Journal of Research in Personality xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Journal of Research in Personality

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

A meta-analysis of the convergent validity of self-control measures

Angela Lee Duckworth *, Margaret L. Kern

Department of Psychology, University of Pennsylvania, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Self-control
Self-regulation
Impulsivity
Multi-method measurement
Convergent validity
Meta-analysis

ABSTRACT

There is extraordinary diversity in how the construct of self-control is operationalized in research studies. We meta-analytically examined evidence of convergent validity among executive function, delay of gratification, and self- and informant-report questionnaire measures of self-control. Overall, measures demonstrated moderate convergence (r_{random} = .27 [95% CI = .24, .30]; r_{fixed} = .34 [.33, .35], k = 282 samples, N = 33,564 participants), although there was substantial heterogeneity in the observed correlations. Correlations within and across types of self-control measures were strongest for informant-report questionnaires and weakest for executive function tasks. Questionnaires assessing sensation seeking impulses could be distinguished from questionnaires assessing processes of impulse regulation. We conclude that self-control is a coherent but multidimensional construct best assessed using multiple methods.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The construct of self-control has attracted substantial attention from psychologists working within a variety of theoretical and methodological frameworks. At present, more than 3% of all publications are indexed in the PsycInfo database by the keywords self-control, impulsivity, or related terms.[1] However, operational definitions of self-control vary widely, begging the question: do these varied measures tap the same underlying construct? For instance, does the Eysenck Impulsiveness Questionnaire (Eysenck, Easton, & Pearson, 1984) assess the same trait as the preschool delay of gratification task (Mischel, Shoda, & Rodriguez, 1989)? Do these measures tap the same underlying construct as the Stroop (Wallace & Baumeister, 2002) or go/no-go (Eigsti et al., 2006) executive function tasks?

Evidence of convergent validity (i.e., substantial and significant correlations between different instruments designed to assess a common construct) is a "minimal and basic requirement" for the validity of any psychological test (Fiske, 1971, p. 164). Unfortunately, the rather "modest requirement" of convergent validity in psychological measurement is often assumed rather than tested directly (Fiske, 1971, p. 164). In the current investigation, we meta-analytically synthesized evidence from 282 multi-method samples to examine the convergent validity of self-control measures.

1.1. Defining self-control

Several authors have noted the challenge of defining and measuring self-control (also referred to as self-regulation, self-discipline, willpower, effortful control, ego strength, and inhibitory control, among other terms) and its converse, impulsivity or impulsiveness (e.g., DePue & Collins, 1999; Evenden, 1999; White et al., 1994; Whiteside & Lynam, 2001). As an illustration of the diversity of measures that have been used to assess self-control, consider the following: refraining from pushing a button when a non-target stimulus appears on a computer screen, matching two geometric patterns from a selection of highly similar patterns, choosing between $1 today and $2 one week later, refraining from immediately eating a single marshmallow in order to obtain two marshmallows later, and responding to questionnaire items such as "Do you often long for excitement?" or "I make my mind up quickly." Given the rather extraordinary range of measures used, one might expect a lively interdisciplinary debate in the self-control literature as to whether these measures are, in fact, tapping the same underlying construct. Instead, self-control researchers tend to read and cite the work of others working in their same methodological tradition: "Unfortunately, with a few exceptions, researchers interested in the personality trait of impulsivity, in the experimental analysis of impulsive behavior, in psychiatric studies of impulsivity or in the neurobiology of impulsivity form largely independent schools, who rarely cite one another's work, and consequently rarely gain any insight into their own work from the progress made by others" (Evenden, 1999, pp. 348–349).

* Corresponding author. Address: Department of Psychology, University of Pennsylvania, 3701 Market Street, Philadelphia, PA 19104, United States.

E-mail address: duckworth@psych.upenn.edu (A.L. Duckworth).

[1] A PsycInfo search was conducted for all publications from 2009 to 2010 using the following keywords: self-control, self control, self-discipline, self-discipline, self-regulation, self regulation, delay of gratification, delayed gratification, gratification delay, impulsive, impulsivity, impulsiveness, impulse control, emotional regulation, emotion regulation, ADHD, attention deficit, attention deficit disorder, attention deficit hyperactivity disorder, hyperactivity, cognitive control, executive function, and executive functioning.

0092-6566/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2011.02.004

What do diverse measures of self-control have in common? We suggest that the common conceptual thread running through varied operationalizations of self-control is the idea of voluntary self-governance in the service of personally valued goals and standards. This idea is captured concisely by Baumeister, Vohs, and Tice (2007): "Self-control is the capacity for altering one's own responses, especially to bring them into line with standards such as ideals, values, morals, and social expectations, and to support the pursuit of long-term goals" (p. 351). Tasks and questionnaire items that attempt to measure self-control implicitly or explicitly posit a plurality of mutually exclusive responses (e.g., if I eat my cake now, I cannot save it too). One response is recognized by the individual as superior insofar as it is aligned with their long-term goals and standards (saving the cake for after dinner will make me happier in the long-run), but an alternative response (eating the cake now) is more gratifying or automatic in the short-term. In such situations, self-controlled individuals tend to choose the superior response, whereas impulsive individuals tend to choose the immediately gratifying or automatic response.

Our conceptualization of self-control emphasizes "top-down" processes that inhibit or obviate impulses, and thus implicitly assumes "bottom-up" psychological processes that generate these impulses. While individuals surely vary in what they find tempting (Tsukayama, Duckworth, & Kim, submitted for publication), given that adults and children across cultures reliably rate themselves lower in self-control than in any other character strength (Peterson, 2006), it seems reasonable to assume that almost everyone is tempted by *something*. That is, while attraction to various vices may vary across individuals, we agree with Oscar Wilde (1912/2009) that for each of us "there are terrible temptations which it requires strength, strength and courage, to yield to" (Second Act, Line 42).

1.2. Measurement traditions in the study of self-control

Our review of the self-control literature revealed four distinct approaches to the measurement of self-control: *executive function tasks*, *delay of gratification tasks*, *self-report questionnaires*, and *informant-report questionnaires*. Arguably, each of these approaches assesses voluntary self-governance in the service of goals or standards. Still, diversity both within and across these types of measures is striking. Because of their distinct histories, we review the four measurement traditions separately below.

1.2.1. Executive function tasks

Executive function refers to goal-directed, higher-level cognitive processing in which top-down control is exerted over lower-level cognitive processes (Williams & Thayer, 2009). Emerging first in the neuropsychology literature, executive function is a relatively new construct (Burgess, 1997) that, like the construct of self-control, continues to be inconsistently defined and measured (Banfield, Wyland, Macrae, Munte, & Heatherton, 2004; Miller, 2000). Behavioral tasks designed to assess executive function have been used to assess individual differences in self-control (e.g., Eigsti et al., 2006; White et al., 1994), the presence of clinical levels of impulsivity (Baker, Taylor, & Leyva, 1995), the effect of self-control interventions (Diamond, Barnett, Thomas, & Munro, 2007), and experimental manipulations aimed at taxing self-control (Hagger, Wood, Stiff, & Chatzisarantis, 2010).

