32 questions. For all other scales, we implement a single factor model. We then examine model fit using commonly applied global fit criteria. Acceptable fit of the model is indicated by a Comparative Fit Index (CFI) around 0.95, and a root mean square error approximation (RMSEA) < 0.08, with RMSEA ≤ 0.05 indicating a strong fit (Hu & Bentler, 1999). In addition, we report the Tucker–Lewis Index (TLI), with values near 1.0 indicating good fit and values > 0.90 considered acceptable, and the Akaike Information Criterion (AIC), with higher values preferable. (Akaike, 1987; Cheung & Rensvold, 2002; Dunn, Everitt, & Pickles, 1993; Marsh, Balla, & Hau, 1996; Tucker & Lewis, 1973).

The results from the confirmatory factor analysis of our psychological scales are shown in Columns (4)–(7) of Table 2, which report the CFI, TLI, RMSEA, and AIC, respectively. Recall from above that acceptable fit of the model is indicated by $RMSEA \leq 0.05$, $CFI > 0.95$, and $TLI \geq 0.90$ (Hu & Bentler, 1999). We observe reasonable fit for our executive function and self-efficacy scales, but poor fit for the two discounting scales. BRIEF has the lowest RMSEA at 0.07, as well as highest AIC, with CFI and TLI close to 0.85. GSE has almost identical relative fit indices, but lower AIC. PMS shows somewhat worse fit, with CFI of 0.69, TLI of 0.54, and a relatively high RMSEA of 0.12. For both CFC and DGS, the CFI and TLI are below 0.50. However, RMSEA for these scales is still reasonable at around 0.10, and the AIC is higher than for the self-efficacy scales. No scale satisfies our pre-defined acceptability criteria, but we obtain reasonable evidence for the validity of the factor structure reported in the literature on Western populations of BRIEF, GSE, and, to a lesser extent, PMS. In conjunction with the results for reliability and consistency, we interpret the evidence to suggest BRIEF and GSE are well-designed for use among our target population, and other scales less so.

Table 3
Factor loadings and communality of all measures.

	Factor 1	Factor 2	Factor 3	Factor 4	Communality
Stroop	0.48	−0.21	0.05	−0.45	0.48
TOL	−0.38	−0.08	−0.41	0.53	0.59
β^{Effort}	−0.18	0.72	−0.07	−0.04	0.56
δ^{Effort}	−0.08	0.64	0.30	−0.24	0.57
SE task	0.23	−0.20	0.74	0.19	0.68
BRIEF	−0.66	−0.21	0.24	−0.03	0.53
CFC	0.56	0.05	0.17	0.49	0.58
DGS	0.54	0.31	0.11	0.40	0.56
GSE	0.63	0.04	−0.41	−0.05	0.57
PMS	−0.68	0.08	0.11	0.19	0.52

Notes: Factor loadings and communality scores for each task and scale. Loadings are after oblique rotation, allowing for non-orthognality between factors. The number of factors displayed is equal to the number of hypothesized constructs measured.

3.5. Exploratory factor analysis

In addition to the confirmatory factor analysis, Table 3 reports the results of an exploratory factor analysis, where each of the individual measures are treated as items of an overall meta-instrument. Columns (2)–(5) list the loadings of each of the measures onto the first four common factors. These factors are rotated obliquely, post-estimation, to maximize explanatory power of the factors and allow for non-orthogonality between factors. We note that the two executive function tasks, Stroop and Tower of London, as well as all the scales, load heavily on the first factor. The behavioral measures of temporal discounting, β^{Effort} and δ^{Effort}, and the Deferment of Gratification Scale, load together onto the second factor, but the Consideration of Future Consequences scale does not. The third factor is characterized by loadings for two measures each of executive function (TOL and BRIEF) and self-efficacy (Slider Task and GSE), implying that these two constructs covary in the population and are related to another latent characteristic. Finally, the fourth factor exhibits strong explanatory power for TOL and the two temporal discounting scales, for which there is no straightforward interpretation.

Column (6) reports the communality of each instrument, or the proportion of total variation which can be accounted for by the four common factors. We find communality scores to be clustered in a narrow band between the Stroop task (0.48) and the Slider task (0.68). The model explains approximately half of the variation for each of the constructs measured with moderate and similar communality scores, implying that no instrument is orthogonal to others.

3.6. Construct validity: pairwise correlations

Construct validity for both scales and tasks is assessed using convergent and discriminant validity. We hypothesize that tasks (in terms of overall performance, defined separately for each task) and scales will correlate with other tasks and scales within the same construct (convergent validity), but not with scales and tasks corresponding to other constructs (discriminant validity). To test this, we construct a

Table 4
Construct validity: Correlations.

Type	Construct		Stroop	TOL	β^{Effort}	δ^{Effort}	SE task	BRIEF	CFC	DGS	GSE	PMS
Tasks	Exec. Funct.	Stroop	1.00									
		TOL	−0.20***	1.00								
	Discounting	β^{Effort}	−0.06	0.06	1.00							
		δ^{Effort}	0.01	−0.07	0.13**	1.00						
	Self-Efficacy	SE task	0.13***	−0.12***	−0.09	0.00	1.00					
Scales	Exec. Funct.	BRIEF	−0.15***	0.13***	0.03	0.02	−0.10***	1.00				
	Discounting	CFC	0.16***	−0.05	−0.03	−0.01	0.12***	−0.15***	1.00			
		DGS	0.07	−0.06	0.01	0.00	0.05	−0.28***	0.20***	1.00		
	Self-Efficacy	GSE	0.18***	−0.03	−0.02	−0.05	0.07	−0.34***	0.19***	0.14***	1.00	
		PMS	−0.23***	0.15***	0.10*	−0.03	−0.13***	0.27***	−0.19***	−0.14***	−0.35***	1.00

Notes: Pairwise Pearson correlations between psychological scales and behavioral tasks. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

correlation matrix of all measurement instruments. We expect $|\rho| \geq 0.3$ for instruments within the same construct and $|\rho| < 0.3$ for instruments relating to different constructs.

