Spring 2012

Using longitudinal mean and covariance structures (LMACS) analysis to assess construct stability over two time points: An example with psychological entitlement

Bozhidar Mihaylov Bashkov
James Madison University

Using Longitudinal Mean and Covariance Structures (LMACS) Analysis to Assess Construct Stability over Two Time Points: An Example with Psychological Entitlement

Bozhidar M. Bashkov

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Psychological Sciences

May 2012

Acknowledgements

I would like to acknowledge several people who have supported me in completing this project.

First, I would like to thank my family and friends for providing me with all the moral support a graduate student could ask for. I am especially grateful to my mother, who gave me strength and courage in the hardest of times.

Second, I would like to say a big thank you to my thesis committee members, who put aside many hours of their time to contribute to my project. Their expertise, ideas, and feedback have greatly enriched the content of this document and have allowed me to learn a lot in the process.

Last but not least, I would like to express my deepest gratitude to my academic advisor, Dr. Sara J. Finney, for all of her guidance, support, and contribution to this work. I honestly could not ask for a better mentor, nor can I think of a more dedicated faculty member that I have met in my life. Thank you so much!

Table of Contents

List of Tables

List of Figures

Abstract

The current paper reviewed two widely used approaches to assessing construct stability over two time points (rank-order and mean-level consistency), as well as common misconceptions about what each indicates. In addition, the application of longitudinal mean and covariance structures (LMACS) analysis as a modern approach to assessing construct stability was explained and demonstrated by assessing the stability of psychological entitlement over 1.5 years measured via the Psychological Entitlement Scale (PES).

Confirmatory factor analysis supported a one-factor solution for the PES at both time points, and reliability of scores was adequate (ω = .88 and .89). Full configural and metric invariance and partial scalar invariance were established for the PES. Rank-order consistency of factor scores was moderate (r = .61) and the latent mean difference in psychological entitlement across time was not statistically significant.

Results provided construct validity evidence for the PES regarding measurement invariance and also indicated that psychological entitlement tended to be stable on average but not at the individual level over 1.5 years. Discussion of the effects of differential item functioning (DIF) of scalar non-invariant items on mean difference testing and use of the PES with observed scores, the change in psychological entitlement at the individual level, and the advantages of LMACS analysis as a unified approach to assessing construct stability is also provided.

Chapter I

Introduction

Assessing the Stability of a Construct over Two Time Points

Definition of longitudinal stability. A fair amount of psychological research has focused on studying psychological constructs over time. Developmental psychologists, for example, often measure constructs multiple times throughout a person's life, from early childhood, through adolescence, to adulthood. One of the reasons researchers are interested in studying constructs over time is to assess how constructs of interest change and evolve (Chan, 1998). That is, researchers are often interested in whether scores representing a construct remain stable over time (McCrae, Terracciano, & Khoury, 2007). Longitudinal or temporal stability of a construct is essentially the lack of change in the construct over time.

How is longitudinal stability typically evaluated? The evaluation of longitudinal stability usually involves a repeated measures design, whereby data representing the construct of interest are collected from the same sample of participants at least twice, allowing a certain amount of time to lapse in between the data collection occasions (Chan, 1998). The amount of time between data collection occasions varies and is determined by the researchers based on the construct under study. With a paired set of scores from just two time points, researchers often evaluate longitudinal stability in one of two ways.

One method used to assess longitudinal stability is to compute the Pearson product moment correlation coefficient between scores from Time 1 and Time 2. The correlation coefficient representing the relationship between scores collected at two time points is often called test-retest consistency and indicates the stability of the rank order of scores across time (i.e., rank-order consistency). That is, a high test-retest consistency coefficient would indicate that respondents remained in the same position relative to others from Time 1 to Time 2, whereas a low test-retest consistency coefficient would indicate that respondents changed their relative position in the rank order of scores across the two time points.

Another method of assessing stability is to compare the magnitude of the observed score means at Time 1 and Time 2 via a dependent samples (repeated measures) *t*-test. If the difference between means from the two data collection occasions is not significantly different from zero, researchers often conclude that there was no (mean-level) change in the construct from Time 1 to Time 2.

Typically, one of these two methods is used to evaluate stability. That is, either rank-order consistency or mean-level change is used to make inferences as to whether a construct is stable over two time points. Unfortunately, simply assessing rank order-consistency or mean-level change to infer construct stability can be misleading, in that the information necessary to make such an inference is incomplete (McCrae et al., 2007). Below, I review the common misconceptions associated with the use of these methods to assess stability.

