Assessing the Implicit Theory of Willpower for Strenuous Mental Activities Scale: Multigroup, Across-Gender, and Cross-Cultural Measurement Invariance and Convergent and Divergent...

2 authors:

Christopher M Napolitano
University of Illinois, Urbana-Champaign
35 PUBLICATIONS **400** CITATIONS

SEE PROFILE

Veronika Job
Technische Universität Dresden
47 PUBLICATIONS **1,080** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Implicit theories of willpower View project

Family-Based Alternative Life Paths (ALPs) View project

Assessing the Implicit Theory of Willpower for Strenuous Mental Activities Scale:

Multigroup, across-gender, and cross-cultural measurement invariance and convergent and

divergent validity

Christopher M. Napolitano[1,2] and Veronika Job[2]

[1] University of Illinois at Urbana-Champaign, Department of Educational Psychology

[2]University of Zurich, Department of Psychology

Abstract

Why do some people struggle with self-control (colloquially called *willpower*) whereas others are able to sustain it during challenging circumstances? Recent research showed that a person's implicit theories of willpower – whether they think self-control capacity is a limited or nonlimited resource – predict sustained self-control on laboratory tasks and on goal-related outcomes in everyday life. The present research tests the **I**mplicit **T**heory of **W**illpower for Strenuous **M**ental Activities Scale (or ITW-M) Scale for measurement invariance across samples and gender within each culture, and 2 cultural contexts (the U.S. and Switzerland / Germany). Across a series of multi-group confirmatory factor analyses, we found support for the measurement invariance of the ITW-M scale across samples within and across two cultures, as well as across men and women. Further, the analyses showed expected patterns of convergent (with life-satisfaction and trait-self-control) and discriminant validity (with implicit theory of intelligence). These results provide guidelines for future research and clinical practice using the ITW-M scale for the investigation of latent group differences, e.g. between gender or cultures.

Public Significance Statement

A person's implicit theories of willpower – whether they think self-control is a limited or nonlimited resource – predict self-control assessed with laboratory tasks and in everyday life. This study demonstrates that a six-item scale invariantly measures a person's implicit theory of willpower for strenuous mental activities across ten datasets, two cultural contexts, and men and women.

Keywords: implicit theories of willpower; self-control; measurement invariance; confirmatory factor analyses; cross-cultural analyses;

Assessing the Implicit Theory of Willpower for Strenuous Mental Activities Scale:

Multigroup, Across-Gender, and Cross-Cultural Measurement Invariance and Convergent

and Divergent Validity

Recent research has shown that the way people think about their self-control capacity (called *willpower* in everyday language) as either a limited or nonlimited resource affects their self-control performance on consecutive tasks in the laboratory and has far reaching consequences on their everyday self-control, their personal goal striving, and their well-being (Bernecker, Herrmann, Brandstätter, & Job, in press; Job, Dweck, & Walton, 2010; Job, Walton, Bernecker, & Dweck, 2015). Much of this research was conducted using a six-item scale that assesses people's beliefs about their willpower in the domain of strenuous mental activity (hereafter the **I**mplicit **T**heory of **W**illpower for Strenuous **M**ental Activities Scale, or ITW-M). This scale assesses if people believe that after engaging in strenuous mental activity their willpower is either *limited*, and they need rest or food to work effectively again, or alternatively, that their willpower is *non-limited*, and that perhaps such strenuous mental activities actually activate their self-control resources.

Some properties of the ITW-M scale (e.g., descriptive, internal consistency, retest-reliability, and correlations with related constructs) have been reported in previous research. However, researchers using the ITW-M scale have assumed that the items index an underlying latent construct of a person's implicit theory about their willpower for strenuous mental tasks, and that across samples, cultural contexts, and participant gender, the ITW-M scale consistently indexes this underlying construct in the same way. We test these assumptions here. The present research tests for multigroup, across-gender, cross-cultural latent measurement invariance of the ITW-M scale across 11 samples, assesses the ITW-M scale in terms of convergent and

discriminant validity, and aims to provide readers with guidelines for how to use the ITW-M

scale to test for group differences in their research and practice.

Implicit Theories

The study of implicit theories about personal attributes has a long tradition (Dweck &

Leggett, 1988; Dweck, 1986; Hong, Chiu, Dweck, Lin, & Wan, 1999; Molden & Dweck, 2006).

Implicit theories are people's beliefs about themselves and their world. Similar to scientists who

develop theories to explain phenomena in the world, laypersons hold theories and beliefs about

different personal characteristics, such as intelligence, personality, or health (Burnette, O'Boyle,

VanEpps, Pollack, & Finkel, 2013; Dweck, Chiu, & Hong, 1995). Implicit theories provide a

conceptual framework that people use to ascribe meaning to the world around them. Every

experience is interpreted on the basis of a person's implicit theories, determining their

motivational and behavioral responses (Molden & Dweck, 2006; Crum, Salovey, & Achor,

2013).

In contrast to scientific theories, laypersons' theories are often implicit in the sense that

people are not aware of them or how they affect their behavior. However, researchers use

questionnaires to assess people's implicit theories and assume that people can provide

appropriate answers when explicitly asked to think about what they believe. An additional basic

assumption is that answers to these questions measure the same underlying construct across

groups, and thus facilitate the ability to compare group differences. The aim of the present

research was to test this basic assumption with regard to one facet of a person's *implicit theories*

about willpower –whether it is a limited or nonlimited resource (Job et al., 2010).

Implicit Theories About Willpower

Implicit theories about willpower have evoked considerable attention in self-control research because they addressed a highly influential theory: the strength model of self-control (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Baumeister, Vohs, & Tice, 2007). This model posits that self-control, i.e., the effortful inhibition of impulses, relies on a limited energy resource, which is partially depleted by every act of self-control, directly reducing the capacity to exert further self-control—a phenomenon termed ego depletion. An impressive amount of laboratory experiments provided support for this model. In general, this research shows that, after an initial task requiring self-control, like focusing one's attention in a mental task or inhibiting thoughts of a white bear, people's performance on subsequent unrelated tasks requiring self-control can be diminished (for a meta-analysis, see Hagger, Wood, Stiff, & Chatzisarantis, 2010).

The implicit theories about willpower research adopted the implicit theories perspective to the context of self-control. The core hypothesis was that people differ in the degree in which their beliefs about willpower correspond with the strength model of self-control and that differences in these beliefs determine their self-control capacity. To this end, Job and colleagues (2010) developed the ITW-M scale to assess a person's implicit theory about willpower regarding strenuous mental activities. This scale contains statements such as "After a strenuous mental activity, your energy is depleted and you must rest to get it refueled again" (reflecting a limited-resource theory) or "After a strenuous mental activity, you feel energized for further challenging activities" (reflecting a nonlimited-resource theory; see Appendix 1 for the full scale).

A first experimental study showed that scores on the ITW-M scale predicted who would show an ego-depletion effect on two consecutive mental self-control tasks and who would not. Only participants who believed that willpower to stick to mental tasks is a *limited and an easily depleted resource* performed worse on an unrelated task after they had exerted self-control beforehand. These results were found both when measuring ITW-M as an individual difference variable and when manipulating them experimentally, supporting the assumption of their causal role (Job et al., 2010).

More recent research indicates that ITW-M also predicts self-control outcomes in everyday life. When students face high demands, their self-control is impaired and they receive lower grades when they endorse a limited-resource theory as compared to students with a nonlimited theory (Job et al., 2015). For example, in the week prior to final exams, students with a limited theory procrastinate more and eat more unhealthy food. Students with a nonlimited theory are not similarly affected by high demands (Job et al., 2010, Study 4; Job et al., 2015). Further, a recent study showed that a limited theory about willpower is related to worse therapy adherence and psychological adjustment in a sample of patients with Type 2 Diabetes (Bernecker & Job, 2015).

How is it that a scale assessing beliefs about a narrow topic (i.e. strenuous mental activity) has such an effect on a broad range of outcomes, including clinically relevant areas like eating behavior or therapy adherence that do not appear to require much mental effort? Recent research indicates that when people believe that willpower is a limited resource they turn towards rest, recovery, and replenishment of resources once they exerted even a slight amount of mental effort (Job, Bernecker, Miketta, & Friese, 2015). Hence, implicit theories about willpower have an impact on how people respond to the accumulation of mental demands, e.g., when they have

to deal with high workload at work. In sum, ITW-M moderates the effect of mental demands on self-control outcomes. Accordingly, willpower theories are predictive of everyday self-control when students face high demands, such as a heavy course load (Job, Walton, et al., 2015). People who endorse a nonlimited belief about willpower are more resilient to the negative effect of high demands; when stressed by a heavy workload they do not start being careless in their eating behavior and they stay focused on their personal goals (Bernecker & Job, 2015). Therefore, they fare better with regard to a wide range of outcomes and report better psychological adjustment in the face of multiple challenges as compared to people who belief that willpower is a limited resource. In this way, future research on a person's implicit theory of their willpower for strenuous mental activities holds significant promise for future clinical and applied research and interventions.