Table 4 shows the pairwise Pearson correlations between our measures. For executive control, we find that BRIEF-A weakly correlates with both executive control tasks (i.e. TOL and Stroop) but also correlates with some of the other measures. For temporal discounting, CFC weakly correlates with DGS, as expected, but not significantly with any of the behavioral discounting measures. The same is true for DGS, which does not correlate significantly with any of the behavioral measures. For self-efficacy, GSE and PMS are moderately correlated, and PMS also weakly correlates with the self-efficacy task, but GSE does not reach significance. Across constructs, the self-efficacy (SE) task weakly correlates with the TOL and Stroop tasks, suggesting it may be tapping aspects of executive control in addition to self-efficacy. In turn, the Stroop and TOL tasks, both tapping distinct aspects of executive control, weakly correlate with each other as expected. The two discounting parameters do not correlate significantly with other scales or tasks, but they do correlate with each other.

Overall, contrary to what we expected, we find weak correlations between tasks and scales implemented to measure executive control and no correlations between tasks and scales implemented to measure temporal discounting and self-efficacy, respectively. These results suggest that each scale and behavioral task tested in this context might be tapping a distinct process within the domain of the target construct.

4. Discussion

The purpose of this paper was to adapt and test the psychometric properties of a battery of Western psychological scales and behavioral tasks, measuring temporal discounting, self-efficacy, and executive control in Kenya. We report two main findings: First, we find that some scales show better psychometric properties than others; specifically, the BRIEF-A, which we used to measure executive control, and the GSE, which we used to measure self-efficacy, were the only two scales with acceptable values of Cronbach's alpha and test-retest reliability. Indeed, many of the statistics do not reach the levels conventionally considered desirable, highlighting the difficulty of translating constructs, scales, and tasks across cultures. Second, we find low correlations between behavioral tasks and self-report scales hypothesized to measure the same construct, suggesting that these specific assays do not measure the construct, or may be tapping distinct processes within the domain of the target construct.

We speculate that BRIEF-A (which targets executive control) and GSE (which targets self-efficacy) may adapt well because they tap universal constructs that vary little across cultures. Indeed, our findings complement prior work, which has successfully validated these assays across cultures (Scholz, Doña, Sud, & Schwarzer, 2002). In contrast, the measures tested here that do not adapt well may be measuring distinct processes within the domain of the target construct that vary across cultures, and thus may require further adaptation. For example, PMS may not exactly measure self-efficacy, but rather perceived control, which has been shown to carry varying connotations dependent on the culture in which it is measured (Cheng, Cheung, Chio, & Chan, 2013). Similarly, CFC (which targets future orientation) and DGS (which targets impulsivity) tap constructs that offer two distinct explanations for discounting behavior that may be context-dependent (Becker & Mulligan, 1997; Carvalho et al., 2016; Laajaj, 2012), rather than measuring time preferences. Further, although we conducted cognitive interviewing with the target population to reduce the risk of response bias, it is also possible that such biases affected responses to the self-report questionnaires.

One of the central goals of the SOBC network is to ascertain whether specific assays within a target domain are tapping overlapping or distinct processes. The current study tests the hypothesis that psychological scales strongly correlate with behavioral tasks theorized to tap the same construct, but finds poor correlations between self-report and behavioral measures adapted to the Kenyan context. Other studies have also found that self-report assessments of discounting do not correlate with behavioral tasks. For example, in a meta-analysis on self-report and behavioral measures of self-control, Duckworth and Kern (2011) conclude that "self-control is a coherent but multidimensional construct best assessed using multiple methods." However, given the relatively small number of measures used, poor psychometric properties of most of the scales, and the possibility that they tapped distinct processes, it is difficult to draw strong conclusions from this result. Indeed, a limitation of the current study is that the scales and tasks chosen for adaption may have measured the intended target constructs, but separate components. Specifically, CFC and DGS assess trait-like qualities related to temporal discounting rather than true preferences measured with the discounting task. BRIEF-A assesses executive control across four different subdomains (i.e. initiate, working memory, plan/organize, and task monitor). TOL and Stroop assess planning and attentional control, respectively. Finally, our behavioral measure of self-efficacy may be assessing processes other than self-efficacy, measured by the GSE, and mastery by the PMS. Future studies might attempt to develop self-report and behavioral measures that tap identical processes and assess culturally valid manifestations of these processes.

The present study raises several questions for future research. First, though we discuss temporal discounting, self-efficacy, and executive control as potential mechanisms of change mediating the relationship between stress, chlorination, and ANC/PNC adherence in Kenya, we did not study relationships of the tasks and questionnaires with these outcome variables in the current study. Using the adapted measures with acceptable psychometric properties, we are currently in the process of examining the effects of stress on our three targets in a laboratory

study, which induces stress using four different methodologies: hydrocortisone administration (Riis-Vestergaard et al., in press), the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993), the cold pressor task (Hines & Brown, 1936), and centipede game (Haushofer, Jang, & Lynham, 2015). We are also developing and testing several interventions hypothesized to engage our three targets and intend to study how target engagement correlates with health behaviors, such as chlorination of drinking water and ANC/PNC adherence. Second, given the reasonable but not particularly strong psychometric properties of the psychological scales, future work might attempt to develop new scales for these three concepts that are specifically geared to low-income populations in developing countries.