Misconceptions about the assessment of construct stability. There are three common misconceptions with respect to the evaluation of a construct's stability over two time points. First, a high rank-order consistency coefficient is erroneously interpreted as sufficient evidence for the argument that a construct is stable over time. A test-retest consistency coefficient is simply an indicator of whether the rank order of respondents' scores on the construct is consistent from one testing occasion to the next (i.e., those scoring *relatively* high at Time 1 also score *relatively* high at Time 2, and those scoring *relatively* low at Time 1 also score *relatively* low at Time 2). Consistency in the rank order of scores does not imply that the mean level of the construct remains stable over time. For example, it is possible that respondents' scores increased over time (i.e., level of the construct increased), but did so by approximately the same amount (i.e., high rank-order consistency). Likewise, it is possible that respondents' scores decreased over time (i.e., level of the construct decreased), but did so by approximately the same amount (i.e., high rank-order consistency).

A second misconception regarding the assessment of the stability of a construct over time is that the lack of mean change from one testing occasion to the next implies no individual change. This misconception is equivalent to erroneously stating that lack of mean change across time also reflects high rank-order consistency. Recall that rank-order consistency indicates that respondents remained in relatively the same position compared to others from one testing occasion to the next. Such consistency does not indicate the absence or presence of *average* change in the construct across time. On the other hand, mean-level change on the construct indicates change in respondents' scores from Time 1 to Time 2 *on average*, but provides no information as to how *individuals* changed across time.

To better illustrate the distinction between rank-order consistency and mean-level consistency, consider the following four scenarios (see Figure 1) when crossing rank-order consistency results (yes vs. no) with mean consistency results (yes vs. no).

1) There is rank-order consistency and mean consistency across time. Thus, respondents preserved their relative position compared to others across time because all respondents' scores on the construct remained the same across time.

2) There is rank-order consistency but not mean consistency. Respondents' relative position remained the same; those relatively high at Time 1 were relatively high at Time 2. Moreover, on average, scores increased or decreased across time, which would be reflected in a mean-level difference across time. However, the average change does not imply that all respondents increased or decreased the same amount across time; this would only be true if the consistency in rank order was 1.0. For example, if the average change across time was 5 units, and the rank-order consistency was 1.0, this would indicate that every respondent increased by 5 units across time. As the consistency coefficient decreases from 1, this indicates that individuals change more or less than each other across time.

3) There is no rank-order consistency, but there is mean consistency. That is, respondents changed their relative position in the rank order of scores, but because some respondents increased and others decreased, there is no change *on average*.

4) There is no consistency in either the rank order or average across time. This would be the case if individuals changed over time (e.g., some increased, whereas others decreased; all increased; or all decreased) but by different amounts; thus, resulting in change *on average* (i.e., no mean level consistency), in addition to individual change (i.e., no rank-order consistency). Recall that a mean difference indicates change in the construct *on average*—it does not reflect the type of *individual* change that occurred as

respondents scores were "reshuffled" at Time 2 (as indicated by the lack of rank-order consistency).

Despite the clear distinction between rank-order consistency and mean-level consistency, researchers often report either of these two very different indicators as evidence for the stability of a construct over two time points. The four possible scenarios highlight the importance of examining and correctly interpreting *both* rank-order consistency and mean-level change when assessing stability (Chan, 2003).

Clearly, it is important to avoid misconceptions about rank-order and mean-level change. However, the third misconception is more fundamental and potentially has a more serious impact on the validity of inferences about construct stability. Many researchers erroneously assume the *measurement* of a construct remains stable or invariant over time (Chan, 1998; 2003). When tested, this assumption is not always empirically supported. Moreover, in order to accurately interpret mean-level or rank-order change, one must rule out change in the measurement of the construct over time. Below I review how these three types of stability (e.g., stability in measurement, stability in the average level of a construct, and stability in rank order) have been conceptualized in the measurement literature and how to assess each empirically when modeling scores from two time points. It is important to note that empirically supporting any one type of change (i.e., change in measurement, change in average level, or change in rank order) is not necessarily better or worse than other types of change. Instead, these are merely different types of change, all of which are worthy of interpretation (Chan, 2003).

Types of Stability and Contemporary Methods to Assess Them

There are three indicators of the stability (or lack of change) of a construct over two time points: 1) stability of factor structure (no gamma change) and stability of measurement parameters (no beta change), 2) stability of average latent scores, which is represented by no mean change in the construct across time (no alpha change), and 3) stability of rank order, which is represented by a high correlation of factor scores across time. Each of these indicators of longitudinal stability will be defined and the importance of each when empirically evaluating stability across two time points will be explained.

Factor structure and measurement parameter stability. Factor structure stability is the extent to which the factor structure underlying a measure's scores (i.e., the number and form of the factors) does not change from one administration of the measure to the next. Measurement parameter stability is the degree to which factor pattern coefficients and intercepts remain the same or invariant from one administration of the measure to the next. Both factor structure and measurement parameter invariance must be satisfied before one can evaluate the other two indicators of longitudinal stability of a construct: no latent mean change in the construct and high rank-order consistency.