There is growing evidence that executive function is not unitary in nature, but rather a collection of distinct processes associated with the frontal lobes, including working memory, attention, response inhibition, and task switching (Kramer, Humphrey, Larish, & Logan, 1994; Miller, 2000; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Because any single executive function task tends to assess a plurality of these cognitive processes (Burgess, 1997; Zaparniuk & Taylor, 1997), it was not feasible to organize executive function tasks according to any of the proposed taxonomies of executive function. As an alternative, we noted that authors often explicitly referred to measures as belonging to one of 12 subtypes of executive function task (e.g., sun–moon Stroop, color-word Stroop, counting Stroop) and categorized measures accordingly (see Table 1).

1.2.2. Delay of gratification tasks

Whereas executive function tasks have their roots in the neuropsychology literature and the study of neurological impairment, delay of gratification tasks were first developed to understand normative, age-related changes in child development. The ability to delay the discharge of impulses figured prominently in Freud's (1922) psychoanalytic theory of ego development. Early attempts to operationalize the capacity to delay gratification for the sake of long-term gain included coding images of humans in action from responses to Rorschach ink blots (Singer, 1955). Such projective measures of delay of gratification generally demonstrated poor reliability and validity and were supplanted by more direct measures developed by Walter Mischel in the 1960s. Performance in delay tasks has been shown to predict academic achievement (Evans & Rosenbaum, 2008; Mischel et al., 1989), drug use (Kirby, Petry, & Bickel, 1999), and aggressive and delinquent behavior (Krueger, Caspi, Moffitt, White, & Stouthamer-Loeber, 1996).

Mischel's research included three subtypes of delay tasks (see Table 2). In a *hypothetical choice delay task*, subjects make a series of choices between smaller, immediate rewards and larger, delayed rewards, most or none of which they expect to actually receive. For instance, children answer questionnaire items such as, "I would rather get ten dollars right now than have to wait a whole month and get thirty dollars then" (Mischel, 1961, p. 3). More recently, similar questionnaires have been used to calculate a discount rate for each individual that relates the subjective value of a delayed reward to the delay required to receive it (e.g., Green, Fry, & Myerson, 1994; Kirby et al., 1999). In a *real choice delay task*, subjects make an actual (i.e., not hypothetical) choice between a small, immediate reward and a larger, delayed reward (e.g., Duckworth & Seligman, 2005; Mischel, 1958). This decision happens at a single point in time, after which the decision cannot be revoked. The third subtype, the *sustained delay task,* differs from hypothetical and real choice tasks in that subjects first choose the preferred delayed reward, which is clearly "worth the wait". Subsequently, the ability to delay gratification is measured as the time subjects can resist the smaller, immediate reward in order to obtain the larger, deferred reward (e.g., Mischel et al., 1989; Solnick, Kannenberg, Eckerman, & Waller, 1980).

A fourth subtype of delay task not used by Mischel and colleagues is the *repeated trials delay task*, in which subjects complete a series of brief trials, in each of which they choose between a smaller, more immediate reward and a larger, delayed reward (e.g., Newman, Kosson, & Patterson, 1992). As in choice delay procedures, subjects receive actual rewards (e.g., nickels or candy) and cannot revoke their decision.

1.2.3. Self- and informant-report personality questionnaires

In individual difference and clinical psychology research, self-control is most often measured by pen-and-paper personality questionnaires completed by the participant or a close informant (e.g., parent). Questionnaire measures of self-control have been shown to predict academic achievement (Duckworth, Tsukayama, & May, 2010), physical health (Moffitt et al., 2011; Tsukayama, Toomey, Faith, & Duckworth, 2010), wealth (Moffitt et al., 2011), juvenile delinquency (Benda, 2005), criminal activity in adulthood (Moffitt et al., 2011), and even longevity (Kern & Friedman, 2008).

Please cite this article in press as: Duckworth, A. L., & Kern, M. L. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* (2011), doi:10.1016/j.jrp.2011.02.004

Table 1
Executive function task subtypes, with frequency and convergence with other types of self-control measures.