Acknowledgements

This work was supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institute of Nursing Research (UH2NR016378). For more information on this study's role in the Science of Behavior Change Program, please visit our Open Science Framework page: https://osf.io/twbu8/. Oxford research assistance was supported by an anonymous donor under grant LER00130, to KO. We are grateful to the staff of the Busara Center for Behavioral Economics for excellent research assistance and data collection. We are also grateful to our NIH Project Scientist, Dr. Rosalind King, for her scientific oversight. Contributions: KO and JH directed the study; KE, DM, PJ, JA, CJ, and MRV designed tasks and data collection instruments; KE, PJ, CJ, and DM collected data; KE, HU, CT, and KO developed protocols for and conducted cognitive interviews; HU, CT, DM, GN, JA, KO, and JH developed analysis methods and wrote code; DM and JH ran analysis; KE, DM, KO, and JH wrote the paper.

Appendix

Fig. A.1. Stroop Task Interface.

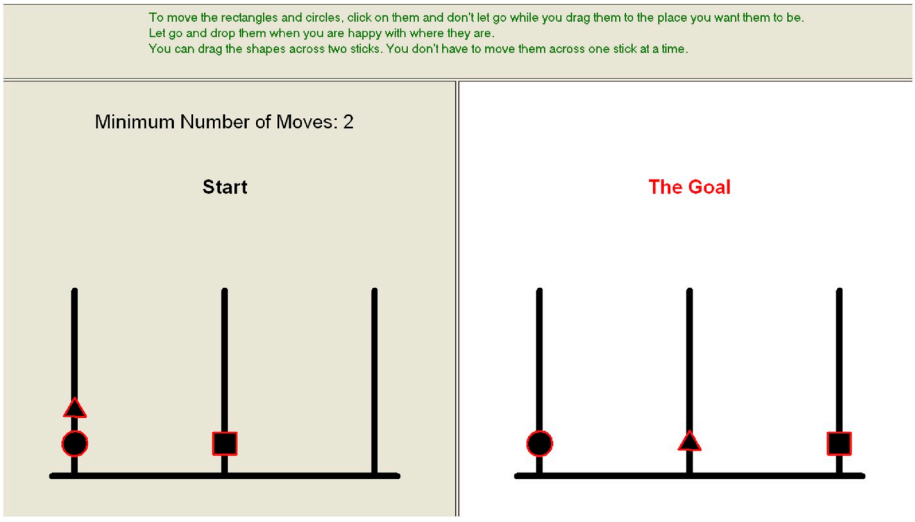

Fig. A.2. Tower of London Task Interface.

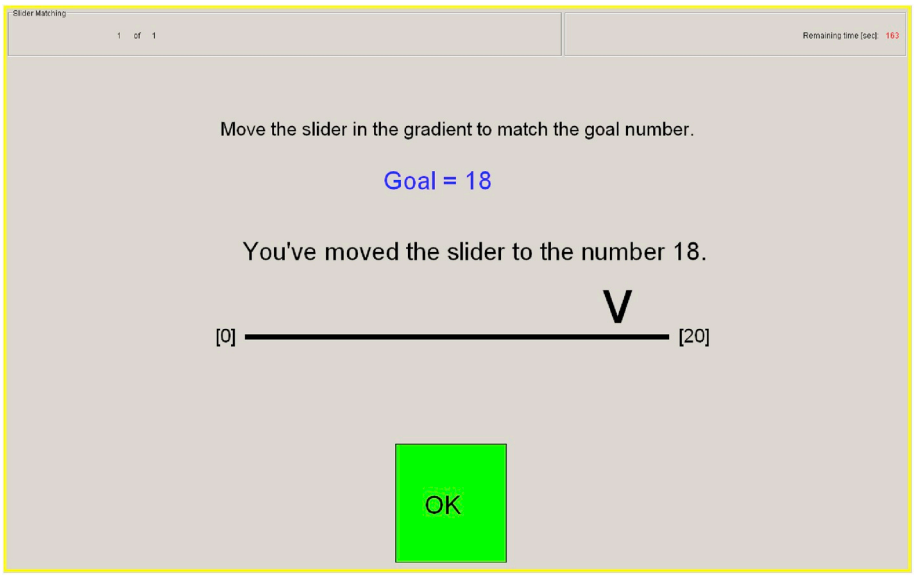

Fig. A.3. Self-Efficacy Task Interface.

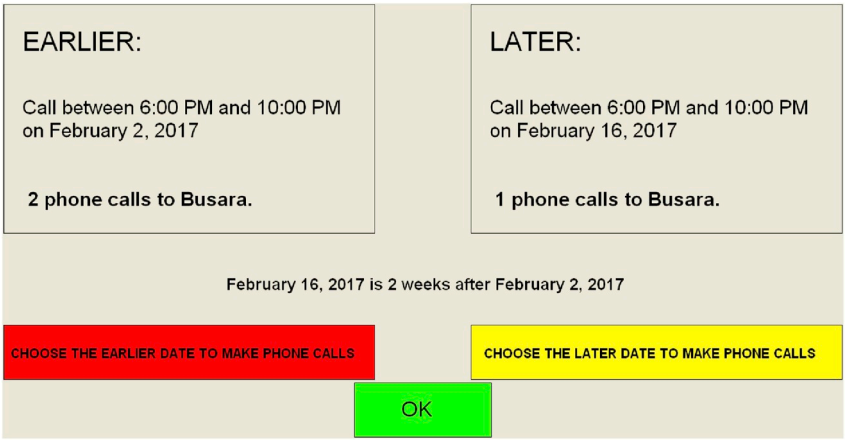

Fig. A.4. Effort Discounting Task Interface.