A popular framework for organizing some of the indicators of longitudinal stability is the conceptualization of gamma, beta, and alpha (ΓBA) change (Golembiewski, Billingsley, & Yeager, 1976), where gamma and beta change pertain to factor structure and measurement parameter instability, and alpha change refers to latent mean change in the construct of interest across time. A contemporary method of testing longitudinal measurement invariance and latent mean differences involves a multi-step process known as longitudinal mean and covariance structures (LMACS) analysis. LMACS analysis is considered a structural equation modeling technique (Vandenberg &

Lance, 2000). More specifically, LMACS analysis is considered an extension of longitudinal factor analysis (Chan, 2003). This LMACS analysis approach will be presented alongside the gamma, beta, alpha change framework given that this invariance testing process simply explicates the specific models that are tested when assessing ΓBA change.

Gamma change refers to an inconsistency in respondents' understanding of the construct over time. Gamma change has been defined as "a redefinition or reconceptualization of some domain, a major change in the perspective or frame of reference within which phenomena are perceived and classified, in what is taken to be relevant in some slice of reality" (Golembiewski et al., 1976, p. 135). For example, an employee completing a communication skills inventory before participating in a training aimed at fostering those very skills may perceive communication skills as a unidimensional construct, but when completing the same measure after the training, s/he may conceptualize the construct as multidimensional (e.g., interpersonal skills, writing skills). Empirically, gamma change results in a different number of factors or items representing different factors at different time points (Riordan, Richardson, Schaffer, & Vandenberg, 2001). In a LMACS analysis, gamma change can be tested by specifying and estimating the same factor structure at each time point. If there is no change in the number of factors or in the items serving as indicators for different factors at the different time points, *configural invariance* is established (Millsap & Meredith, 2007). Given configural invariance, one can next examine if the values representing the relationship between the factor and the indicator (e.g., factor pattern coefficients) are equivalent over time (Vandenberg & Lance, 2000). However, if there is a change in the number of factors or items are representing different factors at different time points, then one cannot proceed to this next stage of assessing longitudinal stability because one would be comparing conceptually different constructs (Chan, 2003). Thus, it is crucial that there is no gamma change (i.e., configural invariance is established across time) in order to assess beta change.

Beta change occurs when the respondents' definition and conceptualization of the construct being measured is the same across testing occasions (i.e., no gamma change), but respondents use the response scale differently across time. Specifically, respondents use the scale one way initially, but then they "shrink" or "stretch" the scale during subsequent administrations of the measure, thus recalibrating their use of the scale (Golembiewski et al., 1976). For instance, beta change would occur if a respondent's understanding and level of the latent construct entitlement did not change across time, yet the respondent chose different response categories when responding to the same items representing entitlement at two time points. Thus, the respondent's observed item scores change across occasions, even though his or her actual latent level on the construct remains the same. In LMACS framework, such inconsistency is assessed via metric and scalar invariance tests (Vandenberg & Lance, 2000). *Metric invariance* requires that the values of factor pattern coefficients (i.e., factor loadings) be of the same magnitude across testing occasions, which implies the items have equivalent saliency to the factor across time (Millsap & Meredith, 2007). In other words, establishing metric invariance implies that an item is equally representative of the underlying construct at each time point. At least partial metric invariance needs to be established before one could proceed with subsequent invariance testing (Vandenberg & Lance, 2000). *Scalar invariance* requires that the values of intercepts (i.e., the observed item score when the latent variable underlying the item has a value of zero) of corresponding items be of the same magnitude across time, which means that given an equal representation of the construct by the items across time points (i.e., metric and scalar invariance), responses to the items will be equally indicative of the construct (factor) they represent (Millsap & Meredith, 2007). For example, given configural, metric, and scalar invariance, a person whose latent level of entitlement truly did not change across two time points should provide the same observed responses at Time 1 and Time 2. By contrast, if s/he actually increased or decreased on entitlement (change at the latent level), s/he should respond differently to the items across time. Thus, establishing both metric and scalar invariance implies that *observed* mean change is reflecting *latent* mean change. That is, the absence of beta change allows the researcher to make valid inferences about latent mean stability of the construct, based on presence or absence of changes in observed scores over time (Riordan et al., 2001).

Latent mean stability. Given no gamma or beta change, one can assess alpha change. Alpha change "involves a variation in the level of some existential state, given a constantly calibrated measuring instrument related to a constant conceptual domain" (Golembiewski et al., 1967, p. 134). That is, the instrument measures the construct equally well from one administration to the next, yet respondents' scores vary across testing occasions (mean change in observed scores across time is present). In the case of alpha change, a change in a respondent's observed score across time is inferred to reflect a change in the amount of the underlying construct across time. Thus, by definition, alpha change requires the absence of both beta and gamma change (Riordan et al., 2001).