Task	Example	General description	Exec function	Delay tasks	Questionnaire measures	
					Self	Informant
Go/No-go task	Continuous performance task (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956)	The subject develops a prepotent motor response (e.g., hitting the spacebar) to frequently appearing targets, and then must inhibit this response when a less frequently appearing non-target appears	r = .16 [.14, .17] k = 64, j = 131 N = 4855	r = .12 [.05, .19] k = 10, j = 17 N = 523	r = .11 [.08, .15] k = 30, j = 47 N = 1969	r = .15 [.11, .18] k = 23, j = 46 N = 1883
Stroop task	Stroop task (Stroop, 1935)	The subject must respond to a series of stimuli in a way that requires inhibition of a previously overlearned response	r = .14 [.13, .16] k = 97, j = 153 N = 7819	r = .16 [.08, .23] k = 4, j = 4 N = 663	r = .12 [.06, .18] k = 9, j = 10 N = 1919	r = .09 [.05, .13] k = 10, j = 23 N = 1284
Set switching task	Wisconsin card sorting task (Heaton & Pendleton, 1981)	The subject learns an initial set of rules, which change during subsequent trials. The task requires inhibition of previously learned rules and the adoption of a new set of rules	r = .15 [.13, 17] k = 74, j = 120 N = 6525	r = -.06 [−.32, .20] k = 1, j = 1 N = 60	r = .18 [.10, .26] k = 4, j = 6 N = 373	r = .23 [.15, .31] k = 7, j = 9 N = 430
Reflection task	Matching familiar figures task (Kagan, 1964)	A stimulus (e.g., a geometric pattern) is presented, and the subject must choose the correct response (e.g., the identical pattern) among very similar responses	r = .18 [.14, .22] k = 22, j = 33 N = 1646	r = .11 [−.01, .24] k = 3, j = 3 N = 244	r = .07 [.01, .13] k = 13, j = 16 N = 991	r = .14 [.09, .18] k = 15, j = 28 N = 1117
Stop-signal task	Stop-signal paradigm (Logan, 1994)	The subject performs a primary task and is presented with periodic signals, in response to which they must temporarily stop performing the primary task	r = .11 [.08, .14] k = 19, j = 41 N = 1982	r = .17 [.02, .31] k = 5, j = 5 N = 189	r = .17 [.08, .25] k = 8, j = 10 N = 402	r = .13 [.08, .18] k = 4, j = 13 N = 506
Motor inhibition task	Draw A line slowly task (Maccoby, Dowley, Hagen, & Degerman, 1965)	The subject must control or slow motor behavior	r = .17 [.13, .20] k = 17, j = 32 N = 1529	r = .11 [.04, .18] k = 4, j = 5 N = 665	r = .04 [−.01, .10] k = 8, j = 13 N = 919	r = .07 [−.01, .15] k = 4, j = 4 N = 564
Tower tasks	Tower of London test (Shallice, 1982, 1988)	The subject must plan ahead and resist immediate action in order to solve a problem	r = .14 [.11, .17] k = 22, j = 43 N = 1840	–	r = .16 [.02, .20] k = 1, j = 3 N = 24	r = .21 [.00, .40] k = 2, j = 2 N = 90
Trails task	Trail making task (Reitan & Wolfson, 1985)	Subjects first connect numbered circles in sequential order, and in a subsequent trial, connect numbers in an alternating pattern. Differences in performance between these two trials are recorded	r = .18 [.14, .21] k = 16, j = 27 N = 2035	r = .11 [.02, .20] k = 1, j = 1 N = 430	r = .05 [−.04, .14] k = 1, j = 1 N = 430	r = .14 [.08, .20] k = 3, j = 9 N = 604
Porteus maze task	Porteus maze (Porteus, 1942)	The subject completes a series of mazes of increasing complexity. Successfully completing the maze requires looking ahead and avoiding dead ends	r = .15 [.11, .19] k = 10, j = 21 N = 914	–	r = .11 [.03, .19] k = 3, j = 6 N = 278	–
Attention Task	Flanker Task (Eriksen & Eriksen, 1974)	Subjects must sustain attention to a target stimulus while ignoring distracters	r = .19 [.10, .28] k = 5, j = 10 N = 221	–	–	r = .33 [.22, .42] k = 3, j = 6 N = 140
Iowa Gambling Task	Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994)	The subject chooses among four decks of cards. Each card results in a monetary gain or loss, and some decks yield more long-run gains than others	r = .17 [.05, .27] k = 6, j = 10 N = 219	r = .00 [−.15, .15] k = 2, j = 2 N = 172	r = -.02 [−.10, .07] k = 2, j = 3 N = 409	–
Risk Task	Balloon Analogue Risk Task (Lejuez et al., 2002)	Subjects play a game in which rewards are steadily accrued but the risk of losing all accumulated rewards increases with each trial	r = .11 [−.04, .25] k = 2, j = 3 N = 109	r = -.05 [−.28, .19] k = 1, j = 1 N = 70	r = .16 [.05, .27] k = 2, j = 4 N = 168	–

Note. r = average correlation coefficient, based on a fixed effects model (weighted by the inverse variance); 95% confidence interval is given in brackets [ ]; k = number of samples included in average correlation; j = number of effect sizes included in average correlation; N = number of participants included in average correlation.

Our literature search revealed over 100 unique self- and informant-report questionnaires, most designed as stand-alone measures and a few as subscales of omnibus personality, temperament, or psychopathology inventories. Items on these questionnaires suggested considerable heterogeneity in the underlying constructs assessed. For instance, the Eysenck I_7 Impulsiveness Scale includes items about doing and saying things without thinking (Eysenck et al., 1984). The Self-Control Scale (Tangney,

Please cite this article in press as: Duckworth, A. L., & Kern, M. L. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* (2011), doi:10.1016/j.jrp.2011.02.004

Table 2
Delay task subtypes, with frequency and convergence with other types of self-control measures.

Task	Example	General description	Delay tasks	Exec function	Questionnaire measures	
					Self	Informant
Hypothetical delay	Kirby Delay discounting questionnaire (Kirby et al., 1999)	The subject makes hypothetical choices, each of which includes a smaller, more immediate reward and a larger, delayed reward	r = .20 [.07, .33] k = 3, j = 3 N = 226	r = .07 [.00, .13] k = 10, j = 18 N = 481	r = .16 [.13, .20] k = 15, j = 22 N = 1774	r = .27 [.20, .35] k = 2, j = 4 N = 304
Repeated trials delay	Newman task (Newman, et al., 1992) Single key impulsivity paradigm (Dougherty, Mathias, Marsh, & Jagar, 2005)	The subject plays a game where one type of response is immediately rewarded and another mutually exclusive type of response yields a delayed, but larger reward	r = .20 [−.03, .41] k = 2, j = 2 N = 76	r = .14 [.09, .18] k = 4, j = 15 N = 546	r = .10 [.02, .18] k = 4, j = 6 N = 534	r = .24 [.18, .30] k = 2, j = 4 N = 581
Sustained delay	Snack delay task (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996)	The subject must wait for the larger reward, while the smaller, immediate reward remains accessible	–	r = .13 [.04, .21] k = 4, j = 7 N = 310	r = .39 [.09, .62] k = 1, j = 1 N = 41	r = .11 [−.01, .23] k = 4, j = 6 N = 193
Real choice delay	Choice delay task (Kendall & Wilcox, 1979)	The subject chooses between a smaller, immediate reward or a larger, delayed reward	r = .23 [.08, .37] k = 1, j = 1 N = 164	r = .07 [−.06, .20] k = 2, j = 3 N = 165	r = .08 [−.02, .19] k = 2, j = 3 N = 196	r = .12 [.02, .21] k = 2, j = 3 N = 274

Note. For delay measures that offered hypothetical choices as well as choices between physical rewards, we coded them according to the category under which the majority of the choices fell. For example, if a task presented 10 different hypothetical delay questions, and the subject was aware that they would receive only 1 of their choices, the measure was coded as hypothetical delay. r = average correlation coefficient, based on a fixed effects model (weighted by the inverse variance); 95% confidence interval is given in brackets [ ]; k = number of samples included in average correlation; j = number of effect sizes included in average correlation; N = number of participants included in average correlation.

Baumeister, & Boone, 2004) casts a wider net, including items about acting "without thinking through all the alternatives," as well as "resisting temptation," and "concentrating." Likewise, the Barratt Impulsiveness Scale Version 11 (BIS-11) includes separate scales for motor impulsiveness, non-planning impulsiveness, and cognitive impulsiveness (Barratt, 1985).

1.3. Control of impulses vs. generation of impulses

In an attempt "to bring order to the myriad of measures and conceptions of impulsivity" (p. 684) in the individual difference and clinical psychology literatures, Whiteside and Lynam (2001) administered several previously published self-control questionnaires to a large sample of undergraduates. Item-level factor analyses produced four distinct factors (UPPS) interpreted as "discrete psychological processes that lead to impulsive-like behaviors" (Whiteside & Lynam, 2001; p. 685): *Urgency* is the inability to override strong impulses (e.g., "I have trouble controlling my impulses"). (Lack of) *premeditation*, is similar to Eysenck's conception of acting before thinking (e.g., "My thinking is usually careful and purposeful" (reverse-scored)). (Lack of) *perseverance* refers to the inability to focus on boring or difficult tasks (e.g., "I tend to give up easily"). Finally, *sensation seeking* refers to an attraction to exciting and risky activities (e.g., "I'll try anything once").