Table A.1
Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A) Scale Items.

Scale	1 = "Never a problem"...7 = "Always a problem"
1	I do not have problems completing my work
2	I make careless mistakes when completing work
3	I have trouble being attentive while working (such as household chores, reading or work).
4	I need to be reminded to start a task even when it's my own will.
5	I get overwhelmed by large tasks.
6	I have trouble with jobs or tasks that have more than one step.
7	I have trouble getting ready for the day.
8	When I have many important things to do, I have trouble deciding which activity to start first.
9	I forget what I am doing in the middle of things/activities.
10	I don't inspect my work for mistakes.
11	I lay around in the house a lot.
12	I start work (such as cooking, projects) without the right tools.
13	I fail to judge how difficult or easy work will be.
14	I have trouble starting anything on my own.
15	I have trouble staying on the same topic when talking.
16	I don't plan early for future activities.
17	I concentrate for a short time.
18	I have goals that are unachieveable.
19	I make mistakes carelessly.
20	I have difficulty being excited about things.

21	I forget instructions easily.
22	I have good ideas but cannot put my ideas into action.
23	I have trouble getting started on tasks.
24	I have trouble finishing tasks (such as chores, work).
25	I start things at the last minute (such as assignments, chores, tasks).
26	I have difficulty finishing a task on my own.
27	I have trouble remembering things, even for a few minutes (such as directions, phone numbers).
28	I have trouble coming up with ideas for what to do with my free time.
29	I don't plan early for tasks.
30	I have problems organizing activities.
31	I have trouble doing more than one thing at a time.
32	I have trouble organizing work.
34	I do not make careless mistakes when completing my work.

Table A.2
Consideration of Future Consequences (CFC) Scale Items.

Scale	1 = ”Not at all like me”...7 = ”Very much like me”
1	I think about how things would be in days to come, and try to influence those things in my daily behavior.
2	I often involve myself in a specific behavior to achieve results that may not come until many years later.
3	I only act to satisfy immediate needs, thinking the future will take care of itself.
4	My behavior is influenced by the immediate outcomes of my actions (i.e. within a few days or weeks).
5	When I take action or make decisions, I am more likely to choose an option that involves little trouble or effort.
6	I am ready to sacrifice my current happiness or wellbeing in order to achieve future results.
7	I think it's important to take warnings on bad outcomes as a very weighted issue even if the bad outcome may not happen until many years later.
8	I think it is important to work on something with important future consequences than on something with less important immediate consequences.
9	In general, I ignore warnings about problems that can possibly happen later because I think those problems will be solved before reaching a crisis level.
10	I think that sacrificing now is not a must because later outcomes can be dealt with at a later time.
11	I only act to satisfy immediate concerns, figuring that I will take care of the future problems that may occur at later date.
12	Since my everyday work has specific outcomes, it is more important to me than behavior that has distant outcomes.
13	When I make a decision, I think about how it might affect me in the future.
14	My behavior is generally influenced by future consequences.

Table A.3
Deferment of Gratification (DGS) Scale Items.

Scale	1 = ”Strongly disagree”...7 = ”Strongly agree”
1	I am good in saving my money instead of spending it at once.
2	I enjoy something more when I have to wait for it and plan for it.
3	When I was a child, I saved any pocket money that I had.
4	When I am in the market, I usually buy a lot of things that I had not planned to buy.
5	I am constantly without money.
6	I agree with the philosophy: ”Eat, drink, and be happy, for tomorrow we may all be dead”.
7	I would describe myself as often acting without thinking for my own good.
8	I often think it is important to wait and think things over before deciding.
9	I like spending my money immediately after I get it.
10	It is hard for me to avoid losing my temper when someone gets me very angry.
11	Most of the time, it is easy for me to be patient when I am kept waiting for things.
12	I am good at planning things ahead.

Table A.4
General Self-Effiacy (GSE) Scale Items.

Scale	1 = ”Strongly disagree”...7 = ”Strongly agree”
1	I cannot usually handle whatever comes my way.
2	I can always solve difficult problems if I try hard enough.
3	If someone is against me, I can find means and ways to get what I want.
4	It is easy for me to stick to my aims and accomplish my goals.
5	I am confident that I could deal appropriately with unexpected events.
6	Thanks to my skillful and creative thinking, I know how to handle unforeseen situations.
7	I solve most problems if I put in the necessary effort.
8	I can remain calm when I am facing difficulties because I can rely on my abilities to cope.
9	When I am faced/confronted with a problem, I can usually find several solutions.
10	If I am in trouble, I can usually think of a solution.
11	I can usually handle whatever comes my way.
12	I cannot always solve difficult problems if I try hard enough.

Table A.5
Pearlin Mastery Scale (PMS) Scale Items.