To sum, research has only recently begun to explore the role of a person's theory of their willpower for strenuous mental activity in predicting important outcomes in academic, health-related, and clinical contexts. Future research and practice will aim at explaining different kinds of self-control failures (e.g., impulsive spending, aggression and violence) by willpower theories and, in the long term, at developing interventions and therapeutic techniques to promote a nonlimited theory about willpower in various clinical populations. The previous studies noted above provided initial evidence for the predictive validity of the ITW-M scale; however, one critique of this research is that the psychometric properties of the scale have not yet been scrutinized from a measurement invariance approach (Meredith, 1993; Reise, Widaman, & Pugh, 1993). If the ITW-M scale is to be useful to researchers across a variety of disciplines and cultural contexts, two important questions must be tested.

First, can the six items included in the ITW-M scale be invariantly modeled across existing samples and groups of interest? For future researchers who plan to use the ITW-M scale, it is important to know whether one can expect the latent structure of construct to be the same across the groups that they test. In a similar vein, future research addressing questions with regard to possible study-based, time-bound, cultural, or gender differences in ITW-M requires that ITW-M can be invariantly measured across groupings of interest.

Second, does a latent ITW-M factor display the expected patterns of convergence and divergence across existing samples? This question addresses issues of validity. Previous research suggests that ITW-M is related to several constructs including life-satisfaction and trait self-control. Further, if ITW-M is a distinct lay belief, then it should diverge from other implicit theories like implicit theories of intelligence.

Therefore, the present research is a first test of whether the ITW-M scale invariantly measures what it purports to measure, and whether it does so consistently across different samples within the same cultural context, across two cultural contexts (the U. S. and Switzerland / Germany), and across male and female participants. Affirmative findings to these questions would support researchers and clinicians who wish to include the ITW-M scale as a predictor, covariate, or outcome in a wide variety of questionnaire-based studies or diagnostics involving latent analyses.

Measurement invariance

We address these issues using a confirmatory factor analysis (CFA) based approach to measurement invariance. Because this approach is covered in great detail elsewhere (e.g., Little, 1997; Meredith, 1993; Millsap, 2011) we only briefly describe it here. A central assumption for tests of mean differences across groups is that the structure and measure of the latent factor of

interest does not significantly differ across samples, time points, or other data groupings. This assumption can be tested through measurement invariance analyses, which typically involve a series of multi-group CFAs with increasingly restrictive parameters. In this research, we adopt common names for each hierarchical step of this procedure: a first test is for *configural* (pattern) invariance; a second test is for *weak* (loading) invariance, and a third test is for *strong* (intercept) invariance. Configurally invariant data can be modeled with the same number of factors and with the same items associated with their respective factors (Meredith, 1993). Data that attain weak invariance additionally have the same pattern of loadings, and, for in strong invariant models, all indicator intercepts can be equated across groups.

In this research, we assess whether the ITW-M scale invariantly measures the same factor first across samples within one of two cultural contexts (i.e., across the 7 American samples, and across the 4 Swiss or German samples, respectively); next across men and women within a shared cultural context; and finally, *across* the American and Swiss/ German contexts. Such a finding would support future work focused on describing or explaining group-level differences in ITW-M across these groupings.

Convergent and Discriminant Validity

The second question for this research is whether the invariant factor measured by the ITW-M scale (should this be the case) has a pattern of relations with other constructs that is consistent with theoretical and empirical research on implicit theories of willpower. In other words, does the latent factor indexed by the ITW-M scale "behave" as it should, if it were in fact a latent construct for implicit theories of willpower for strenuous mental activities? We assess these tests across three theoretically and empirically relevant constructs.

Life satisfaction. Because a limited theory about willpower impairs everyday self-control and goal striving in everyday life, specifically when people face high demands, it can have far reaching consequences for a person's well-being. Supporting this assumption, recent research indicates that ITW-M is associated with life satisfaction (hereafter, LS). Specifically, participants with a limited theory reported lower LS than participants with a nonlimited theory (Bernecker et al., in press). Moreover, theories about willpower predicted a change in LS in situations where participants faced an increase in self-regulatory demands (i.e., during examination periods). Therefore, in the present research, we expected a significant negative correlation between ITW-M and LS.

Trait self-control. Trait self-control (hereafter, TSC) describes an individual's personal capacity to exert self-control and not to act impulsively across a variety of self-control contexts. It is typically assessed with self-report scales on impulsivity, like the Barratt Impulsivity Scale (e.g., "I act on impulse"; Barratt, Patton, & Stanford, 1975) or the Trait Self-Control scale by Tangney and colleagues (e.g., "I am good at resisting temptations"; Tangney, Baumeister, & Boone, 2004). Research suggests that there is substantial interindividual variance in people's self-control capacity and that TSC is related to a number of positive life outcomes such as psychological adjustment, eating behavior, drug abuse, and interpersonal relationships (Tangney et al., 2004).

Previous research involving self-control tasks in the laboratory and self-control in everyday life demonstrated a negative relationship between ITW-M and TSC. The more participants in these studies believed that willpower for mental activity is a limited resource the lower were their scores on the TSC scale (Bernecker et al., in press, Study 2; Job et al., 2015).

Therefore, in the present research, we expected that a more limited theory would be associated with lower levels of TSC.

  Implicit theories of intelligence. As a test of discriminant validity of the ITW-M scale, the present research examined its relationship with implicit theories of intelligence (hereafter, ToI). We suggest that ToI and ITW-M are distinct beliefs because they refer to distinct concepts (intelligence vs. self-control) and different time frames of change for these concepts. Whereas a person's theory of their intelligence addresses the possibility to grow intelligence with practice over larger periods of times (days to years), a theory about willpower refers to the depletability of self-control capacity in time ranges covering several-minutes to a few hours. Therefore, a person can think of intelligence as malleable, but may believe that willpower is limited and that one needs frequent rest to perform well. Hence, although ITW-M and ToI both predict positive outcomes in academic contexts (e.g., Blackwell et al; Dweck, 2012; Job et al., 2015; Molden & Dweck, 2006), we assume that the two belief-systems are largely independent. Accordingly, we expected no significant correlation between ITW-M and ToI.

The Present Research

  The analyses and results for the present research are organized into four sections: (a) Multigroup measurement invariance for ITW-M across samples *within* the U.S. and Swiss/German contexts, respectively; (b) Across-gender measurement invariance for ITW-M within cultural contexts; (c) Cross-cultural measurement invariance for ITW-M; and (d) Patterns of cross-culture latent correlations between ITW-M and LS, TSC, and ToI. As such, the first three sections provide information on the invariance of the ITW-M measure, while the final section provides initial information on the validity of the ITW-M measure.

In general, we hypothesize that ITW-M can be invariantly modeled across samples within the same cultural context, across men and women within the same cultural context, and across cultural contexts, and that the resulting factor is characterized by at least partial strong measurement invariance. We expect the same pattern of invariance for LS, TSC, and ToI (these analyses are presented in Appendix 3). Finally, we expect that across samples and cultural contexts: a more limited ITW-M should be associated with lower levels of LS (less satisfaction), a more limited ITW-M should be associated with lower levels of TSC (less self-control), and a more limited ITW-M should *not* be significantly associated with one's ToI.

Method

Participants and Procedure

This research uses data from 11 samples. Some of these samples have been used in published research on ITW-M, whereas other samples were collected for other purposes (e.g., unrelated research questions like leadership or everyday desires; pilot studies; Master's theses). We used samples that were available to us by February 2015 and used final, non-pilot language for all scales. All studies included in this research were approved by the local Institutional Review Board or Ethics Committee. In Table 1, we report the key demographic information and characteristics of each sample, as well as the means and standard deviations of the measures.

Measures

This research involves four sets of items used, in varying combinations, across 11 samples: the Implicit Theory of Willpower Scale for Strenuous Mental Activities Scale (ITW-M, Job et al. 2010), the Satisfaction with Life Scale (LS; Diener, Emmons, Larsen, & Griffin, 1985), the Trait Self Control Scale (TSC; Tagney, Baumeister, & Boone, 2004), and the Theories of

Intelligence scale (ToI; Dweck, 1999). The full text of each scale (in English and in German) is provided in Appendix 1.