Whiteside and Lynam's (2001) UPPS model has been validated in subsequent studies (e.g., Miller, Flory, Lynam, & Leukefeld, 2003; Whiteside, Lynam, Miller, & Reynolds, 2005) but is not the only multidimensional model for self-control. Indeed, at least a dozen different factor structures for self-control (e.g., Barratt, 1985; Buss & Plomin, 1975; Miller, Joseph, & Tudway, 2004; White et al., 1994) have been suggested. One attractive feature of the UPPS is that it situates facets of self-control within the five-factor model of personality, relating urgency to neuroticism, perseverance and planning to conscientiousness, and sensation seeking to extraversion. Also notable is the similarity between the UPPS and Buss and Plomin's (1975) four-factor model, and at least partial overlap with other proposed factor structures for self-control. Finally, the distinction between sensation seeking impulses and a variety of psychological processes that direct behavior away from those impulses is consistent with dual-system models of self-control (Carver, Johnson, & Joormann, 2009; Eisenberg et al., 2004; Hofmann, Friese, & Strack, 2009; Metcalfe & Mischel, 1999; Steinberg, 2008). While dual-system models vary somewhat in their particulars, they all posit two opponent systems underlying the *generation* of quick, involuntary, and often consummatory impulses on the one hand, and the *control* of these impulses by deliberate, volitional processes impulses on the other.

1.4. Expectations about convergent validity

Our initial, qualitative survey of the self-control literature indicated considerable heterogeneity in the targeted psychological processes and, in addition, in the *level* of intended description. Some tasks and questionnaire items, it seemed, were designed to assess aggregate self-controlled behavior (i.e., ultimately behaving in accordance with long-term goals and standards at the expense of short-term gratification). For instance, the Self-Control Scale (Tangney et al., 2004) includes the item, "People would say that I have iron self-discipline." Other tasks and questionnaire items, in contrast, seemed to target the component psychological processes that precede and contribute to self-controlled behavior. In addition to the four processes specified by Whiteside and Lynam's (2001) UPPS model, we suggest that self-control may be facilitated by accurately weighing long-term and short-term consequences (delay discounting questionnaire; Kirby et al., 1999), following through on a decision to resist immediate gratification (preschool delay of gratification task, Mischel et al., 1989), suppressing habitual or automatic responses that conflict with one's goals (go/no-go task; Eigsti et al., 2006), and effectively regulating attention in the face of distractors (Attentional Network Task; Rueda et al., 2004).

Heterogeneity in the targeted dimensions of self-control and in the level of description suggests that correlations among diverse self-control measures may be relatively modest. In addition, measurement error, whether from task-specific or random error variance, should further attenuate estimates of convergent validity. A

meta-analysis of published studies reporting multi-method, multi-trait matrices of correlations found that more than 60% of the variance in personality measures was accounted for by task-specific and random error variance (Cote & Buckley, 1987).

1.5. The current study

In the current investigation, we meta-analytically examined evidence for convergent validity among self-control measures from published and unpublished studies that used at least two different measures of self-control. We had several specific goals in our analyses: First, we sought an overall estimate of the convergent validity among executive function, delay of gratification, and questionnaire self-control measures. Second, we examined sources of heterogeneity in convergent validity estimates, including type and subtype of self-control measure. Finally, we examined our meta-analytically derived correlation matrix for support of the UPPS model (Whiteside & Lynam, 2001).

2. Method

2.1. Literature search

We used two strategies to search the PsycINFO database for published articles and dissertations available by February 2008 that used more than one measure of self-control. First, keyword searches were conducted for self-control related terms and popular self-control measures.[2] Second, we identified relevant studies cited by articles identified in this search and from the library of the first author.

Over 7000 study abstracts were screened, resulting in 1280 potentially-relevant manuscripts. For studies that did not report all correlations among the self-control measures used, we emailed authors to request this information. Of the 542 authors we contacted, 101 authors (18.6%) provided correlation matrices.

2.2. Inclusion and exclusion criteria

Studies selected for this meta-analysis were written in English and reported at least one bivariate correlation coefficient for two different measures of self-control. We excluded studies that reported correlations of $r = 1$ or reported only Spearman rank or partial correlations. We also excluded measures that did not meet our broad definitional criteria for self-control (i.e., the self-governance of responses in order to achieve long-term benefit at the expense of short-term gratification). Finally, we excluded correlations between subscores of a common self-control measure (e.g., correlations between error and latency scores from a single executive function task; subscale scores from a single questionnaire) or between two different versions of a common questionnaire.

2.3. Moderator coding procedure

A total of five trained coders recorded sample characteristics and correlations. Each correlation was coded independently by at least two coders to ensure reliability. Conflicts were resolved by discussion and re-examination of studies in question. In addition to sample sizes and correlation coefficients, the coders recorded the following variables:

2.3.1. Name, type, and subtype of measure

We recorded the name of each measure and classified each according to one of four types: executive function task, delay of gratification task, self-report questionnaire, or informant-report questionnaire. Executive function and delay of gratification subtypes were classified according to the categories in Tables 1 and 2. Questionnaire measures were classified by the name of the scale.

2.3.2. Source

Each correlation was classified as originating from one of three sources: published articles or book chapters ($k = 131$), email correspondence with study authors ($k = 86$), or dissertations ($k = 65$).

2.3.3. Age

Mean ages for $k = 281$ samples were divided into the following ordinal categories: 0–5, 6–12, 13–17, 18–21, 22–29, 30–39, 40–49, 50–59, 60–69, and 70+ years. For samples that reported age ranges but not means, we categorized each sample into the age bracket in which most of the sample fell (e.g., age range of 17–22 was coded as the 18–21 category).

2.3.4. Gender

We recorded the number of male and female participants for every study that reported the relevant descriptive statistics. From these numbers we calculated the percent of females included in the study ($k = 247$).

2.3.5. Sample type

We recorded whether the study sample represented either *non-clinical* or *clinical/mixed* populations. *Non-clinical samples* ($k = 127$) were typically convenience samples of non-clinical individuals. *Clinical/mixed samples* ($k = 155$) included at least some participants with a psychological disorder or other impairment (e.g., ADHD, learning disorder, behavioral problems, delinquency, anxiety, substance abuse, neurological impairment, incarceration). Samples including participants who had been administered psychoactive medication were excluded.

2.4. Data analyses

We used the Pearson correlation coefficient (r) as the effect size (ES) measure. For the vast majority of included studies, multiple correlation coefficients based on a single sample were reported. We computed synthetic effect sizes by aggregating homogeneous dependent effect sizes within samples. This approach assumes that correlations within a sample are based on measures of a common latent variable and produces accurate, if somewhat conservative, meta-analytic estimates of the effect size (Hedges, 2007).