Scale	1 = "Strongly disagree"...7 = "Strongly agree"
1	There's no way I can solve some of problems I have.
2	Sometimes I feel that I am being pushed here and there in life.
3	I have little control over things that happen to me.
4	I can do anything when I put my mind to it.
5	Most of the time, I feel helpless when dealing with problems of life.
6	What happens to me in the future mostly depends on me.
7	There's little I can do to change most of the important things in my life.

Table A.6
Quality of Psychometric Properties in Considered Measures.

	Instrument	Est. #Citations	Reliability			Validity		
	Instrument	Est. # Citations	Internal Consistency	Test-Retest Reliability	Inter-rater Reliability	Construct Validity	Criterion Validity	Cross cultural validity
TEMPORAL DISCOUNTING								
Inventories	Possible Selves Inventory	5587	NA	+ +	+ + +	NA	+ +	+ +
	Zimbardo Time Perspective Inventory	1635	+ +	+ +	NA	+ + +	+ +	+ +
	Consideration of Future Consequences Scale	849	+ +	+ +	NA	+ +	+ +	NA
	Barratt Impulsiveness Scale	551	+ +	+ +	NA	+ +	+ +	+ + +
	Future Time Perspective Inventory	269	NA	NA	NA	±	NA	NA
	Deferment of Gratification Scale	70	+ +	NA	NA	+ +	NA	NA
	Delaying Gratification Inventory	34	+ + +	+ +	NA	+ +	+	NA
	Future Outlook Inventory	19	+	+	NA	NA	+	NA
Tasks	Marshmallow Test	2199	NA	NA	NA	+ + +	+ + +	+
	McClure Juice Task	616	NA	NA	NA	+ +	+ +	NA
	Hot/Cool Go/No-go Task	275	NA	NA	NA	+ +	+ +	NA
	Convex Time Budget	228	NA	NA	NA	+ +	+ +	+
	Disgusting Drink Task	166	NA	NA	NA	+ +	+ +	NA
	Effort (Job) Allocation Task	89	NA	NA	NA	+ +	+ +	+
	Real Effort Slider Task	41	NA	NA	NA	+	+	NA
	Delay of Gratification Test for Adults	22	–	NA	NA	+	+	NA
	Dynamic Inconsistency in Food Choice	9	NA	NA	NA	+	+	NA

	5-Trial Adjusting Delay Discounting Task	9	NA	NA	NA	+	+	NA
	Bucket Task	6	NA	NA	NA	±	NA	NA
SELF-EFFICACY								
Inventories	Rotter I-E Scale	21221	+ +	- -	NA	+ +	+ +	+ +
	COPE Inventory	8176	+ +	±	NA	+ +	+	NA
	Mastery Scale	7238	+ +		NA	+	+	+ +
	Generalized Self-Efficacy Scale	3029	+ +	+ /	NA	+ +	+ +	+ + +
	Sherer General Self-Efficacy Scale	2871	+ +	±	NA	+ +	+	+
	New General Self-Efficacy Scale	1352	+ +	±	NA	+ +	+	+
	Spheres of Control Scale	181	+ +	+ + +	NA	+ +	+ +	NA
	Coping Strategies Inventory	142	+ +	–	NA	+	+	NA
	Self-Control and Self-Management Scale	18	+	+	NA	+	NA	NA
Tasks	Learned Helplessness Task	647	NA	NA	NA	+	+ +	NA
EXECUTIVE CONTROL								
Inventories	Problem-Solving Inventory	971	+ + +	+ +	NA	+ +	+	NA
	Behavior Rating Inventory of Executive Function for Adults	276	+ + +	+ + +	+	+ +	+ +	NA
	Coolidge Axis II Inventory	117	+	+ +	NA	+	NA	NA
Tasks	Stroop Task	12113	NA	+ +	NA	+ + +	+ +	+ +
	Tower of London	3214	NA	NA	NA	+ +	+ +	NA
	Stop Signal Task	1533	NA	+	NA	+	+	NA
	Dots/Flowers Task	870	NA	NA	NA	+	+	NA
	Raven's Matrices	238	+ +	+ + +	NA	+ +	+ +	+ +
	Parametric Go/No-go Task	165	NA	+	NA	+ +	+	NA
	Continuous Performance Task	64	NA	±	NA	+	+	NA

Notes: + + + or — indicates strong evidence positive/negative result; + + or - indicates moderate evidence positive/negative result; + or - indicates limited evidence positive/negative result; +/- indicates conflicting evidence; NA indicates irrelevance to the measure or no evidence found.

References

Adler, N., & Stewart, J. (2007). *The macarthur scale of subjective social status.* http://www.macses.ucsf.edu/research/psychosocial/subjective.php.

Ahuja, A., Kremer, M., & Zwane, A. P. (2010). Providing safe water: Evidence from randomized evaluations. *Annu. Rev. Resour. Econ. 2*(1), 237–256.

Akaike, H. (1987). Factor analysis and aic. *Psychometrika, 52*(3), 317–332.

Alexander, J. K., Hillier, A., Smith, R. M., Tivarus, M. E., & Beversdorf, D. Q. (2007). Beta-adrenergic modulation of cognitive flexibility during stress. *Journal of Cognitive Neuroscience, 19*(3), 468–478.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583–618.

Andrade, L. F., & Petry, N. M. (2012). Delay and probability discounting in pathological gamblers with and without a history of substance use problems. *Psychopharmacology, 219*(2), 491–499.

Andreoni, J., & Sprenger, C. (2012). Estimating time preferences from convex budgets. *The American Economic Review, 102*(7), 3333–3356.