Implicit Theory of Willpower for Strenuous Mental Activities Scale (ITW-M). The scale was developed in preparation for a series of laboratory experiments testing whether beliefs about willpower moderate the ego-depletion effect (Job et al., 2010). For that purpose, a set of 16 items were developed through adapting previous scales measuring other implicit theories (Dweck, 1999). Initial piloting of these items was completed by 95 participants recruited from a community-participant pool (50% female; M_{age} = 20.73, SD = 4.69). From this item pool, six items were selected that showed the highest factor loadings (factor loading > .78) in an exploratory factor analysis. The full scale was translated into German by two researchers from two different research groups who were fluent in both languages. They translated the scale into German independently and then compared the two versions.

Participants responded to the items using a 6-point scale (1 = *strongly agree;* 6 = *strongly disagree*). An example item was "After a strenuous mental activity, your energy is depleted and you must rest to get it refueled again." Item responses were recoded so that all items shared the same direction. Higher values for ITW-M item responses corresponded to a more-limited theory of willpower.

Satisfaction with Life Scale (LS). The scale, developed by Diener and colleagues (Diener et al., 1985), assesses participants' current satisfaction with life with five items (e.g., "In most ways my life is close to my ideal") using a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*). In the German speaking samples, we used the translation from Schumacher and colleagues (2003). Higher values for LS item responses corresponded with greater satisfaction with life.

Trait Self Control Scale (TSC). Levels of trait self-control were assessed using a 9-item version of the TSC scale (e.g. "I often act without reviewing all the alternatives," Tangney et al., 2004). In the German speaking samples, we used the translation from Bertrams and Dickhäuser (2009). Responses were made using a 5-point scale (1 = *Not at all*, 5 = *Very much*). Higher values for TSC item responses corresponded to poorer self-control.

Implicit Theories About Intelligence Scale (ToI). Implicit theories about intelligence were assessed with five items (e.g., "You have a certain amount of intelligence and you really cannot do much to change it" [reverse scored]; Dweck, 1999; German translation: Spinath, 2001). Participants rated how much they agree with each statement on a 6-point scale (1 = *Strongly agree*, 6 = *Strongly disagree*). Higher values for ToI item responses corresponded to a more-fixed theory about one's intelligence.

For our analyses, we used the standard items and scoring protocol for these constructs, with three exceptions. First, we recoded responses to the ITW-M and ToI scales, such that all responses shared the same direction (higher values corresponded to more-limited ITW-M or less-malleable ToI). The second exception involved the TSC items. To parsimoniously index a person's typical level of challenges with self-control, we used a subset of items (9 of 14) that indicated "poor" self-control. Third, we also used a subset of items (5 of 8) for ToI, because these were the items that were consistently included across samples.

Analyses

The analyses presented in this research are based in a multigroup CFA framework. Initial model fit was gauged by Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA) and McDonald's noncentrality index (1989; hereafter, Mc). Indications of good fit were CFI and TLI close to .95, Mc close to .9, and RMSEA close to

.06 (e.g., Hu & Bentler, 1999). All data were modeled in Mplus 7 (Muthén & Muthén, 1998-2012) using the MLR estimator, which is robust to data non-normality. Fit values for CFI and TLI were calculated in reference to an analysis-specific alternative null model, which provides more accurate assessments when compared to the default null model in Mplus (e.g., Widaman & Thompson, 2003). All models used the marker variable approach to identification, which uses the loading of the first indicator to serve as referent for the factor.

We tested for measurement invariance across samples using a CFA-based technique (e.g., Meredith, 1993; Widaman and Reise, 1997). In a first step, we test whether modeling an identical factor structure across groups passes the model fit criteria listed above. If this model does pass these criteria, it was termed configurally invariant. Testing for weak invariance involved equating loadings across groups. Testing for strong invariance involved additionally equating indicator intercepts across groups. We had two criteria for testing weak and strong invariance. Our first criterion was Δ CFI <.01, which is commonly used in invariance testing (Cheung & Rensvold, 2002). We also tested for invariance using the criterion of Δ Mc <.02 (Meade, Johnson, & Braddy, 2008). These criteria were selected instead of χ^2 difference testing because the former are less strongly affected by sample size and number of model constraints. We report comparative fit statistics to the third decimal place to ease model comparison. We used the standardized factor loadings from configurally invariant models to calculate the internal reliability of the factor of interest in terms of ω (e.g., McDonald, 1999).

<div align="center">

Results

</div>

ITW-M Measurement Invariance Across Samples Within Cultural Contexts

     ITW-M measurement invariance in American samples. The existing literature on ITW-M has measured the construct using a single, observed level scale score calculated as the

mean of the six ITW-M items (e.g., Job et al, 2010), consistent with the conceptualization of

ITW-M as a unidimensional construct (i.e., people have more-limited or less-limited implicit

theories of their willpower for strenuous mental activities). Therefore, we first attempted to

model a single ITW-M factor by the six ITW-M indicators across the seven American samples

using a seven-group multiple groups CFA. This initial model (1.A.1) fit poorly (CFI = .863;

TLI= .614, Mc = .826, RMSEA = .206; RMSEA 90% CI = [.189, .223]). Modification indices

showed significant correlations among the first three ITW-M indicators. These indicators are all

worded in terms of a limited ITW-M (i.e. using words such as "exhausts," "depleted," and

"recover"), which suggested a possible method factor.

　　　　Using the same indicators as model 1.A.1, we next modeled ITW-M using a single, six-

indicator factor and a second method factor indicated by items 1-3. The correlation between the

method factor and the ITW-M factor was not estimated. This model is conceptually similar to the

existing literature in that it assesses how well the data fit a single latent ITW-M factor, but

improves upon the prior analysis by accounting for error variance associated with reverse-coded

items. Model fit improved, however Sample 4 had several anomalous significant cross-loadings,

indicating that the structure of the ITW-M data for participants in this sample did not conform to

the pattern found in the six other samples. As a result, we removed this sample from analyses,

and tested the identical model as a six-group multiple groups CFA. The resulting model (1.A.2)

displayed good model fit and served as the basis for the remainder of the analyses Across the six

samples, internal reliability was high (ω = .86 – .95).[1] Fit information for the remaining analyses

in this section are presented in Table 2.

　　　　To test for weak invariance, we next equated factor loadings for the ITW-M factor across

groups. This model (1.A.3) fits well and passes both criteria for testing weak invariance. To test

for strong invariance, we next equated indicator intercepts across groups. This model (1.A.4) also passed both criteria for testing strong invariance.

ITW-M measurement invariance in CH/DE samples. Consistent with the American samples, we first attempted to fit a single-factor model for ITW-M for the four German-speaking samples. As in analysis 1.A.1, this initial model (1.B.1) fit poorly (CFI = .850; TLI= .600, Mc = .817, RMSEA = .212; RMSEA 90% CI = [.198, .225]). Modification indices suggested the same potential reverse-coded method factor was also present in these data.

Using the same structure as the American samples, we next tested whether a single ITW-M factor and method factor fit for the Swiss / German data. The model (1.B.2) fit well, and serves as a basis for the remainder of the analyses in this section. Across the four samples, internal reliability was high (ω = .87 - .91). To test for weak invariance, we next equated ITW-M factor loadings across groups. This model (1.B.3) fits well and passes both criteria for testing weak invariance. To test for strong invariance, we next equated ITW-M indicator intercepts across groups. This model (1.B.4) fits also well, and passes both criteria for testing strong invariance. These results indicate that responses to the ITW-M scale invariantly model the same factor across four samples of German-speaking Swiss and German participants. Unless reported in text, fit information for these analyses is presented in Table 2.

Summary. The results were largely consistent with the expectations. In the American context, ITW-M could be invariantly measured across 6 of 7 samples. In the Swiss/German context, ITW-M could be invariantly measured across all 4 samples. The final models for both cultural contexts were identical: a single six-indicator construct for ITW-M, and a negatively-worded language method factor for a subset of three indicators. These results provide support for

the internal reliability and measurement invariance of the ITW-M items in the American and Swiss/German contexts.

ITW-M Measurement Invariance Across Male and Female Participants

Our next step was to test for measurement invariance across men and women. Because potential gender differences may vary as a function of culture, we tested for invariance across genders and across samples within the American and within the Swiss/German cultural contexts separately. Consistent with the configurally invariant model from analysis 1.A.2, we specified a model (2.A.1) for American participants that included a single ITW-M factor and a method factor across 2 groups: men and women. Model 2.A.1 fit well, indicating configural invariance across men and women. To test for weak invariance, we next equated ITW-M factor loadings across men and women, using model 2.A.1 as a baseline. This model (2.A.2) also fit well, and passes both criteria for testing weak invariance. To test for strong invariance, we next equated all ITW-M indicator intercepts across the male and female groups. This model (2.A.3) also passes the invariance criteria.