An important conceptual question for meta-analyses is whether to assume a fixed or random effects model (cf. Field, 2001; Hunter & Schmidt, 2004; Schulze, 2007). The fixed effects model provides more precise and reliable estimates but cannot be generalized to broader populations (Cooper, 1998). The random effects model allows the amount of variance between and within studies to be considered, but has statistical disadvantages when the number of observations for any particular analysis is small (Hunter & Schmidt, 2004). We report both fixed and random effects analyses for the overall analysis and, because of reduced sample size, fixed effects estimates only for moderator analyses. Since the weights used in the aggregation of correlation coefficients can influence the results, we followed the recommendation of Hedges and Olkin (1985) and used the inverse variance as weights in the analyses.

[2] Keyword searches included at least one self-control related keyword (see Footnote 1) and either: multi-method, multi-source, measure, assess, or validity. Measure searches used all possible pairwise combinations of the following popular self-control measures and terms: matching familiar figures, circle tracing, draw-a-line, walk-a-line, draw-a-star, Stroop, Gordon diagnostic, go/no-go, Wisconsin card sort, trail making, Eysenck impulsiveness, Dickman impulsivity inventory, Barratt impulsiveness, EASI-III impulsivity, child behavior-checklist, Conners rating scale, self-control rating scale, California Q-set, delay of gratification, discount delay, time preference.

Please cite this article in press as: Duckworth, A. L., & Kern, M. L. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* (2011), doi:10.1016/j.jrp.2011.02.004

When possible, it is helpful to correct for artifacts, such as range restriction and lack of reliability, when estimating population effect sizes (Hunter & Schmidt, 2004). However, such corrections require information that was not available in most of the included studies. We therefore did not correct for any artifacts in our analyses, and results should be considered with this limitation in mind.

3. Results

In total, 236 studies met our inclusion criteria, comprised of $k = 282$ independent samples and $N = 33{,}564$ participants. Altogether, $j = 3654$ effect sizes were culled from these study reports, which were aggregated to 282 effect sizes for the overall and moderator analyses (at the sample level), and 907 effect sizes for intertype comparisons. Study characteristics are summarized in Appendix A, with corresponding references in Appendix B.

Based on the total sample of 282 independent effect sizes, the mean effect size across self-control measures was medium in size ($r_{random} = .27$ [95% CI = .24, .30]; $r_{fixed} = .34$ [.33, .35]). As expected, there was substantial heterogeneity, $Q(281) = 2152.20$, $p < .001$.

3.1. Moderation by sample characteristics

In order to account for heterogeneity in effect sizes, we examined available sample characteristics as potential moderators. The number of samples k varied slightly among moderator analyses because information for moderators was missing in a very small proportion of samples.

Overall, sample characteristics explained minimal differences in effect sizes. Year of publication, the source of effect size (dissertation, published study, email correspondence), and sample type (normative vs. clinical/mixed) each reduced heterogeneity variance by about one percent, $Q(1) = 22.60$, $Q(1) = 27.51$, $p < .001$, and $Q(1) = 31.79$ (all $ps < .001$) respectively. Gender composition accounted for less than one percent of the variance in observed effect sizes ($Q(1) = 16.45$, $p < .001$). Although statistically significant, these study characteristic effects were minuscule in magnitude, suggesting that the convergent validity of self-control measures has not strengthened over the past 45 years, publication bias has not favoured larger effect sizes, and effects were fairly constant across clinical and non-clinical populations and across male and female participants.

Age was a statistically significant moderator, $Q(9) = 264.73$, $p < .001$, accounting for 12% of the variance. However, we noted that the type of self-control measure employed varied by age, with younger samples including disproportionately more informant-report questionnaires (completed by teachers and parents) and older samples including disproportionately more executive function measures. When we examined age as a moderator of effect sizes separately for each of the four types of self-control measures, there were no consistent or interpretable trends.

3.2. Convergent validity by type of self-control measure

In contrast to sample characteristics, measure type explained 53% of the overall variance in effect sizes, $Q_{total}(906) = 8049.98$, $p < .001$; $Q_{type}(9) = 4261.09$, $p < .001$. As shown in Table 3, informant-report questionnaires demonstrated the strongest evidence of convergent validity. Correlations among informant-report questionnaires ($r = .54$ [.53, .55]) were slightly higher than correlations among self-report questionnaires ($r = .50$ [.48, .51], $Z = 3.45$, $p = .001$), and much higher than correlations among executive function tasks ($r = .15$ [.14, .17], $Z = 32.69$, $p < .0001$) and among delay tasks ($r = .21$ [.09, .32], $Z = 6.19$, $p < .0001$). Interestingly, delay of gratification tasks, which appeared less frequently in the reviewed literature than other types of measures, demonstrated homogeneity across effect sizes, $Q(3) = .57$, $p = .90$. Delay tasks were more strongly associated with informant-report questionnaires than were executive function tasks ($r_{delay} = .21$ [.17, .25], $r_{exec} = .14$ [.12, .15], $Z = 2.33$, $p = .02$), with a similar trend for self-report questionnaires ($r_{delay} = .15$ [.11, .18], $r_{exec} = .10$ [.08, .12], $Z = 1.87$, $p = .06$).

3.3. Convergent validity by subtype of self-control measure

We next examined subtypes of self-control measures, considering convergence with other measures of the same subtype (e.g., Stroop tasks with other executive function tasks) and with other types of measures (e.g., Stroop tasks with self-report questionnaires). These analyses are summarized in Tables 1 and 2. Marked unevenness in the availability of convergent validity estimates (i.e., higher numbers of effect sizes for some measures than for others) was notable and should be kept mind when considering these results.

Among executive function tasks, go/no-go tasks, Stroop tasks, and set switching tasks were used most frequently; attention tasks, gambling tasks, and risk tasks were used much less frequently. Despite the paucity of multi-method studies using attention tasks, higher than average correlations for attention tasks with informant-report questionnaires ($r = .33$ [.22, .42], $Z = 2.33$, $p = .02$) were notable. Otherwise, there were no particularly striking differences in convergent validity for executive function tasks.

Similarly, there was no salient evidence for the superior convergent validity of one subtype of delay task over another. Within subtype of delay task, correlations (e.g., between two different delay tasks) ranged from $r = .20$ to .23. Convergence with other types of self-control measures was difficult to evaluate because of limited data. None of the four delay task subtypes demonstrated consistently stronger convergent validity with non-delay tasks.