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

Azizan, A., Justine, M., & Kuan, C. S. (2013). Effects of a behavioral program on exercise adherence and exercise self-efficacy in community-dwelling older persons. *Current Gerontology and Geriatrics Research, 2013*, 282315.

Baddeley, A., Della Sala, S., Robbins, T., & Baddeley, A. (1996). Working memory and executive control [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 351*(1346), 1397–1404.

Balakrishnan, U., Haushofer, J., & Jakiela, P. (2015). *How soon is now? evidence of present bias from convex time budget experiments. IZA Discussion Paper*.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122.

Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety Research, 1*(2), 77–98.

Banerjee, A. V. (2001). *Contracting constraints, credit markets and economic development.* 02–17. https://dspace.mit.edu/bitstream/handle/1721.1/63370/contractingconst00bane.pdf?sequence=1.

Banerjee, A. V., Duflo, E., Glennerster, R., & Kothari, D. (2010). Improving immunisation coverage in rural India: Clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ (Clinical Research ed.), 340*, c2220.

Barkley, S. A., & Fahrenwald, N. L. (2013). Evaluation of an intervention to increase self-efficacy for independent exercise in cardiac rehabilitation. *Behavioral Medicine (Washington, D.C.), 39*(4), 104–110.

Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics, 112*(3), 729–758.

Benight, C. C., & Bandura, A. (2004). Social cognitive theory of posttraumatic recovery: The role of perceived self-efficacy. *Behaviour Research and Therapy, 42*(10), 1129–1148.

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction (Abingdon, England), 96*(1), 73–86.

Bickel, W. K., Miller, M. L., Yi, R., Kowal, B. P., Lindquist, D. M., & Pitcock, J. A. (2007). Behavioral and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. *Drug and Alcohol Dependence, 90*(1), S85–S91.

Boker, A., Feetham, H. J., Armstrong, A., Purcell, P., & Jacobe, H. (2012). Do automated text messages increase adherence to acne therapy? Results of a randomized, controlled trial. *Journal of the American Academy of Dermatology, 67*(6), 1136–1142.

Bowling, A. (2014). *Research methods in health: Investigating health and health services.* McGraw-Hill education (UK).

Bradford, W. D. (2010). The association between individual time preferences and health maintenance habits. *Medical Decision Making: An International Journal of the Society for Medical Decision Making, 30*(1), 99–112.

Brock, L. L., Brock, C. D., & Thiedke, C. C. (2011). Executive function and medical non-adherence: A different perspective. *The International Journal of Psychiatry in Medicine, 42*(2), 105–115.

Carducci, B. J. (2009). *The psychology of personality: Viewpoints, research, and applications.* John Wiley & Sons.

Carroli, G., Rooney, C., & Villar, J. (2001). How effective is antenatal care in preventing maternal mortality and serious morbidity? an overview of the evidence. *Paediatric and Perinatal Epidemiology, 15*(s1), 1–42.

Carvalho, L. S., Meier, S., & Wang, S. W. (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *The American Economic Review, 106*(2), 260–284.

Chapman, G. B. (2005). Short-term cost for long-term benefit: Time preference and cancer control. *Health Psychology, 24*(4), S41–S48.

Cheng, C., Cheung, S.-f., Chio, J. H.-m., & Chan, M.-p. S. (2013). *Cultural meaning of perceived control: A meta-analysis of locus of control and psychological symptoms across 18 cultural regions.*

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Davis, C., Patte, K., Curtis, C., & Reid, C. (2010). Immediate pleasures and future consequences. A neuropsychological study of binge eating and obesity. *Appetite, 54*(1), 208–213.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., et al. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science, 325*(5940), 621–625.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268.

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging farmers to use Fertilizer: Theory