Essentially, alpha change is what researchers are typically interested in assessing, and often this is assessed using observed variable techniques such as repeated measures ANOVA to estimate the observed mean difference in scores across time. Fortunately, in a structural equation modeling framework, latent mean stability can be assessed via LMACS analysis by estimating the *latent* mean difference or the difference between means at the construct level across time (Chan, 1998). Thus, LMACS analysis allows for the simultaneous estimation of latent means alongside the parameters analyzed for measurement invariance (factor structure, factor pattern coefficients, and intercepts). Moreover, modeling variables at the latent level accounts for measurement error in the indicators, and therefore estimates of mean difference over time will be more accurate than those produced by observed variable techniques which assume no measurement error (DeShon, 1998; Hancock, 1997). See Chan (2003) and Ployhart and Oswald (2004) for a more detailed review of the advantages of latent variable techniques over traditional observed variable methods in longitudinal research.

Latent rank-order consistency. Another indicator of the stability of a construct is high consistency of scores across time. That is, the degree to which the rank ordering of respondents on the construct is preserved from one time point to another can be estimated by computing a stability coefficient (i.e., the correlation of respondents' scores across two time points). Rank-order consistency should only be estimated after establishing both configural and metric invariance (scalar invariance is not necessary because estimates of consistency do not involve means). Researchers often report and interpret the Pearson product moment correlation coefficient, which is based on observed data. Similar to the difference tests that focus on change in observed scores described above, the correlation coefficient obtained from observed variables will be biased because of measurement error. Alternatively, estimating the correlation between the latent variable (factor) at time 1 and time 2 will result in a more accurate estimate of rank-order consistency.

Example of Assessing Construct Stability: Psychological Entitlement

Understanding the process of assessing longitudinal invariance and factor stability is best facilitated by an example. Below I demonstrate how to apply LMACS analysis to assess the stability of psychological entitlement over two time points, 1.5 years apart. Psychological entitlement is an especially suitable construct, given the current interest in the construct, the need to gather validity evidence for a popular existing measure of entitlement, and most importantly, the lack of empirical support behind the claims that psychological entitlement is stable over time.

Definition of psychological entitlement and its importance. An entitled individual is someone who has “unreasonable expectations of especially favorable treatment or automatic compliance with his or her expectations” (DSM-IV-TR, 2000). With respect to stability, a popular definition of psychological entitlement is “a stable and pervasive sense that one deserves more and is entitled to more than others” (W. K. Campbell, Bonacci, Shelton, Exline, & Bushman, 2004, p. 31). This definition suggests that psychological entitlement is both stable and non-context-specific. Indeed, one can encounter psychologically entitled people in many contexts. For example, children may express a strong feeling of entitlement to Christmas presents, new toys, or candy. Teenagers, on the other hand, may feel entitled to privacy or attention. Young adults entering the work force may feel entitled to a promotion and vacation time during the first few months on the job. These are but a few examples of situations where one can witness a clash of interests due to the perception of “due” favorable treatment by psychologically entitled individuals.

Previous research has linked entitlement to a host of maladaptive attitudes, affects, and behaviors. To name a few, psychological entitlement has been positively related to aggression, acquisitiveness, selfish behaviors in relationships (W. K. Campbell et al., 2004), job-related frustration and coworker abuse (Harvey & Harris, 2010), and negatively related to humility, focus on others (Elliott, 2010), agreeableness, and emotional stability (W. K. Campbell et al., 2004). Clearly, psychological entitlement is a construct worthy of investigation. Moreover, considering the multitude of undesirable attitudes, affects, and behaviors that psychological entitlement has been related to, it is important to empirically test the claim of stability. That is, if psychological entitlement is stable over time, then entitled individuals will most likely remain entitled throughout their lives, which negatively impacts other people. By contrast, if psychological entitlement is malleable (i.e., not stable), experiences that lower levels of psychological entitlement could be sought.

Evidence regarding the stability of entitlement. Authors of the Psychological Entitlement Scale (PES), currently the most popular measure of psychological entitlement, claim that psychological entitlement is “a chronic or stable disposition rather than a response to a specific social situation” (W. K. Campbell et al., 2004, p. 35). Interestingly, the authors of the PES provide no theoretical rationale as to why psychological entitlement is stable. In support of their claim of the stability of entitlement over time, W. K. Campbell and colleagues (2004) conducted two studies, in which the PES was administered to participants twice. In one study, the two test administrations were 1 month apart, and produced a test-retest correlation coefficient of .72. In the other study, the PES administrations were 2 months apart, and the test-retest stability coefficient was .70. Based on this evidence, W. K. Campbell and colleagues (2004) concluded that psychological entitlement is indeed stable over time. No other longitudinal research has been conducted to support or challenge the authors' conclusion. Thus, the empirical evidence supporting the stability of psychological entitlement is limited at best.

Not only is the evidence gathered by W. K. Campbell and colleagues limited, but the interpretation of these test-retest correlation coefficients may be invalid. Recall that in order to assess rank order-consistency, certain assumptions must be met to ensure that no conceptual or measurement inconsistencies affect the results; configural and metric invariance must be established. W. K. Campbell and colleagues did not test these assumptions. Moreover, in order to fully assess the stability of a construct, test-retest stability should be coupled with an assessment of mean-level change. Although W. K. Campbell and colleagues reported observed means at Time 1 and Time 2, they did not interpret or discuss observed mean differences. Of course, the computation and interpretation of mean change requires that the researchers establish configural, metric, and scalar invariance; these invariance tests were not conducted. Considering the insufficiency of evidence regarding the stability of psychological entitlement, further empirical investigation is needed prior to making claims about construct stability.