Unlike executive function and delay tasks, the 104 differently named questionnaire measures included in this meta-analysis did not lend themselves to a defensible *a priori* taxonomy of subtypes. A correlation matrix in which questionnaires were organized simply by name of measure produced 98.4% blank cells (i.e., missing values). Thus, although questionnaire measures demonstrated significant heterogeneity ($Q_{self}(56) = 686.48$ and $Q_{informant}(141) = 1628.17$), we were unable to test whether certain subtypes of questionnaires demonstrated stronger convergent validity than others.

3.4. Evaluating the UPPS structure with self-report questionnaires

Finally, we examined correlations among self-report questionnaires for evidence of the four-factor UPPS structure proposed by Whiteside and Lynam (2001). In consultation with a co-author of the UPPS model (Lynam, personal correspondence November 2009), we reviewed the most popular 80 questionnaires and subscales and recorded the UPPS facet(s) with which they seemed most aligned.[3] We then created a synthesized correlation matrix by aggregating effect sizes for all measures tapping each facet. (Necessarily, the aggregated values were not independent; scales that were rated as assessing multiple facets were included in multiple places in the synthesized matrix.) We expected that average correlations within a facet (i.e., values on the diagonal) to be stronger than correlations across facets (i.e., off-diagonal values).

As shown in Table 4, sensation seeking demonstrated stronger within-facet associations ($r = .48$ between different sensation seeking questionnaires) than associations with other facets (rs with

[3] Available from the first author by request.

Table 3
Correlation matrix for four types of self-control measures.

	Exec function	Delay tasks	Self-report	Informant report
Executive function tasks	r = .15 [.14, .17] k = 147, j = 324 N = 12,118 Q = 536.73, p < .001	r = .11 [.08, .15] k = 19, j = 41 N = 1470 Q = 47.09, p = .21	r = .10 [.08, .12] k = 51, j = 120 N = 3922 Q = 165.59, p = .003	r = .14 [.12, .15] k = 45, j = 140 N = 3802 Q = 216.01, p < .001
Delay of gratification tasks		r = .21 [.09, .32] k = 4, j = 4 N = 270 Q = .57, p = .90	r = .15 [.11, .18] k = 19, j = 33 N = 2317 Q = 42.02, p = .11	r = .21 [.17, .25] k = 9, j = 17 N = 1188 Q = 26.13, p = .05
Self-report questionnaire			r = .50 [.48, .51] k = 47, j = 57 N = 7782 Q = 686.48, p < .001	r = .48 [.46, .50] k = 17, j = 29 N = 3770 Q = 440.11, p < .001
Informant-report questionnaire				r = .54 [.53, .55] k = 44, j = 142 N = 9881 Q = 1628.17, p < .001

Note. r = average correlation coefficient, based on a fixed effects model (weighted by the inverse variance); 95% confidence interval is given in brackets [ ]; k = number of samples included in average correlation; j = number of effect sizes included in average correlation; N = number of participants included in average correlation; Q = Q-statistic for homogeneity test; p = p-value associated with the Q-statistic.

other facets ranging from .36 to .40, all comparison *p*s < .01). The other three proposed facets in the UPPS model were not consistently different from one another. In further support that sensation seeking differs from other aspects of self-control, the correlation between delay tasks and sensation seeking questionnaires (r = .18, [.12, .25], k = 4, j = 12, N = 271) was higher than the correlation between delay tasks and other questionnaires (r = .13, [.11, .15], k = 17, j = 99, N = 2200), whereas the correlation between executive function tasks and sensation seeking questionnaires (r = .07, [.04, .11], k = 9, j = 56, N = 602) was lower than the correlation between executive function tasks and other self-control questionnaires (r = .11, [.10, .12], k = 50, j = 492, N = 3846). However, tests for the significance of these comparisons did not reach significance, suggesting that more studies are needed confirm these trends.

4. Discussion

Across 282 multi-method studies and over 33,000 participants, we found moderate convergence across self-control measures. Correlations did not vary systematically by sample characteristics, including gender, sample type, publication year, or whether the correlations were extracted from published articles, dissertations, or email correspondence with authors. In contrast, over half of the heterogeneity in correlations was explained by the type of self-control measure used. Both within and across types, informant-report questionnaires demonstrated the strong evidence of convergent validity, followed closely by self-report questionnaires, then delay of gratification tasks and, finally, executive function tasks. Notably, all estimates of convergent validity were statistically significant and at least small in magnitude.

Despite the large number of studies and participants included in the meta-analysis, there were insufficient data to draw strong conclusions about the relative convergence of specific subtypes. There was substantial heterogeneity in the convergent validity of executive function tasks, both with other executive function measures and with other types of measures. With this caveat in mind, we note that none of the three most commonly used executive functions tasks (go/no-go, Stroop, and set switching) demonstrated uniformly higher correlations with other executive function tasks or with delay or questionnaire measures of self-control. Other executive function tasks, most notably attention tasks, demonstrated higher convergence, but were too rarely used in multi-method studies to draw firm conclusions. Thus, while there was substantial heterogeneity in the convergent validity of executive function tasks, this heterogeneity was not well-explained by the subtype of executive function task used.

In contrast, we found no evidence for differences in convergent validity among hypothetical, repeated trials, sustained, or real choice delay of gratification tasks. Homogeneity of effect sizes among the four subtypes of delay tasks suggests that these tasks differ less from each other than do executive function tasks. One possible explanation for this pattern of findings is that while different delay tasks may to some extent tap different processes related to delay of gratification (e.g., making the choice to wait for a larger reward vs. sustaining that choice in the face of temptation), as a group they may tap more similar processes than do executive function tasks. Alternatively, since estimates of heterogeneity are dependent upon the number of included effect sizes (Hedges & Olkin, 1985), we cannot rule out the possibility that heterogeneity estimates would have been larger and statistically significant had more data been available from multi-method studies employing delay tasks.

4.1. Distinguishing sensation seeking from other processes relevant to self-control

Our attempt to test the four-factor UPPS structure (Whiteside & Lynam, 2001) using correlations among self-report measures of self-control was constrained by the available data. Nevertheless, *a priori* categorization of questionnaires according to the UPPS factor structure and subsequent analysis of their intercorrelations suggested that sensation seeking could be distinguished from urgency, lack of perseverance, and lack of premeditation. In contrast, we failed to find compelling evidence for separation among the remaining three factors. These analyses support a distinction between sensation seeking tendencies and, broadly, the psychological processes that oppose these tendencies.

Our findings are consistent with Miller et al. (2003), who found that of the four UPPS factors, sensation seeking was the *least* correlated with the remaining UPPS factors. Many authors consider self-control to be coextensive with Big Five conscientiousness (Moffitt et al., 2011), and the UPPS urgency, lack of planning, and lack of persistence facets have been identified (inversely) with Big Five conscientiousness (MacCann, Duckworth, & Roberts, 2009). Likewise, Romer, Duckworth, Sznitman, and Park (2010) found that

Table 4
Testing the UPPS factor structure with self-reported self-control measures.