and experimental evidence from Kenya. *American Economic Review, 101*(6), 2350–2390.
Dunn, G., Everitt, B. S., & Pickles, A. (1993). *Modelling covariances and latent variables using EQS.* CRC Press.
Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions, 28*(2), 212–232.
Ferraz, M. (1997). Cross cultural adaptation of questionnaires: What is it and when should it be performed? *The Journal of Rheumatology, 24*(11), 2066.
Fields, S. A., Lange, K., Ramos, A., Thamotharan, S., & Rassu, F. (2014). The relationship between stress and delay discounting: A meta-analytic review. *Behavioural Pharmacology, 25*(5 and 6), 434–444.
Fields, S. A., Ramos, A., & Reynolds, B. A. (2015). Delay discounting and health risk behaviors: The potential role of stress. *Current Opinion in Psychology, 5*, 101–105.
Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Frick, P. A., Lavreys, L., Mandaliya, K., & Kreiss, J. K. (2001). Impact of an alarm device on medication compliance in women in mombasa, Kenya. *International Journal of STD & AIDS, 12*(5), 329–333.
Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 385–400.
Ghosal, S., Jana, S., Mani, A., Mitra, S., & Roy, S. (2013). *Sex workers, stigma and self-belief: Evidence from a psychological training program in India. Warwick university working paper*.
Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review, 102*(1), 469–503.
Hartley, L., & Adams, R. (1974). Effect of noise on the stroop test. *Journal of Experimental Psychology, 102*(1), 62.
Haushofer, J., Collins, M., de Giusti, G., Njoroge, J. M., Odero, A., Onyango, C., et al. (2014). *A methodology for laboratory experiments in developing countries: Examples from the busara center. Working Paper*.
Haushofer, J., Cornelisse, S., Seinstra, M., Fehr, E., Joëls, M., & Kalenscher, T. (2013). No effects of psychosocial stress on intertemporal choice. *PloS One, 8*(11), e78597.
Haushofer, J., Jang, C., & Lynham, J. (2015). *Stress and temporal discounting: Do domains matter? Working Paper*.
Haushofer, J., & Orkin, K. (2017). *Validating psychological scales and behavioral tasks for urban poor populations in Kenya. AEA RCT Registry*.
Hines, E. A., & Brown, G. E. (1936). The cold pressor test for measuring the reactibility of the blood pressure: Data concerning 571 normal and hypertensive subjects. *American Heart Journal, 11*(1), 1–9.
Hiroto, D. S. (1974). Locus of control and learned helplessness. *Journal of Experimental Psychology, 102*(2), 187.
Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. *The Quarterly Journal of Experimental Psychology, 22*(1), 28–36.
Holden, S. T., Shiferaw, B., & Wik, M. (1998). Poverty, market imperfections and time preferences: Of relevance for environmental policy? *Environment and Development Economics, 3*(1), 105–130.
Hou, M. Y., Hurwitz, S., Kavanagh, E., Fortin, J., & Goldberg, A. B. (2010). Using daily text-message reminders to improve adherence with oral contraceptives: A randomized controlled trial. *Obstetrics and Gynecology, 116*(3), 633–640.
Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.
Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the royal statistical society. Series a (General,* 42–73.
Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). *Getting to the top of mind: How reminders increase saving. Technical report.* National Bureau of Economic Research.
Kenya National Bureau of Statistics, Ministry of Health, National Aids Control Council, Kenya Medical Research Institute, National Council for Population and Development, The DHS Program, & ICF International (2014). *Kenya demographic and health survey 2014. Technical report.* Kenya National Bureau of Statistics.
Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The trier social stress test–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1–2), 76–81.
Kremer, M., Miguel, E., Mullainathan, S., Null, C., & Zwane, A. P. (2009). Making water safe: Price, persuasion, peers, promoters, or product design. *Unpublished Manuscript*.
Laajaj, R. (2012). *Closing the eyes on a gloomy future: Psychological causes and economic consequences. In Pacific Development Economics Conference.* Davis: University of California.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443–478.
Langebeek, N., Gisolf, E. H., Reiss, P., Vervoort, S. C., Hafsteinsdóttir, T. B., Richter, C., et al. (2014). Predictors and correlates of adherence to combination antiretroviral therapy (ART) for chronic HIV infection: A meta-analysis. *BMC Medicine, 12*(1), 142.
Lawrence, I., & Lin, K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics,* 255–268.
Lester, R. T., Ritvo, P., Mills, E. J., Kariri, A., Karanja, S., Chung, M. H., et al. (2010). Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): A randomised trial. *The Lancet, 376*(9755), 1838–1845.
Liston, C., McEwen, B. S., & Casey, B. (2009). Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proceedings of the National Academy of Sciences, 106*(3), 912–917.
Luethi, M., Meier, B., & Sandi, C. (2008). Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioral Neuroscience, 2*.
Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The general self-efficacy scale: Multicultural validation studies. *The Journal of Psychology, 139*(5), 439–457.
MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: A meta-analysis. *Psychopharmacology, 216*(3), 305–321.
Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341*(6149), 976–980.
Marsh, H. W., Balla, J. R., & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. *Advanced Structural Equation Modeling: Issues and Techniques,* 315–353.
Marshall, G. N., & Lang, E. L. (1990a). Optimism, self-mastery, and symptoms of depression in women professionals. *Journal of Personality and Social Psychology, 59*(1), 132.
Marshall, G. N., & Lang, E. L. (1990b). Optimism, self-mastery, and symptoms of depression in women professionals. *Journal of Personality and Social Psychology, 59*(1), 132.
Mausbach, B., Patterson, T., Känel, R. V., Mills, P., Dimsdale, J., Ancoli-Israel, S., et al. (2007). The attenuating effect of personal mastery on the relations between stress and alzheimer caregiver health: A five-year longitudinal analysis. *Aging & Mental Health, 11*(6), 637–644.
McDonagh, M. (1996). Is antenatal care effective in reducing maternal morbidity and mortality? *Health Policy and Planning, 11*(1), 1–15.
Minor, T. R., Jackson, R. L., & Maier, S. F. (1984). Effects of task-irrelevant cues and reinforcement delay on choice-escape learning following inescapable shock: Evidence for a deficit in selective attention. *Journal of Experimental Psychology: Animal Behavior Processes, 10*(4), 543.
Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much.* Macmillan.
Nieuwlaat, R., Wilczynski, N., Navarro, T., Hobson, N., Jeffery, R., Keepanasseril, A., et al. (2014). *Interventions for enhancing medication adherence. The Cochrane database of systematic reviews. Vol. 11*, CD000011.
Pagiola, S. (1996). Price policy and returns to soil conservation in semi-arid Kenya. *Environmental and Resource Economics, 8*(3), 225–271.
Pearlin, L. I., & Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior,* 2–21.
Perkins, S., & Jenkins, L. S. (1998). Self-efficacy expectation, behavior performance, and mood status in early recovery from percutaneous transluminal coronary angioplasty. *Heart & Lung: The Journal of Acute and Critical Care, 27*(1), 37–46.
Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the tower of london task. *The Quarterly Journal of Experimental Psychology Section a, 54*(2), 579–597.
Pop-Eleches, C., Thirumurthy, H., Habyarimana, J. P., Zivin, J. G., Goldstein, M. P., De Walque, D., et al. (2011). Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: A randomized controlled trial of text message reminders. *AIDS (London, England), 25*(6), 825.
Ray, J. J., & Najman, J. M. (1986). The generalizability of deferment of gratification. *The Journal of Social Psychology, 126*(1), 117–119.
Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling. *Behavioural Pharmacology, 17*(8), 651–667.
Riis-Vestergaard, M. I., van Ast, V., Cornelisse, S., Seinstra, M., Joëls, M., and Haushofer, J. The effect of hydrocortisone administration on intertemporal choice. Psychoneuroendocrinology, in press.
Rodrigues, R., Shet, A., Antony, J., Sidney, K., Arumugam, K., Krishnamurthy, S., et al. (2012). Supporting adherence to antiretroviral therapy with mobile phone reminders: Results from a cohort in South India. *PloS One, 7*(8), e40723.
Rogers, R. D., Moeller, F. G., Swann, A. C., & Clark, L. (2010). Recent research on impulsivity in individuals with drug use and mental health disorders: Implications for alcoholism. *Alcoholism, Clinical and Experimental Research, 34*(8), 1319–1333.
Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *BRIEF-A: Behavior rating inventory of executive function–adult version.* Psychological Assessment Resources.
Roth, R. M., Lance, C. E., Isquith, P. K., Fischer, A. S., & Giancola, P. R. (2013). Confirmatory factor analysis of the behavior rating inventory of executive function-adult version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology, 28*(5), 425–434.
Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., et al. (2002). Executive control function: A review of its promise and challenges for clinical research. A report from the committee on research of the american neuropsychiatric association. *The Journal of Neuropsychiatry and Clinical Neurosciences, 14*(4), 377–405.
Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*(6), 1047–1063.
Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242.
Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2010). Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *Journal of Neuroscience, 30*(24), 8190–8196.
Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience, 29*(22), 7191–7198.
Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology, 35*(7), 977–986.
Schwarzer, R., & Jerusalem, M. (2010). The general self-efficacy scale (gse). *Anxiety, Stress and Coping, 12*, 329–345.
Seghers, J., Van Hoecke, A.-S., Schotte, A., Opdenacker, J., & Boen, F. (2014). The added