The Current Study

Purpose. The purpose the current study was to demonstrate the assessment of construct stability by examining psychological entitlement over the course of approximately 1.5 years. The multi-step process of LMACS analysis was employed, as these steps align with the assessment of stability, which are reflected in the three research questions below.

Research question 1: Is the unidimensional model of entitlement championed by the creators of the PES empirically supported at two time points? To answer this question, confirmatory factor analyses were conducted on PES scores collected at Time 1 and 1.5 years later at Time 2. Given the same factor model fit the PES scores at both time points, configural invariance was supported, and the analyses proceeded to the next stage (Figure 2).

Research question 2: Does the PES function equivalently across time? Given a comparable factor structure at both time points, I tested for equal factor pattern coefficient values across time (metric invariance) and equal item intercepts across time (scalar invariance). Based on these results, full metric and partial scalar invariance were established, and the analyses proceeded to the final stage (Figure 3).

Research question 3: Given longitudinal measurement invariance of the PES, does psychological entitlement increase or decrease over 1.5 years? Because only partial scalar invariance was established, a test of the latent mean difference in entitlement from Time 1 to Time 2 was conducted using LMACS analysis both with and without scalar non-invariant items contributing to the latent mean to determine whether entitlement changes on average and to assess the effect of scalar non-invariance. Moreover, using LMACS analysis, the correlation between the entitlement factors across time was examined to assess whether the rank order of respondents remained consistent across time.

Chapter II

Review of the Literature of Psychological Entitlement

The term “entitlement” was first used in the early 1800’s in the context of legal affairs, pertaining to one’s rights to social benefits as regulated by law or social norms. This original meaning has been fully preserved to this day, as we often hear people talk about entitlement programs (e.g., Social Security, Medicare). What has occurred since that time is the emergence of a different, yet related term—“psychological entitlement”.

The term “psychological entitlement” emerged fairly recently (early 90s), but the concept dates back to 1916, when Freud first labeled a group of patients, who felt they were special, “the exceptions” (Freud, 1916). Psychological entitlement is different from the conceptualization of entitlement as regulated by law or social norms in that psychological entitlement represents a sense or expectation that one should receive favors and/or special treatment from others that entitled individuals deem reasonable and justified, when in fact they aren’t. According to Freud, these patients must have been wronged in the past, and thus they expect special treatment.

Jacobson (1959) agreed with Freud’s theory that these patients must have been wronged in the past, but he also added that individuals like Freud’s patients perceived themselves as “blessed” with unusual beauty and talent, and thus, were not to be expected to adhere to rules like everyone else. Horney (1950) labeled these expectations of special treatment “neurotic claims” and stated that for these patients “a wish or need, in itself quite understandable, turns into a claim” (p. 42). In order to claim something, one must have a somewhat legitimate reason to make that claim. A problem arises with psychologically entitled individuals, however, because they deem their claims or expectations justified or legitimate, when, in fact, they aren't. For example, a student may feel entitled to graduate from college simply because s/he was admitted to the university and paid tuition. Such an expectation is certainly not justified. In comparison, an example of justified entitlement is public education in the U.S. Every American citizen is entitled to public education, as regulated by law. As such, justified entitlement is typically *inclusive*—justified entitlement pertains to everyone in a group defined by law or social norms (e.g., social security). Unjustified or psychological entitlement, on the other hand, tends to be *exclusive*—psychologically entitled individuals believe they stand out from the group, they are special, and therefore they should be treated preferentially.

Because entitled individuals are not isolated from others, these unjustified expectations affect others. In the context of distributive justice, or justice as it pertains to the allocation of goods in society, such expectations may lead to a situation of inequity. According to equity theory (Adams, 1965), people experience inequity when they perceive that the ratio of their outcomes (e.g., salary, benefits) to inputs (e.g., effort, skills) and the ratio of others' outcomes to inputs are unequal. Research on equity sensitivity empirically categorized people according to their level of tolerance to inequity: a) benevolent, people who prefer their outcome/income ratios to be lower than those of others, b) equity sensitive, people who prefer their outcome/income ratios to be the same as those of others, and c) entitled, people who experience distress when their outcome/income ratios are not higher than those of others (Huseman, Hatfield, & Miles, 1987; Miles, Hatfield, & Huseman, 1989). The last category defines individuals who are psychologically entitled, because their expectations and claims of a greater portion of the pie are not justified. Having established the distinction between justified (e.g., entitlement programs such as social security) and unjustified (psychological) entitlement, the origins and initial study of psychological entitlement will be reviewed next, since it is the construct under investigation in the current study.