	Urgency	Lack of perseverance	Lack of planning	Sensation seeking
Urgency	r = .46 [.45, .48] k = 32, j = 56 N = 6112	r = .48 [.46, .49] k = 22, j = 51 N = 4827	r = .45 [.44, .47] k = 34, j = 80 N = 6287	r = .40 [.38, .42] k = 20, j = 67 N = 3008
Lack of perseverance		r = .47 [.44, .49] k = 10, j = 16 N = 2721	r = .47 [.45, .49] k = 23, j = 57 N = 4766	r = .39 [.36, .41] k = 11, j = 30 N = 1707
Lack of planning			r = .40 [.39, .42] k = 26, j = 60 N = 4188	r = .36 [.34, .37] k = 16, j = 74 N = 2902
Sensation seeking				r = .48 [.46, .51] k = 3, j = 13 N = 827

Note. r = mean effect size, weighted by sample size; 95% confidence intervals are given in brackets [ ]; k = number of samples included in average correlation; j = number of effect sizes included in average correlation; N = number of participants included in average correlation. Average correlations are not independent; effects were allowed in multiple categories.

sensation seeking peaks sharply during late adolescence and then falls in early adulthood, whereas the developmental trajectories for future time perspective and delay of gratification over the same period are monotonically positive.

Recent neuroscience research suggests that sensation seeking impulses may be generated by dopaminergic subcortical structures whose activity normatively spikes during adolescence, whereas the psychological processes associated with inhibitory control, premeditation, and perseverance correspond to slowly maturing frontal areas (Steinberg, 2008). Collectively, this evidence is consistent with dual-system models of self-control positing impulse-generating and impulse-controlling systems (Carver et al., 2009; Eisenberg et al., 2004: Hofmann et al., 2009; Metcalfe & Mischel, 1999; Steinberg, 2008).

4.2. Implications for research and theory

How do these cross-method correlations compare to those observed for traits other than self-control? Meyer et al. (2001) compiled meta-analytic estimates of cross-method convergent associations for a wide range of psychological constructs, providing a benchmark by which to judge the current estimates. Generally, correlations between self- and informant-report questionnaire measures in Meyer's review were medium in size. For instance, Meyer reported correlations between parent reports of children's behavioral and emotional problems and either self-reports or teacher reports were r = .29, correlations between self-report and spouse/partner reports of personality and mood were r = .29, and correlations between supervisor-report and peer-report ratings of job performance were r = .34. In contrast, Meyer found correlations between task and questionnaire measures to be small in size. For instance, self-report and cognitive test measures of memory problems correlated r = .13; self-report and cognitive test measures of attentional problems correlated r = .06. Overall, it seems that the evidence for convergent validity among self-control measures in the present meta-analysis compares favorably to Meyer's meta-analytic estimates for other psychological constructs.

The dramatically stronger evidence for convergent validity among questionnaire measures, in both Meyer's review and the current investigation, has practical implications for self-control researchers. In particular, researchers facing time and budget constraints may be advised to choose a single informant- or self-report questionnaire over any single executive function or delay of gratification task measure. Task measures, of course, have important advantages over questionnaires (e.g., objective performance outcomes that are difficult if not impossible to fake). However, the comparatively weaker evidence of convergent validity for task measures points to substantial random and task-specific error variance, notoriously problematic for executive function tasks in particular (Rabbitt, 1997) but also well-known for performance task measures in general (Epstein, 1979). In the time required to administer a single executive function or delay task, *many* questionnaire items can be administered. Multiple measures reduce error variance (Brown, 1910; Spearman, 1910), and furthermore, the response to any particular questionnaire item (e.g., "I have trouble resisting temptation") implicitly asks the respondent for an aggregate judgment of behavior across *multiple* situations and observations.

When using task measures of self-control, therefore, we recommend aggregating across measures in order to reduce error variance. For instance, Beck, Carlson, and Rothbart (2011) recently demonstrated that average performance across three vs. six executive function tasks correlated r = .22 and r = .30, respectively, with informant-report questionnaire measures of self-control. Notably, both of these observed convergent validities exceeded our meta-analytic average for single executive function tasks, r = 14.

Perhaps the optimal measurement strategy is to include both task and questionnaire measures. For instance, Duckworth and Seligman (2005, Study 2) measured self-control using a battery of self- and informant-report questionnaires, as well as two delay of gratification measures. Estimates of convergent validity were consistent with those in the current meta-analysis, and the composite measure of self-control predicted objectively measured academic performance better than did any single measure alone.

While we did not find compelling evidence in the current investigation for the distinction among other facets proposed by Whiteside and Lynam's (2001) UPPS model, absence of evidence is not evidence of absence. Our opportunistic analyses using the available meta-analytic correlation matrix were far from a definitive test of the UPPS factor structure. One promising direction for future research would be a more systematic, multimethod investigation of the components of self-control. Ideally, separate measures assessing sensation seeking and other "impulsive" processes would be administered, along with measures designed to assess the psychological processes posited to modulate those impulses. Including research participants of diverse ages would allow researchers to trace the developmental trajectories of these distinct psychological processes over the life course; divergent developmental trajectories would provide evidence, in addition to conventional factor analyses, for the separation of self-control processes. Finally, processes might be distinguished from each other by differential predictive validity for theoretically-relevant outcomes.

Please cite this article in press as: Duckworth, A. L., & Kern, M. L. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* (2011), doi:10.1016/j.jrp.2011.02.004

5. Conclusion

The promise of psychology as a cumulative science depends not only upon field-unifying theories and well-designed studies, but also upon valid, consensually understood measures (Mischel, 2009). On the basis of the current meta-analysis, we suggest that evidence for the convergent validity of self-control measures is adequate – and as strong as the evidence of convergent validity for other psychological measures. Looking to the future, we hope the current investigation encourages collaboration among researchers of diverse methodological traditions. Such interdisciplinary partnerships should dramatically accelerate our understanding of the coherent, yet complex, construct of self-control.

Acknowledgements

This study was funded by the National Institute on Aging K01 mentored research scientist award (Duckworth PI) and the John Templeton Foundation (Seligman PI; Duckworth Co-PI). The authors acknowledge equal contribution to the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jrp.2011.02.004.

References

Baker, D. B., Taylor, C. J., & Leyva, C. (1995). Continuous performance tests: A comparison of modalities. *Journal of Clinical Psychology, 51*, 548–551.

Banfield, J., Wyland, C. L., Macrae, C. N., Munte, T. F., & Heatherton, T. F. (2004). The cognitive neuroscience of self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *The handbook of self-regulation* (pp. 62–83). New York: Guilford Press.

Barratt, E. S. (1985). Impulsive subtraits: Arousal and information processing. In J. T. Spence & C. E. Izard (Eds.), *Motivation, emotion, and personality* (pp. 137–146). North Holland: Elsevier Science Publishers.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351–355.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7–15.