We followed the same procedure for assessing invariance between men and women in the Swiss / German sample. Models 2.B.1, 2.B.2, and 2.B.3 fit well and passed the criteria for configural, weak, and strong invariance, respectively. Men did not significantly differ in their latent means for ITW-M compared to women in either the Swiss/German (latent $M = -.04$; $p = .39$) or American cultural context (latent $M = .04$; $p = .51$). Fit information for these analyses is presented in Table 3.

Summary. Across samples within both the American and Swiss/German cultural contexts, respectively, the ITW-M scale was characterized by strong invariance across men and women. These results provide support for future analyses that seek to describe or explain any

gender-related differences in ITW-M between men and women (based on experimental

manipulations, for example) within these cultural contexts.

ITW-M Measurement Invariance Across Cultural Contexts

Our next step was to test for invariance across cultural contexts. To do so, we specified a

two-group across-culture ITW-M model (3.1), modeling a single ITW-M factor and a method

factor. The first group included samples from the American context, while the second group

included participants from the Swiss/German context. Model 3.1 fit well, indicating configural

invariance across cultural contexts and serving as the basis for the remainder of the analyses in

this section. To test for weak invariance, we next equated ITW-M factor loadings across all

groups. This model (3.2) also fits well and passes both weak invariance criteria. To test for strong

invariance, we next equated all ITW-M indicator intercepts across groups. Although the Model

3.3 fit acceptably, it just exceeds the Δ CFI invariance criterion. Modification indices show that

model misfit derived from equating the intercepts for ITW-M item 6 across cultural contexts.

In a next step, we specified a model that was identical to Model 3.3, except that the

intercept for ITW-M item 6 as freely estimated across cultural contexts. This model allowed for

the testing of partial measurement invariance (e.g., Byrne, Shavelson, & Muthén, 1989; McArdle

& Cattel, 1994; Steenkamp & Baumgartner, 1998). This model (3.4) passes the both invariance

criteria for invariance testing, illustrating partial invariance across cultural context.[2] Fit

information for these analyses is presented in Table 3. Achieving partial strong invariance allows

for cautious interpretation of cultural differences in latent means (Byrne et al, 1989). For these

data, American participants reported a significantly more limited ITW-M for strenuous mental

activities (latent $M = -.34$; $p > .001$). We computed effect size based on Hancock's guidelines

(2001), which allow calculating standardized effect sizes comparable to Cohen's *d*. For this test, the effect size is small (*d* = .289).

Summary. The results from this section indicated that ITW-M can be measured with partial strong invariance across two cultural contexts. The intercept for ITW-M item 6 was not invariant across the U. S. and Swiss/German samples. Although "full" strong invariance was a preferable outcome, partial strong invariance (with five of six intercepts equated) still provides some ability to compare relations across groups (e.g., Steenkamp & Baumgartner, 1998). For these data, American participants reported a significantly more limited ITW-M than did Swiss or German participants.

Alternative models

An alternative approach to testing for invariance would be to additionally constrain item loadings for the method factor during the weak invariance test, and equate the reverse-coded method factor means across groups in the strong invariance test. Such findings would provide support that ITW-M observed score mean differences specifically reflect differences in the ITW-M factor mean, and would further justify interpreting across-group ITW-M differences. We provide details of these analyses in Appendix 2.

The results generally supported our hypotheses. The across-culture model achieved partial strong invariance, again by allowing the intercept of ITW-M item 6 to be freely estimated across cultures. The across-gender, within culture models were characterized by strong invariance. The U.S. samples were characterized by partial strong invariance (the intercept for ITW-M item 2 was freely estimated in Sample 2). The Swiss/German samples were characterized by strong invariance when allowing the latent mean of the method factor to be freely estimated. We interpret these results in the discussion.

Patterns of ITW-M Convergent and Discriminant Validity

The purpose of this final set of analyses is to determine whether the ITW-M displays the expected patterns of convergence and discriminant relations with factors for life satisfaction (LS), trait self-control (TSC) and theories of intelligence (ToI). All analyses were conducted in the following way. Using the partial strong invariant cross-cultural ITW-M model (3.4) as a base, we first freely estimated the correlations between ITW-M and the construct of interest for all samples with the relevant items. We additionally controlled for potential interindividual differences by regressing LS, TSC, and ToI on participant age and gender.[3] Model fit information for these tests is presented in Table 4. Latent correlation parameters are presented in Table 5.

For each construct of interest, we also assessed internal reliability and conducted invariance tests across groups and cultural contexts. For a report of these tests, please see Appendix 3. To briefly summarize these analyses, LS, TSC, and ToI were characterized by strong invariance across all groups and both cultural contexts.

Correlations between ITW-M and LS. Three samples (3, 5, and 8) measured both ITW-M and LS. Based on prior findings (Bernecker et al., in press; Bernecker & Job, 2015), we hypothesized that the ITW-M and LS factors should be negatively correlated, such that a more limited implicit theory of willpower should be related to lower life satisfaction. Using the strong invariant models for ITW-M and LS as a baseline, we estimated this correlation, and the resulting model (4.A) fit acceptably. Consistent with our hypothesis, ITW-M and LS were significantly negatively correlated in Samples 5 and 8 (rs = -.322 and -.417, respectively; see Table 5). However, this correlation was not significant in Sample 3. We interpret this finding in the general discussion section.

Correlations between ITW-M and TSC. Samples 2, 5, and 11 measured both ITW-M and TSC. Given prior research (Bernecker et al., in press; Job et al., 2015), we hypothesized that ITW-M and TSC would be positively correlated, such that a more limited willpower theory would be correlated with lower levels of trait self-control. Model 4.B freely estimated this correlation and fit acceptably well. As expected, across Samples 2, 5, and 11, a more limited ITW-M was correlated with lower levels of trait self-control ($r = .244$, .298 and .474, respectively; see Table 5). These results supported our hypotheses. Across three groups, a more limited willpower theory is consistently correlated with lower levels of trait self-control.

Correlations between ITW-M and ToI. Samples 1, 6, 9, and 11 measured both ITW-M and ToI. We hypothesized that the ITW-M and ToI factors would not be systematically correlated, indicating that the two factors do not "overlap." We tested this hypothesis using the same procedure as in Models 4.A and 4.B. The freed correlation model (4.C) fit acceptably. As expected, in three of the four samples (1, 6, and 11), there were no significant correlations between ITW-M and ToI. However, there was a significant correlation between ITW-M and ToI for Sample 9 ($r = .270$; see Table 5). We will also interpret this finding in the general discussion.

Summary. Findings were generally consistent with our hypotheses and provide initial support for the validity of the ITW-M scale. In 2 of the 3 Samples, a more limited ITW-M was correlated with lower levels of life satisfaction. Across 3 samples, a more limited ITW-M was correlated with lower levels of trait self-control. Finally, in 3 of 4 samples, ITW-M and a person's theories of intelligence were not significantly correlated.

Discussion

The aim of the present research was to test measurement invariance as well as discriminant and convergent validity of responses to a recently developed scale assessing

people's beliefs about willpower for strenuous mental tasks (the ITW-M scale). We expected that

participants' responses to the ITW-M scale were invariant across samples and cultural contexts,

and that the resulting latent factor displayed a pattern of relations consistent with theoretical and

empirical research on implicit theories of willpower. The target audience for this research

includes researchers and practitioners who want to use the ITW-M scale in questionnaire-based

studies involving latent analyses. Given the increasing interest in the study of self-control and

willpower-related constructs (Duckworth & Gross, 2014; Kotabe & Hofmann, 2015; Mann, de

Ridder, & Fujita, 2013), we believe that the reported analyses and findings will provide a useful

source and guideline to the adoption of the scale in research and practice.

In general, our expectations were supported. The results can be categorized into two

major categories. First, the ITW-M items, as specified by a single, six-item ITW-M factor and a

negative-worded method factor for a subset of items, displayed strong (or intercept) invariance

across samples within cultures and across men and women, and achieved partial strong

invariance across cultures. In addition, the resulting ITW-M latent factor was characterized by

strong internal reliability ($\omega = .86$-$.95$) across samples. Together, these findings provide support

for using these items to describe or explain group-level differences in ITW-M between adult men

and women and within and across the U.S. and Swiss / German cultural context. Given the

burgeoning interest in the study of self-control, we encourage others interested in assessing ITW-

M to adopt the modeling approach presented here.