The Emergence of Psychological Entitlement as a Construct

Psychological entitlement as a component of narcissism. Feelings of being "exceptional", "special", and "unique" as exhibited by psychologically entitled individuals, also appear to be symptoms of a personality disorder that was given much attention by clinical psychologists and psychoanalytic practitioners in the last century—the narcissistic personality disorder (NPD). The interest in NPD granted the disorder its inclusion in the third edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III; American Psychiatric Association, 1980). The DSM-III (1980) describes a narcissistic individual as one displaying "characteristic disturbances in interpersonal relationships, such as feelings of entitlement, interpersonal exploitativeness, relationships that alternate between the extremes of overidealization and devaluation, and lack of empathy" (p. 315). According to this definition, psychological entitlement can be considered a component of narcissism. This conceptualization is further supported by certain criteria in the DSM-III (1980) for diagnosing someone with NPD, such as "grandiose sense of self-importance and uniqueness, e.g., exaggeration of achievements and talents" and "preoccupation with fantasies of unlimited success, power, brilliance, beauty, or ideal love" (p. 317). These criteria align with Freud's categorizing of patients expecting special treatment as the "exceptions" (Freud, 1916), as well as Jacobson's addition of the self-perceived possession of unseen beauty and talent by such individuals (Jacobson, 1959). Whether it is because narcissists believe they are simply "the exception" or worth more than others, they typically feel entitled to things: "Entitlement, the expectation of special favors without assuming reciprocal responsibilities, is usually present" (DSM-III, 1980, p. 316). This clinical definition of psychological entitlement is consistent with Huseman, Hatfield, and Miles' (1987) description of entitled people, as expecting more than they contribute. As more research was conducted on narcissism and entitlement, entitlement grew into an area of interest of its own, rather than simply as a component of narcissism.

Psychological entitlement as an independent construct worthy of study. Over the last couple of decades, interest in psychological entitlement has increased. This rising interest is reflected in both the popular press and psychological research. A Lexis/Nexis™ (2010) search of the term "sense of entitlement" returned 831 results in major newspapers and magazines between the years 1990 and 2000, compared to over 3000 results between 2000 and 2010. The increased focus on entitlement has resulted in a better understanding of the construct. For example, although entitlement is still listed as a criterion for the diagnosis of narcissistic personality disorder in the most current version of the DSM (DSM-IV-TR), entitlement's definition has changed. Formerly, entitlement was defined as "the expectation of special favors without assuming reciprocal responsibilities" (DSM-III, 1980). Currently, an entitled individual is defined as someone who has "unreasonable expectations of especially favorable treatment or automatic compliance with his or her expectations" (DSM-IV-TR, 2000). The current definition of entitlement is different from the former definition in that the current definition amplifies the scope of expectations inherent in entitled individuals. The simple idea of expecting special favors without reciprocating has grown into a set of beliefs that are “unreasonable” and involve “especially favorable treatment” and “automatic compliance”. As such, the current definition of entitlement reflects a deeper understanding of entitlement among researchers, and more importantly, one that portrays entitlement more negatively than before.

In addition to this revised definition of entitlement, conceptualizations of both general psychological entitlement and context-specific entitlement have emerged. A construct that is considered *general* is not specific to any domain or situation, but is rather universal, manifesting itself in a wide range of settings, virtually independent of the situation. By contrast, a construct that is considered *context-specific* is heavily dependent on the situation and is only typical of the domain to which it was specified. One will likely encounter people with high levels of entitlement in a wide variety of contexts. At the workplace, for instance, an employee may feel entitled to extra vacation days or approval of late arrivals to work. Such attitudes and behaviors are rarely tolerated, and can often lead to interpersonal tension or even conflict with supervisors and coworkers. Another example is academic entitlement or “a sense of entitlement specific to education” (Kopp, Zinn, Finney & Jurich, 2011). An entitled student may expect a passing grade in the course simply because s/he attended class throughout the semester. These are two *different* types of context- specific entitlement. That is, think about a student who is working while attending college. The working student may feel entitled in one context, but not the other. For example, the student may be academically entitled, but not feel entitled at the workplace, or vice versa.

Alternatively, a person may feel entitled across a variety of contexts, thus exemplifying the aforementioned general, non-context-specific entitlement. General attitudes and affects are lasting beliefs, convictions, or viewpoints that are independent of the context or the situation at hand. A person's level on a latent variable that is non-context-specific should remain fairly similar in magnitude (i.e., intensity) and direction (e.g., positive or negative) *across* situations or contexts.

It is important to study entitlement with respect to specificity. That is, empirical studies of a construct at different levels of measurement specificity allow the researcher to examine the extent to which the construct is general, context-specific, or both (Baranik, Barron, & Finney, 2010). With respect to psychological entitlement, research employing measures of both general and context-specific entitlement is very limited (Chowning & N. J. Campbell, 2009; Greenberger, Lessard, Chen, & Farruggia, 2008; Kopp, Zinn, Finney, & Jurich, 2011). Nonetheless, in all three papers, general entitlement and context-specific (e.g., academic) entitlement are moderately correlated. Therefore, it is important that researchers continue to study general psychological entitlement in addition to context-specific entitlement.