value of a brief self-efficacy coaching on the effectiveness of a 12-week physical activity program. *Journal of Physical Activity & Health, 11*(1), 18–29.
Shallice, T. (1982). Specific impairments of planning. Philosophical transactions of the royal society of london B. *Biological Sciences, 298*(1089), 199–209.
Shiferaw, B., & Holden, S. T. (2001). Farm-level benefits to investments for mitigating land degradation: Empirical evidence from Ethiopia. *Environment and Development Economics, 6*(3), 335–358.
Shnek, Z. M., Foley, F. W., LaRocca, N. G., Gordon, W. A., DeLuca, J., Schwartzman, H. G., et al. (1997). Helplessness, self-efficacy, cognitive distortions, and depression in multiple sclerosis and spinal cord injury. *Annals of Behavioral Medicine, 19*(3), 287–294.
Shorey, S., Chan, S. W. C., Chong, Y. S., & He, H.-G. (2015). A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. *Journal of Advanced Nursing, 71*(6), 1260–1273.
Simoni, J. M., Huh, D., Frick, P. A., Pearson, C. R., Andrasik, M. P., Dunbar, P. J., et al. (2009). Peer support and pager messaging to promote antiretroviral modifying therapy in seattle: A randomized controlled trial. *Journal of Acquired Immune Deficiency Syndromes, 52*(4), 465–473.
Stilley, C. S., Bender, C. M., Dunbar-Jacob, J., Sereika, S., & Ryan, C. M. (2010). The impact of cognitive function on medication management: Three studies. *Health Psychology, 29*(1), 50.
Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*(4), 742.
Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*(3), 217–222.
Tan-Kristanto, S., & Kiropoulos, L. A. (2015). Resilience, self-efficacy, coping styles and depressive and anxiety symptoms in those newly diagnosed with multiple sclerosis. *Psychology, Health & Medicine, 20*(6), 635–645.
Thirumurthy, H., Hayashi, K., Linnemayr, S., Vreeman, R. C., Levin, I. P., Bangsberg, D. R., et al. (2015). Time preferences predict mortality among hiv-infected adults receiving antiretroviral therapy in Kenya. *PloS One, 10*(12), e0145245.
Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10.
Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica, 113*(1), 45–65.
WHO, W. H. O. (2014). *World health statistics 2014. Technical report, WHO*.
Willis, G. B. (1999). *Cognitive interviewing: A how to guide.* Research Triangle Park, NC: Research Triangle Institute.
Wühr, P. (2007). A stroop effect for spatial orientation. *The Journal of General Psychology, 134*(3), 285–294.
Zolfaghari, M., Mousavifar, S. A., Pedram, S., & Haghani, H. (2012). The impact of nurse shortmessage services and telephone follow-ups on diabetic adherence: Which one is more effective? *Journal of Clinical Nursing, 21*(13–14), 1922–1931.
Zurovac, D., Sudoi, R. K., Akhwale, W. S., Ndiritu, M., Hamer, D. H., Rowe, A. K., et al. (2011). The effect of mobile phone text-message reminders on kenyan health workers' adherence to malaria treatment guidelines: A cluster randomised trial. *The Lancet, 378*(9793), 795–803.