One exception to the otherwise consistent measurement invariance warrants further

discussion. First is the cross-cultural intercept variability for ITW-M item 6. Participants in the

Swiss/German sample tended to agree less with this item as compared to participants in the

American sample. We believe that this is based in translation issues, rather than any systematic

cultural differences in the conceptualization of ITW-M between American and Swiss / German samples. Our interpretation is that the German expression "voller Kraft" might have been perceived as describing a stronger state of "power" as compared to the English "energized". Future work could test if adapting the language of ITW-M item 6 (and the similarly-worded item 5) provides strong invariance across U.S. and Swiss / German samples, and additionally whether this variation in interpretation exists across other cultures. However, regardless this limitation or pending future research, partial strong invariance (with 5 of 6 indicators) across 10 samples and 2,550 participants still provides the ability to cautiously compare the latent means for ITW-M (as well as other latent relations) across groups (e.g., Byrne, Shavelson, & Muthén, 1989; Reise, Widaman, & Pugh, 1993).

Despite the relatively weak effect, future work should explore the robustness of and the potential processes underlying the finding that American participants reported a significantly more-limited implicit theory of willpower for strenuous mental activities. Given the earlier reported negative effects of such an implicit theory, this represents a potentially-fruitful avenue for theory, research, and application.

The second major category of findings illustrated that the ITW-M factor generally displayed the expected pattern of convergent and discriminant validity with several constructs (which, in separate analyses, were each found to be strong invariant across cultures). In two of three samples, a more-limited implicit theory of willpower was associated with lower levels of self-reported trait self-control and life-satisfaction. This finding is consistent with other research that used observational-level data. The previous research, however, suggests that willpower theories predict well-being depending on changing levels of demands (Bernecker et al., in press). This could explain why the correlation between ITW-M and life satisfaction did not reach

significance in Study 3. The sample of Study 3 consisted of university students who completed

the questionnaire at a relatively-relaxed time period at the beginning of the term (Job et al., 2010,

Study 4). We assume that the relationship would become more pronounced when demands

accumulate and students with a limited theory perceive that they do not make as much progress

on their personal goals as they expected.

Finally, we found that in two of three studies ITW-M was not significantly associated

with implicit theories of intelligence, suggesting that the two constructs do not "overlap" in a

systematic way. We did not expect to find a strong overlap between the two constructs because

they depict people's theories about two distinct phenomena (intelligence vs. willpower) and with

regard to different dimensions (malleability in the long-term vs. short term depletability).

However, in one of the three examined studies, the relationship between ITW-M and theory of

intelligence was significant. This result suggests that there might be conditions under which the

two constructs converge, with "limited theorists" having more of a fixed theory about

intelligence. Future research has to investigate whether this convergence is caused by a third

variable (e.g., social background; study characteristics) or whether the two might be related in

some heretofore-unexpected conceptual way.

A Note Regarding Reverse-Coded Item Method Factors

Before discussing limitations, we will briefly describe the decision to model the ITW-M

data using a reverse-coded item method factor. There is quite some literature describing the pros

and cons of creating scales of items with reversed or negated items (for an informative review,

see Weijters & Baumgartner, 2012). To briefly sum, when used properly, reversed and negated

items can improve the quality of data by minimizing acquiescence bias (Greenleaf, 1992;

Baumgartner & Steenkamp, 2001) and disrupting non-substantive participant response patterns.

However, such items often introduce shared error variance that results in model misfit for latent analyses of "simple" hypothesized constructs (e.g., the single-factor Models 1.A.1 and 1.B.1 reported here).

The existing theoretical and empirical literature on ITW-M describes the construct unidimensionally (Job et al., 2010). That is, people have theories of willpower that vary in the degree to which they are "limited;" those with low levels are described as "non-limited," whereas those with high levels are described as having a "limited" theory of their willpower. A central aim for this research was to assess if ITW-M, as it is conceptualized in the existing literature, can be invariantly modeled across several samples and two cultures. We believe that modeling ITW-M using a reverse-coded method factor is the most appropriate technique given the existing unidimensional conceptualization of ITW-M and the patterns of shared variance among reverse-coded items. Although future theoretical and empirical work may elect to explore this possibility (that is, that what we have modeled here as method variance is instead a construct-relevant second dimension of ITW-M), we believe that attempting to model separate "limited" and "non-limited" factors with these data would likely result in the specification of non-informative "artifactors" (e.g., Marsh, 1996).

Limitations

Although the results largely supported our hypotheses, a first limitation involves the alternative modeling approach, which equated method factor parameters across groups. Using this approach, the Swiss/German across-samples model achieved strong invariance when freeing the latent mean of the method factor across samples. All other analytical groupings achieved strong or partial strong invariance when equating the method factor's latent mean. Therefore, we suggest that researchers specifically interested in assessing sample-by-sample differences in

observed level ITW-M means of Swiss or German samples do so with caution. The

heterogeneous study designs and location of the ITW-M items within these studies may affect the

way participants respond to reverse-coded items. Future work should systematically investigate

which sample- (e.g., age) or study-specific characteristics (e.g., experimental vs. observed

design) may underlie this potential variation.

A second limitation was the lack of overlap between ITW-M and the LS, TSC, and ToI

across all samples. We only used existing data for this research, and these data were not collected

for the purposes of testing the ITW-M scale's latent relations. This limitation extends to the lack

of demographic variable overlap across studies reported here; future work will benefit from

understanding how a diverse array of demographic characteristics may or may not affect a

person's ITW-M. Broadly, a beneficial next step would involve systemically assessing these (and

other) factors across a program of research. Generalizing an invariant latent correlation between

ITW-M and LS for people in a Western cultural context, for instance, may not yet be advisable.

We encourage researchers to systematically assess the relations between ITW-M and other

factors across samples and cultural contexts.

A third limitation was the cross-sectional nature of these data. These data were not

sufficient to address the enticing question of whether a person's implicit theories of willpower

vary developmentally across the life span. Future work should assess whether ITW-M is

developmentally invariant by collecting data from multiple, developmentally-meaningful

measurement points. A final limitation involved Sample 4, which did not pass the configural

invariance step, and thus did not share the same latent pattern for ITW-M with any other sample.

This finding was unexpected, and further work should explore the possible sources of variation

(e.g., in the sample, design, or other questionnaire items) in this data that may contribute to a different factor structure.

Conclusion

Self-control failures are at the heart of a wide range of behavioral problems that can have severe personal consequences and cause high societal expenses (e.g., overeating, crime and violence, overspending; Baumeister & Heatherton, 1996; Gottfredson & Hirschi, 1990; Vohs & Faber, 2007). The research on implicit theories about willpower suggests that people's beliefs about the nature of their self-control capacity play a crucial role in explaining these self-control failures. Future clinical research and practice might be interested in explaining and promoting better self-control by targeting people's beliefs about willpower. Therefore, researchers and practitioners alike will need to rely on a reliable and validated measure. The analyses presented her support the use of the ITW-M scale and provide guidelines on how to treat it when modeling latent factors.

Footnotes

1. Some recent work (e.g., Rodriguez, Riese, & Haviland, 2016) has suggested reporting

   reliability for bifactor models using ω_h (see Zinbarg, Revelle, Yovel, & Li, 2005). For the

   American samples, ITW-M ω_h ranged from .65 – .86. For the Swiss German samples,

   ITW-M ω_h ranged from .73 – .80. Using the two-group CFA approach reported in Model

   3.1, American and Swiss German ITW-M ω_h were .84 and .76, respectively.

2. Freeing the intercepts of ITW-M item 6 is sufficient to achieve partial strong invariance.

   However, modification indices also suggested some model misfit deriving from equating

   the intercepts of ITW-M item 5. This item shares some language with ITW-M item 6

   (both include the words "energized" or "voller Kraft"). Thus, an alternative way for

   testing for partial strong measurement invariance would be to freely estimate the

   intercepts of ITW-M items 5 and 6 across cultures. This model fits well (CFI = .994;

   TLI= .987, Mc= .994, RMSEA = .038; RMSEA 90% CI = [.026, .051]) and also achieves

   partial strong invariance. See the discussion section for further elaboration and potential

   future directions.

3. We also tested whether participant age was associated with ITW-M across all studies.

   Participant age was not a significant predictor of ITW-M in any model.

References

Barratt, E. S., Patton, J., & Stanford, M. (1975). *Barratt Impulsiveness Scale*. Barratt-Psychiatry

Medical Branch, University of Texas.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the

active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252–65.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview.

Psychological Inquiry, 7, 1-15.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control.

Current Directions in Psychological Science, *16*, 351–355.