Justified Entitlement vs. Unjustified Entitlement vs. Deservingness

It is important to clarify the differences between three related concepts—justified entitlement, unjustified (psychological) entitlement, and deservingness. Above, the distinction between justified and unjustified entitlement was established. However, the introduction of the concept "deservingness" often causes confusion among all three terms, and thus requires some clarification.

In everyday language, the terms “entitlement” and “deservingness” are often used interchangeably. For example, one may say “I deserve a raise” or “I am entitled to a raise”. Both statements convey one’s expectation of a higher salary in the future. However, there are conceptual differences as to whether one is said to be “deserving of” something, or to be “entitled to” something. Moreover, recall that entitlement can be justified (e.g., social security) or unjustified (e.g., psychological entitlement). A good approach to highlighting the distinction between justified entitlement, deservingness, and unjustified (psychological) entitlement is asking *why*. Why should I expect a raise? If I expect a raise because I have worked extra hard, and have increased sales by 20%, then I may truly *deserve* a raise. Alternatively, if I expect a raise because I have been with the company for two years and such is the company’s rule (to increase salary after two years of employment), then my entitlement may be *justified*. However, if I expect a raise because I feel that I am inherently worth more than other employees, or because I am innately special, then I am likely exhibiting a sense of *unjustified* or *psychological entitlement*.

There is a fair amount of research to support the first of the two distinctions—justified entitlement vs. deservingness. Feather (2003) proposed that judgments of deservingness relate to outcomes that are “earned or achieved as products of a person’s actions”, whereas judgments of entitlement are guided by “agreed-upon body of law, social norms, and formal and informal rules”. Defined this way, Feather’s conceptualization of entitlement aligns with the justified entitlement presented above. Following this framework of defining entitlement and deservingness, it appears that the term “deservingness” is to be used when someone is personally responsible for the outcome, whereas the term “entitlement” (whether justified or not) applies to situations that are independent of one’s actions. For example, an employee may feel that s/he should receive a bonus at the end of the year for two different reasons: a) because s/he put forth a great amount of effort to produce high-quality work and demonstrated performance on the job beyond expectations, or b) company policy grants every employee with an end-of-year bonus calculated as a percentage of an employee’s wages, regardless of work performance. In the first case, the employee is said to be deserving of the bonus because of his or her actions (i.e., producing high-quality work, exceeding expectations). In the second case, the employee is said to be entitled to the bonus because the company is supposed to give a bonus to every employee, independent of the quantity and quality of work done. In other words, the entitled employee perceives that the positive outcome (i.e., the bonus) is not so much his or her own responsibility, but rather an expected outcome based on the company’s policy (i.e., end-of-year bonus for all).

To test the conceptual distinction between justified entitlement and deservingness empirically, Feather (2003) conducted two studies to examine whether participants were actually able to distinguish between entitlement and deservingness. In one of the studies, participants were given a scenario of a student running for election in a national student organization, where the student exerted high vs. low effort, was eligible or not eligible to run for election by virtue of age, and was either elected or not elected (Feather, 2003). In the other study, participants were presented with a scenario, where a person suffering from an illness was to decide how to divide his will among his son, nephew, or friend, who provided him with help or limited help. In both studies participants’ reactions to the respective scenario were evaluated by asking participants to complete ratings of justified entitlement and deservingness. Results from both studies showed that participants were able to distinguish between justified entitlement and deservingness by giving high ratings of justified entitlement when the outcome involved an external network of rights, norms, and social norms (e.g., age eligibility), and giving high ratings of deservingness when the outcome involved personal responsibility (e.g., actions such as effort).

The distinction between deservingness and unjustified (psychological) entitlement is equally important, yet it has not been supported by any research. What is more, the distinction between these two concepts should be even more apparent than the distinction between deservingness and justified entitlement that was just presented. As discussed above, deservingness is associated with personal responsibility and merit. Someone who feels deserving of a reward actually did the work to deserve the reward. By contrast, unjustified or psychological entitlement has nothing to do with personal responsibility or merit—it is entirely based on one's self-concept of being inherently special, precious, and exceptional. For example, a person may feel they should receive an award for two very different reasons: 1) because s/he put forth a great amount of effort and contributed to many accomplishments in the area; 2) because s/he is inherently a special person who should receive awards regardless of accomplishments. In the first case, the person is said to be deserving of the award because of his or her actions (i.e., effort, accomplishments). In the second case, the person is said to be feeling unjustified entitlement for the award because s/he had no contributions or accomplishments. In other words, the entitled individual expects the positive outcome (i.e., the award) without assuming responsibility for the work or conducting the work; s/he expects the positive outcome because s/he feels innately special. This shift of responsibility has been conceptualized as externalized responsibility. In the context of education, externalized responsibility has been defined as the lack of responsibility for one's education (Chowning & N. J. Campbell, 2009). Not surprisingly, several measures of general (non-context-specific) unjustified entitlement correlated positively with externalized responsibility (*r*s = .29 to .38) indicating that unjustified entitlement is related to a lack of a sense of personal responsibility (Chowning & N. J. Campbell, 2009). Because the current research is focused on psychological (unjustified) entitlement, understanding the distinction between deservingness and unjustified entitlement is important. As explained below, distinguishing between these two constructs when constructing psychological entitlement measures is an area that needs further attention.