Beck, D. M., Carlson, S. M., & Rothbart, M. K. (2011). *Executive function, effortful control and parent report of children's temperament*. University of Washington.

Benda, B. B. (2005). The robustness of self-control in relation to form of delinquency. *Youth & Society, 36*, 418–444.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 91–116). East Sussex: Psychology Press.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. Oxford: Wiley Interscience.

Carver, C. S., Johnson, S. L., & Joormann, J. (2009). Two-mode models of self-regulation as a tool for conceptualizing effects of the serotonin system in normal behavior and diverse disorders. *Current Directions in Psychological Science, 18*, 195–199.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Beverly Hills, CA: Sage.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance. Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*, 315–318.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*, 491–569.

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

Dougherty, D. M., Mathias, C. W., Marsh, D. M., & Jagar, A. A. (2005). Laboratory behavioral measures of impulsivity. *Behavior Research Methods, 37*, 82–90.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939–944.

Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychology and Personality Science, 1*, 311–317.

Eigsti, I.-M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M., et al. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science, 17*, 478–484.

Eisenberg, N., Spinrad, T. L., Fabes, R. A., Reiser, M., Cumberland, A., Shepard, S., et al. (2004). The relations of effortful control and impulsivity to children's resiliency and adjustment. *Child Development, 75*, 25–46.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097–1126.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Evans, G. W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly, 23*, 504–514.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology. Special Issue: Impulsivity, 146*, 348–361.

Eysenck, S. B., Easton, G., & Pearson, P. R. (1984). Age norms for impulsiveness, venturesomeness and empathy in children. *Personality and Individual Differences, 315*, 321.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A monte carlo comparison of fixed- and random-effects methods. *Psychological Methods, 6*, 161–180.

Fiske, D. W. (1971). *Measuring the concepts of personality*. Chicago, Illinois: Aldline Publishing Co..

Freud, S. (1922). *Beyond the pleasure principle*. New York: Liveright.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science, 5*, 33–36.

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

Heaton, R. K., & Pendleton, M. G. (1981). Use of neuropsychological tests to predict adult patients' everyday functioning. *Journal of Consulting and Clinical Psychology, 49*, 807–821.

Hedges, L. V. (2007). Meta-analysis. In C. R. Rao & S. Sinharay (Eds.). *Handbook of statistics* (Vol. 26, pp. 919–953). Amsterdam: Elsevier.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from dual-systems perspective. *Perspectives on Psychological Science, 4*, 162–176.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Kagan, J. (1964). *Matching familiar figures test*. Cambridge, MA: Harvard University.

Kendall, P. C., & Wilcox, L. E. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology, 47*, 1020–1029.

Kern, M. L., & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology, 27*, 505–512.

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

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507.

Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9*, 491–512.

Krueger, R. F., Caspi, A., Moffitt, T. E., White, J., & Stouthamer-Loeber, M. (1996). Delay of gratification, psychopathology, and personality: Is low self-control specific to externalizing problems? *Journal of Personality, 64*, 107–129.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*, 75–84.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (1st ed., pp. 189–239). Academic Press.

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences, 19*, 451–458.

Maccoby, E. E., Dowley, E. M., Hagen, J. W., & Degerman, R. (1965). Activity level and intellectual functioning in normal preschool children. *Child Development, 36*, 761–770.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*, 3–19.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *Amercian Psychologist, 56*, 128–165.

Miller, E. K. (2000). The prefrontal cortex: No simple matter. *Neuroimage, 2*, 447–450.

Miller, E., Joseph, S., & Tudway, J. (2004). Assessing the component structure of four self-report measures of impulsivity. *Personality and Individual Differences, 37*, 349–358.

Miller, J., Flory, K., Lynam, D., & Leukefeld, C. (2003). A test of the four-factor model of impulsivity-related traits. *Personality and Individual Differences, 34*, 1403–1418.

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

Mischel, W. (1961). Preference for delayed reinforcement and social responsibility. *Journal of Abnormal & Social Psychology, 62*, 1–7.

Mischel, W. (2009). Becoming a cumulative science. *APS Observer, 22*, 18.

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Moffitt, T., Arseneault, L., Belsky, D., Dickson, N., Hancox, R., Harrington, H. L., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1010076108.

Newman, J. P., Kosson, D. S., & Patterson, C. M. (1992). Delay of gratification in psychopathic and nonpsychopathic offenders. *Journal of Abnormal Psychology, 101*(4), 630–636.

Peterson, C. (2006). *A primer in positive psychology*. New York: Oxford University Press.

Porteus, S. D. (1942). *Qualitative performance in the maze test*. Vineland, NJ: The Smith Printing House.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–38). East Sussex: Psychology Press.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan neuropsychological test battery: Therapy and clinical interpretation*. Tucson, AZ: Neuropsychologycal Press.

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

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20*, 343–350.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., et al. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*, 1029–1040.

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie/Journal of Psychology, 215*, 90–103.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 298*, 199–209.

Shallice, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.

Singer, J. L. (1955). Delayed gratification and ego development: implications for clinical and experimental research. *Journal of Consulting Psychology, 19*, 259–266.

Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation, 11*, 61–77.

Spearman, Charles (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Development Review, 28*, 78–106.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*, 271–322.

Tsukayama, E., Duckworth, A. L., & Kim, B. E. (submitted for publication). Resisting everything except temptation: An explanation for domain-specific impulsivity. *European Journal of Psychology*.

Tsukayama, E., Toomey, S. L., Faith, M. S., & Duckworth, A. L. (2010). Self-control protects against overweight status in the transition from childhood to adolescences. *Archives of Pediatrics and Adolescent Medicine, 164*, 631–635.

Wallace, H. M., & Baumeister, R. F. (2002). The effects of success versus failure feedback on further self-control. *Self and Identity, 1*, 35–41.

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal of Abnormal Psychology, 103*, 192–205.

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*, 669–689.

Whiteside, S. P., Lynam, D. R., Miller, J. D., & Reynolds, S. K. (2005). Validation of the UPPS impulsive behaviour scale: A four-factor model of impulsivity. *European Journal of Personality, 19*, 559–574.

Wilde, O. (2009). An ideal husband. The Project Gutenberg eBook #885. <www.gutenberg.org/files/885/885-h/885-h.htm> (original work published 1912).

Williams, P. J., & Thayer, J. F. (2009). Executive functioning and health: Introduction to the special series. *Annals of Behavioral Medicine, 37*, 101–105.

Zaparniuk, J., & Taylor, S. (1997). Impulsivity in children and adolescents. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment and treatment* (pp. 158–179). New York: Guilford Press.

Please cite this article in press as: Duckworth, A. L., & Kern, M. L. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* (2011), doi:10.1016/j.jrp.2011.02.004