Bernecker, K. (2015). Implicit theories about willpower and their consequences for achievement,

health, and well-being. (Doctoral Dissertation). Retrieved from Zurich Open Repository and

Archive database.

Bernecker, K., Herrmann, M., Brandstätter, V., & Job, V. (in press). Implicit theories about

willpower predict subjective well-being. *Journal of Personality.*

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of Intelligence

Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an

Intervention. *Child Development, 78,* 246-263.

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets

matter: a meta-analytic review of implicit theories and self-regulation. *Psychological

Bulletin*, *139*(3), 655–701.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor

covariance and mean structures: The issue of partial measurement invariance. *Psychological

Bulletin, 105,* 456–466.

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

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: the role of mindsets in determining the stress response. *Journal of Personality and Social Psychology, 104*(4), 716–33.

Duckworth, a., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science, 23,* 319–325.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Dweck, C. S. (1996). Implicit theories as organizers of goals and behavior. In Gollwitzer, P. M., & Bargh, J. A.: *The Psychology of Action: Linking Cognition and Motivation to Behavior*. New York, NY: The Guilford Press.

Dweck, C. S. (1999). *Self-Theories*. Lillington, NC: Taylor & Francis.

Dweck, C. S. (2012). Mindsets and Human Nature: Promoting Change in the Middle East, the Schoolyard, the Racial Divide, and Willpower. *American Psychologist, 67,* 614-622.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role on judgements and reactions: a world from two perspectives. *Psychological Inquiry, 6,* 267–285.

Dweck, C. S., yi Hong, Y., & yue Chiu, C. (1993). Implicit theories individual differences in the likelihood and meaning of dispositional inference. *Personality and Social Psychology Bulletin, 19*, 644-656.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49,* 74-75.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford, CA: Stanford University Press.

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*(4), 495–525.

Hancock, G. R. (2001). Effect size, power, and sample size determinants for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66, 378-388.

Heller, S. (2013). Der Einfluss von transformationaler Führung auf Burnout von Mitarbeitern - die Rolle von impliziten Theorien über Willenskraft, beruflicher Selbstwirksamkeits- erwartung und wahrgenommener Arbeitsbelastung [The effect of transformation leadership on burnout – the role of implicit theories about willpower, self-efficacy, and perceived job demands]. University of Zurich: unpublished master thesis.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M.-S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*, 588–599.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion--is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21,* 1686–93.

Job, V., Walton, G. M., Bernecker, K., & Dweck, C. S. (2015). Implicit theories about willpower

    predict self-regulation and grades in everyday life. *Journal of Personality and Social*

    Psychology, 108, 637–647.

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control.

    Perspectives on Psychological Science, 10, 618–638.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data:

    Practical and theoretical issues. *Multivariate Behavioral Research, 32,* 53–76.

Mann, T., de Ridder, D., & Fujita, K. (2013). Self-regulation of health behavior: Social

    psychological approaches to goal setting and goal striving. *Health Psychology, 32,* 487–98.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful

    distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810.

McArdle, J. J., & Canell, R. B. (1994). Structural equation models of factorial invariance in

    parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral*

    Research, 29, 61-101.

McDonald, R. P. (1989). An index of goodness-of-fit based on non-centrality. *Journal of*

    Classifiation, *6*, 97-103.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum Associates.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit

    indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance.

    Psychometrika, 58, 525–543.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York: Routledge.

Molden, D. C., & Dweck, C. S. (2006). Finding "Meaning" in Psychology: A Lay Theories

Approach to Self-regulation, Social Perception, and Social Development. *American Psychologist, 61,* 192-203.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide*. *Seventh Edition*. Los Angeles, CA: Muthén & Muthén

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 223–237. http://doi.org/10.1080/00223891.2015.1089249

Schumacher, J., Klaiberg, A., & Brähler, E. (2003). Diagnostische Verfahren zu Lebensqualität und Wohlbefinden [*Assessment of life quality and well-being*]. Göttingen, Germany: Hogrefe.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78–90.

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,* 271–324.

Vohs, K. D., & Faber, R. J. (2007). Spent resources: Self-regulatory resources availability affects impulse buying. *Journal of Consumer Research, 33,* 537-547.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*, 82-96.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys:

A review. *Journal of Marketing Research*, *49*(5), 737–747.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological

instruments: Applications in the substance use domain. In K. Bryant, M. Windle, & S. West

(Eds.), *The science of prevention; Methodological advances from alcohol and substance*

abuse research (pp. 281-324). Washington, DC: American Psychological Association.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit

indices in structural equation modeling. *Psychological Methods, 8* (1), 16-37.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and

Mcdonald's ωH: their relations with each other and two alternative conceptualizations of

reliability. *Psychometrika*, *70*, 123–133. http://doi.org/10.1007/s11336-003-0974-7

Table 1

Sample Demographics, Sampling Information, Relevant Citations, and Observed-level Descriptive Data

S	Country	N	% ♀	Age M (SD)	Sampling	Citation	ITW-M M (SD)	LS M (SD)	TSC M (SD)	ToI M (SD)
1	U.S.	66	65.2	20.67 (2.67)	Student participant pool	Job et al. (2010), Study 1	4.13 (0.84)	-	-	4.08 (0.95)
2	U.S.	176	57.4	21.21 (2.62)	Student participant pool	Job et al. (2015)	3.88 (0.88)	-	4.04	-
3	U.S.	101	74.5	24.67 (9.81)	Student participant pool	Job et al. (2010), Study 4	3.67 (0.92)	4.35 (1.60)		
4	U.S.	240	62.9	43.83 (14.84)	Amazon Mechanical Turk (MTurk)	Unpublished Data	3.75 (1.09)	-	4.62 (0.83)	-
5	U.S.	160	43.8	35.19 (11.27)	MTurk	Unpublished Data	3.86 (0.99)	4.23 (1.54)	3.93 (1.05)	-
6	U.S.	222	52.3	34.51 (10.66)	MTurk	Unpublished Data	4.23 (1.11)	-	-	3.80 (1.36)
7	U.S.	179	51.1	35.09 (11.79)	MTurk	Unpublished Data	4.09 (1.08)	-	-	
8	CH	269	63.9	39.24 (10.33)	Online forums	Bernecker et al. (in press), Study 1	3.78 (0.87)	3.37 (1.36)	-	-
9	CH	247	81.0	22.01 (5.34)	Psychology 1st year students	Bernecker (2015)	4.04 (0.79)	-	-	
10	DE	871	60.7	37.61 (11.16)	Online forums, Networks (e.g., XING)	Heller (2013)	3.36 (0.91)	-	-	-
11	CH	259	77.3	21.51 (5.14)	Psychology 1st year students	Unpublished Data	3.87 (0.81)	-	3.19 (0.60)	3.75 (1.17)

Note. Sample 4 was not used in final analyses. U.S. = American sample; CH = Swiss sample; DE = German sample.

ITW-M = Implicit theory of willpower for strenuous mental activities; responses ranged from 1 (strongly agree) to 6 (strongly disagree); higher values indicate a more-limited theory of willpower for strenuous mental activities

LS = Satisfaction with life; responses ranged from 1 (strongly disagree) to 7 (strongly agree); higher values indicate greater satisfaction with life

TSC = Trait self-control; responses ranged from 1 (not at all) to 5 (very much); higher values indicate poorer self-control

ToI = Implicit theories of intelligence; responses ranged from 1 (strongly agree) to 6 (strongly disagree); higher values indicate a more-fixed implicit theory of intelligence

Table 2

Implicit Theory of Willpower for Strenuous Mental Activities Invariance Testing Fit Statistics for Final Multigroup Models Assessing Invariance Across Samples, within Cultural Contexts

Model	Test	Context	$\chi^{2\dagger}$	df	RMSEA	RMSEA 90% CI	Mc	ΔMc	CFI	ΔCFI	TLI
1.A.2	Configural	U.S.	42.985	36	.036	.000, .071	.996		.997		.988
1.A.3	Weak	U.S.	67.166	61	.026	.000, .057	.997	.001	.997	<.001	.994
1.A.4	Strong	U.S.	105.530	80	.046	.015, .068	.986	.011	.989	.008	.980
1.B.2	Configural	CH/DE	35.864	24	.035	.000, .057	.996		.997		.989
1.B.3	Weak	CH/DE	57.908	39	.034	.012, .052	.994	.002	.996	.001	.989
1.B.4	Strong	CH/DE	93.155	50	.046	.031, .060	.987	.007	.990	.006	.981

Note. Invariance criteria: Δ CFI <.01 and Δ Mc <.02 per invariance step. U.S. = American samples (1-3/5-7), CH/DE= Swiss and German samples (8-11)

[†] In MPlus, the χ^2 test statistic computing using the MLR estimator (used in all models in this research) is asymptotically equivalent to the Yuan-Bentler T2* test statistic.

Table 3

Implicit Theory of Willpower for Strenuous Mental Activities Invariance Testing Fit Statistics for Final Multigroup Models Assessing Invariance Across Genders (within Cultural Contexts), and Assessing Invariance Across Cultural Contexts

Model	Test	Context	Description	χ^2	df	RMSEA	RMSEA 90% CI	Mc	ΔMc	CFI	ΔCFI	TLI
2.A.1	Configural	U.S.	Across gender	26.202	12	.051	.024, .078	.992		.993		.975
2.A.2 [a]	Weak	U.S.	Across gender	28.283	17	.038	.007, .062	.994	.002	.994	.001	.986
2.A.3	Strong	U.S.	Across gender	32.652	21	.035	.000, .057	.994	.000	.994	<.001	.988
2.B.1 [b]	Configural	CH/DE	Across gender	14.476	12	.016	.000, .035	.999		.999		.998
2.B.2	Weak	CH/DE	Across gender	18.959	17	.012	.000, .035	.999	.000	.999	<.001	.999
2.B.3	Strong	CH/DE	Across gender	30.743	21	.024	.000, .041	.997	.002	.997	.002	.995
3.1 [c]	Configural	all	Across culture	21.652	12	.026	.005, .043	.998		.998		.994
3.2	Weak	all	Across culture	49.042	17	.039	.027, .052	.993	.005	.994	.004	.986
3.3	Strong	all	Across culture	116.352	20	.063	.052, .074	.981	.012	.983	.011	.964
3.4 [d]	Par. strong	all	Across culture	64.742	19	.044	.033, .056	.991	.002	.991	.003	.982

Note. Invariance criteria: Δ CFI <.01 and Δ Mc <.02 per invariance step**.** U.S. = American samples (1-3/5-7), CH/DE= Swiss and German samples (8-11)

[a-b] Models 2.A.1 – 2.B.3 were two-group CFAs. Participants were grouped by gender
[c] Models 3.1-3.4 were two group CFAs. Participants were grouped by cultural context
[d] In model 3.4, the intercept for item 6 was not equated across cultural contexts

Table 4

Fit Information for Multigroup Models with Implicit Theory of Willpower for Strenuous Mental Activities and Constructs of Interest

Model	Description	χ^2	*df*	RMSEA	RMSEA 90% CI	Mc	CFI	TLI
4.A.1	ITW-M and LS	306.657	214	.051	.037, .063	.912	.969	.958
4.B.1	ITW-M and TSC	212.5347	140	.051	.036, .064	.941	.975	.963
4.C.1	ITW-M and ToI	393.501	270	.047	.036, .057	.928	.976	.966

Note. LS = Satisfaction with Life (Samples 3, 5, & 8); TSC = Trait Self-control (Samples 2, 5, & 11); ToI = Theories of Intelligence (Samples 1, 6, 9, 11)

Table 5

Latent Correlations Between Implicit Theory of Willpower for Strenuous Mental Activities and Constructs of Interest

Model	Description	Sample 1	Sample 2	Sample 3	Sample 5	Sample 6	Sample 8	Sample 9	Sample 11
4.A	ITW-M / LS			-.130	-.322***		-.417***		
4.B	ITW-M / TSC		.244*		.298**				.474***
4.C	ITW-M / ToI	.149				.119		.270**	.002

Note. LS = Satisfaction with Life (Samples 3, 5, & 8); TSC = Trait Self-control (Samples 2, 5, & 11); ToI = Theories of Intelligence (Samples 1, 6, 9, 11)

* p < .05; ** p < .01; *** p < .001

Appendix 1

Item Text

1. Implicit Theories About Willpower for Strenuous Mental Activities Scale (ITW-M)

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion - is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, *21*, 1686–1693.

1. Strenuous mental activity exhausts your resources, which you need to refuel afterwards (e.g. through breaks, doing nothing, watching television, eating….).
2. After a strenuous mental activity, your energy is depleted and you must rest to get it refueled again.
3. When you have completed a strenuous mental activity, you cannot start another activity immediately with the same concentration because you have to recover your mental energy again.
4. Your mental stamina fuels itself. Even after strenuous mental exertion, you can continue doing more of it. (r)
5. When you have been working on a strenuous mental task, you feel energized and you are able to immediately start with another demanding activity. (r)
6. After a strenuous mental activity, you feel energized for further challenging activities. (r)

1 = strongly agree; 6 = strongly disagree

German translation:

Job, V., Bernecker, K., Miketta, S., & Friese, M. (2015). Implicit theories about willpower predict the activation of a rest goal following self-control exertion. *Journal of Personality and Social Psychology, 109,* 694-706.

1. Anstrengende mentale Tätigkeiten (z.B. Nachdenken über eine schwierige Fragestellung, sich stark auf Etwas konzentrieren) erschöpfen meine Kraftressourcen, welche ich im Anschluss wieder auftanken muss (z.B. durch Pausen, Nichtstun, Fernsehen oder Essen).
2. Nach einer anstrengenden mentalen Tätigkeit ist meine Energie erschöpft und ich muss mich erholen, um sie wieder aufzutanken.
3. Nach einer anspruchsvollen geistigen Tätigkeit kann ich nicht mit derselben Konzentration eine neue Tätigkeit angehen, weil ich mich zuerst erholen muss.
4. Meine Willenskraft ist unerschöpflich. Auch nachdem ich mich eine Zeit lang auf Etwas konzentriert habe, kann ich mit etwas Anspruchsvollem fortfahren. (r)
5. Wenn ich an einer anstrengenden mentalen Aufgabe gearbeitet habe, fühle ich mich voller Kraft und bin im Stande, sofort eine neue herausfordernde Aufgabe anzupacken. (r)
6. Nach einer anstrengenden mentalen Tätigkeit, fühle ich mich voller Kraft für weitere herausfordernde Aktivitäten. (r)

1 = stimme ausgesprochen zu; 6 = stimme überhaupt nicht zu

2. Satisfaction with Life Scale

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49,* 74-75.

1.   In most ways my life is close to my ideal.
2.   The conditions of my life are excellent.
3.   I am satisfied with my life.
4.   So far I have gotten the important things I want in life.
5.   If I could live my life over, I would change almost nothing.

1 = strongly disagree; 7 = strongly agree

German translation:

Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the satisfaction with life scale (SWLS). Psychometric properties, validity, and population-based norms. *European Journal of Psychological Assessment, 27,* 127–132.

1.   In den meisten Bereichen entspricht mein Leben meinen Idealvorstellungen.
2.   Meine Lebensbedingungen sind ausgezeichnet.
3.   Ich bin mit meinem Leben zufrieden.
4.   Bisher habe ich die wesentlichen Dinge erreicht, die ich mir für mein Leben wünsche.
5.   Wenn ich mein Leben noch einmal leben könnte, würde ich kaum etwas ändern.

1 = stimme genau zu; 7 = stimme überhaupt nicht zu

3. Trait Self-Control (TSC)

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*(2), 271–324.

1. I have a hard time breaking bad habits.
2. I am lazy.
3. I say inappropriate things.
4. I do certain things that are bad for me, if they are fun.
5. I wish I had more self-discipline.
6. Pleasure and fun sometimes keep me from getting work done.
7. I have trouble concentrating.
8. Sometimes I can't stop myself from doing something, even if I know it is wrong.
9. I often act without reviewing all the alternatives.

1 = not at all; 5 = very much

German translation:

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität: Eine deutsche Adaptation der Kurzform der Self-Control Scale (SCS-K-D). *Diagnostica*, *55*, 2–10.

1. Es fällt mir schwer schlechte Gewohnheiten abzulegen.
2. Ich bin faul.
3. Ich sage unangemessene Dinge.
4. Ich tue manchmal Dinge, die schlecht für mich sind, wenn sie mir Spass machen.
5. Ich wünschte ich hätte mehr Selbstdisziplin.
6. Angenehme Aktivitäten und Vergnügen hindern mich manchmal daran meine Arbeit zu machen.
7. Es fällt mir schwer mich zu konzentrieren.
8. Manchmal kann ich mich selbst nicht daran hindern, etwas zu tun, obwohl ich weiss, dass es falsch ist.
9. Ich handle oft ohne die Alternativen durchdacht zu haben.

1 = völlig unzutreffend; 5 = trifft genau zu

4. Implicit Theories of Intelligence (ToI)

Dweck, C. S. (1999). *Self-Theories*. Lillington, NC: Taylor & Francis.

1. You have a certain amount of intelligence, and you can't really do much to change it.
2. Your intelligence is something about you that you can't change very much.
3. No matter who you are, you can significantly change your intelligence level. (r)
4. To be honest, you can't really change how intelligent you are.
5. You can always substantially change how intelligent you are. (r)
6. You can learn new things, but you can't really change your basic intelligence.
7. No matter how much intelligence you have, you can always change it quite a bit. (r)
8. You can change even your basic intelligence level considerably. (r)

1 = strongly agree; 6 = strongly disagree

German translation:

Spinath, B. (2001). *Implizite Theorien über die Veränderbarkeit von Intelligenz und Begabung als Bedingungen von Motivation und Leistung [Implicit theories about the malleability of intelligence and abilities as conditions for motivation and achievement]*. Lengerich: Pabst.

1. Man hat ein bestimmtes Maß an Intelligenz und kann nicht viel tun, um es zu verändern.
2. Die Intelligenz einer Person ist etwas, was sie nicht verändern kann.
3. Man kann zwar neue Dinge lernen, seine grundlegende Intelligenz kann man jedoch nicht verändern.
4. Egal wer man ist, man kann seine eigene Intelligenz verändern. (r)
5. Man kann immer die eigene Intelligenz wesentlich verändern. (r)

1 = stimme ausgesprochen zu; 6 = stimme überhaupt nicht zu

Appendix 2: Results for Alternative Models of Implicit Theory of Willpower for Strenuous Mental Activities Invariance

Table A.1

Implicit Theory of Willpower for Strenuous Mental Activities Invariance Testing Fit Statistics for Alternative Models

Model	Test	Context	Description	χ^2	df	RMSEA	RMSEA 90% CI	Mc	ΔMc	CFI	ΔCFI	TLI
A.1	Configural	All	Across culture	21.652	12	.026	.005, .043	.998		.998		.994
A.2	Weak	All	Across culture	55.767	19	.040	.028, .052	.993	.005	.994	.004	.986
A.3	Par. Strong	All	Across culture	111.510	23	.056	.046, .067	.982	.011	.984	<.010	.971
B.1	Configural	U.S.	Across gender	26.202	12	.051	.024, .078	.991		.993		.975
B.2	Weak	U.S.	Across gender	29.786	19	.035	.000, .059	.993	.002	.995	.002	.988
B.3	Strong	U.S.	Across gender	34.429	24	.031	.000, .053	.994	.001	.995	<.001	.991
C.1	Configural	CH/DE	Across gender	14.476	12	.016	.000, .041	.999		.999		.998
C.2	Weak	CH/DE	Across gender	19.814	19	.007	.000, .032	1.000	.001	.999	<.001	.999
C.3	Strong	CH/DE	Across gender	51.177	24	.037	.023, .051	.992	.008	.992	.007	.987
D.1	Configural	U.S.	Across sample	42.985	36	.036	.000, .071	.996		.997		.988
D.2	Weak	U.S.	Across sample	78.528	71	.027	.000, .056	.995	.001	.997	<.001	.983
D.3	Strong	U.S.	Across sample	133.510	95	.052	.029, .071	.976	.019	.984	.013	.975
D.4	Par. Strong	U.S.	Across sample	123.142	94	.045	.018, .066	.991	.004	.988	.009	.981
E.1	Configural	CH/DE	Across sample	35.864	24	.035	.000, .057	.996		.997		.989
E.2	Weak	CH/DE	Across sample	71.286	45	.038	.020, .054	.992	.004	.994	.003	.987
E.3	Strong	CH/DE	Across sample	248.799	59	.088	.077, .100	.944	.048	.957	.037	.930
E.4	Strong, free latent means	CH/DE	Across sample	118.331	56	.052	.039, .065	.981	.009	.986	.008	.976

Note. Invariance criteria: Δ CFI <.01 and Δ Mc <.02 per invariance step. U.S. = American samples (1-3/5-7), CH/DE= Swiss and German samples (8-11)

In weak invariance step, reverse-coded method factor loadings are equated. In strong invariance step, reverse-coded method factor means are equated, except for Model E.4. Model A.3 is consistent with Model 3.4 in Table 3. In Model D.4, the intercept for item ITW-M 2 is freed in Sample 2.

Appendix 3: Measurement invariance for other constructs of interest across American and Swiss/German samples

Life Satisfaction (LS) measurement invariance across American and Swiss/German participants

We tested for configural invariance of LS by specifying a single five-item LS factor, using data from Samples 3, 5, and 8. This initial model (F.1) had good fit, and serves as the basis for invariance testing. Across all samples, internal reliability was high (ω = .91 – .93). To test for weak invariance, we next equated LS factor loadings across groups, using model 4.A.1 as a baseline. This model (F.2) also fits well, and passes both criteria for testing weak invariance. To test for strong invariance, we next equated LS indicator intercepts across the three samples. This model (F.3) passes the invariance criteria. Therefore, in three samples across both cultural contexts, LS is characterized by strong invariance. The fit statistics for all analyses in this appendix can be found in Table A.2.

TSC measurement invariance across American and Swiss/German participants

We parceled (Little, Willingham, Shahar, & Widaman, 2002) the 9 indicators of TSC into three arbitrary parcels across Samples 2, 5, and 11. Consistent with the positions of Little and colleagues (Little, Rhemtulla, Gibson, & Schoemann, 2013) our decision to parcel was informed by attempts at model parsimony, as well as the TSC items' widespread use (e.g., de Ridder, Lensvelt-Mulders, Finkenauer, Stok, & Baumeister, 2012). Using these three parcels as indicators results in a saturated model (G.1) from which we base the subsequent invariance tests. Across the four samples, internal reliability ranged from acceptable to good (ω = .71 [Sample 11] — .90). To test for weak invariance, we next equated the loadings of the parcels across samples. This model (G.2) passes both criteria for testing weak invariance. To test for strong invariance, we equated TSC parcel intercepts across the three samples. This model (G.3) passes invariance test

criteria. The TSC factor is characterized by strong invariance across three samples and two cultural contexts.

Theories of Intelligence (ToI) measurement invariance across American and Swiss/German participants

The ToI items are written similarly to the ITW-M items; 3 of 5 items are worded "negatively." Therefore, we replicated the basic structure of models 1.A.2 and 1.B.2 to test for configural invariance. Across 4 samples (1, 6, 9, and 11), the 5 indicators loaded onto a single ToI construct, and 3 of these indicators also loaded onto a method factor (which did not correlate with the ToI factor). This initial model (H.1) fit acceptably, and provides the basis for testing for weak and strong invariance. Across the three samples, internal reliability was high ($\omega = .89 – .96$; $\omega_h = .78 – .94$). To test for weak invariance, we next equated ToI factor loadings across the three samples. This model (H.2) passed both criteria for testing weak invariance. To test for strong invariance, we next equated ToI indicator intercepts across the three samples. This model (H.3) passed the invariance criteria, thus displaying strong invariance across four samples and both cultural contexts.

Results Summary

We tested for measurement invariance for three constructs that will be used to test for patterns of ITW-M convergent and discriminant validity. In these analyses, LS, TSC, and ToI displayed strong invariance across all samples and both cultural contexts.

Appendix 3 References

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

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151–173.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300. http://doi.org/10.1037/a0033266

Table A.2

Satisfaction with Life, Trait Self-control and Theories of Intelligence Invariance Testing Fit Statistics

Model	Test	Construct	χ^2	df	RMSEA	RMSEA 90% CI	Mc	ΔMc	CFI	ΔCFI	TLI
F.1	Configural	LS	12.861	15	.000	.000, .061	1.000		1.000		1.000
F.2	Weak	LS	19.644	23	.000	.000, .051	1.000	<.001	1.000	<.001	1.000
F.3	Strong	LS	42.768	31	.046	.000, .078	.989	.011	.991	.009	.985
G.1	Configural	TSC	**	**	**	**	**		**		**
G.2	Weak	TSC	8.020	4	.077	.000, .154	.996		.995		.972
G.3	Strong	TSC	17.415	8	.083	.027, .127	.991	.005	.988	.007	.967
H.1	Configural	ToI	12.995	8	.056	.000, .109	.997		.998		.978
H.2	Weak	ToI	35.583	20	.062	.026, .095	.990	.007	.992	.006	.972
H.3	Strong	ToI	43.768	28	.053	.017, .082	.990	<.001	.992	.004	.980

Note. LS = Satisfaction with Life (Samples 3, 5, & 8); TSC = Trait Self-control (Samples 2, 5, & 11); ToI = Theories of Intelligence (Samples 1, 6, 9, 11)

** = saturated model