It is extremely important to characterize measures of "entitlement" as to what they actually represent: deservingness, unjustified (psychological) entitlement, or some combination of the two. A measure of psychological entitlement would have a different nomological net than a measure of deservingness. For example, one would expect people scoring high on a measure representing psychological entitlement to score low or moderate on a measure of effort, because effort has practically nothing to do with the expectation of rewards or resources. On the contrary, people scoring high on a measure representing deservingness should score high on a measure of effort because unlike entitlement, deservingness and effort go hand in hand. When a measure confounds deservingness and psychological entitlement, one is unable to hypothesize the relationship between scores on the confounded measure and scores on external measures (e.g., effort). Confounding of this kind would potentially prevent the researcher from accurately establishing or expanding a construct's nomological net. To ensure that this does not take place, measures should be carefully examined with respect to construct validity before use.

Measurement of Psychological Entitlement

With the growing interest in psychological entitlement and its emergence from simply a component of narcissism to an important independent construct, researchers have created several instruments to measure the construct. Because the focus of the current study is on general, non-context-specific psychological entitlement, only measures representing general psychological entitlement are reviewed.

NPI Entitlement subscale (Raskin & Terry, 1988). The Narcissistic Personality Inventory (NPI) was developed by Raskin and Terry in 1988 as a measure of narcissism in non-clinical populations. The NPI consists of seven factors: entitlement, authority, exhibitionism, exploitativeness, self-sufficiency, superiority, and vanity. The Entitlement subscale consists of six forced-choice items (e.g., "I will never be satisfied until I get all that I deserve" versus "I will take my satisfactions as they come"; Raskin & Terry, 1988). The authors' definition of entitlement was "the expectation of special privileges over others and special exemptions from normal social demands." Unfortunately, this definition does not appear to align very well with the content of the NPI Entitlement items. Items such as "If I ruled the world it would be a better place" and "I have a strong will to power" indicate seeking of authority, dominance, and power—not a sense of entitlement. Further, the item "I will never be satisfied until I get all that I deserve" certainly confounds entitlement and deservingness. If a respondent is truly deserving of outcomes and thus strongly endorses this item, then the item is indicative of the respondent's sense of deservingness, not his or her expectations of special privileges and exemptions (what the authors define as entitlement). Conceptual confounding of this type, in addition to the mismatch between the definition of the construct and the content of the items used to measure the construct (lack of face validity), as well as psychometric issues associated with the development of the NPI (see W. K. Campbell et al., 2004 for review) have resulted in concerns about the quality of the NPI Entitlement subscale as a measure of entitlement (W. K. Campbell et al., 2004; Pryor, Miller, & Gaughan, 2008; Watson & Biderman, 1993). Despite the measure's flaws, it is important to mention the NPI when discussing the measurement of psychological entitlement because before the emergence of entitlement as an independent construct, the NPI Entitlement subscale was the most popular instrument used to study entitlement, and thus it cannot be ignored. Yet, the issues associated with the measure indicate that research on psychological entitlement conducted using the NPI may be difficult to interpret. Thus, a better measure of psychological entitlement is needed.

Psychological Entitlement Scale (W. K. Campbell et al., 2004). With the increasing interest in entitlement as an independent construct and in an attempt to resolve some of the issues associated with the NPI Entitlement subscale, W. K. Campbell and colleagues (2004) developed a stand-alone measure of entitlement—the Psychological Entitlement Scale (PES). The PES consists of 9 Likert items with responses ranging from 1 ("strong disagreement) to 7 ("strong agreement"). The authors defined psychological entitlement as "a stable and pervasive sense that one deserves more and is entitled to more than others" (W. K. Campbell et al., 2004). With respect to the distinction between justified and unjustified, as well as general and context-specific entitlement, the PES can be considered a measure of unjustified, non-context-specific entitlement, based on the following statement by the authors: "Our concept of psychological entitlement is intrapsychically pervasive or global; it does not necessarily refer to entitlement that results from a specific situation (e.g., "I am entitled to social security because I paid into the system," or "I deserve an 'A' because I performed well in class")" (W. K. Campbell et al., 2004, p. 31).

With respect to the distinction between entitlement and deservingness, the authors state: "our concept of psychological entitlement includes both the experience of being deserving and entitled." Although they acknowledge the difference between the two terms ("deservingness typically reflects the expectation of a reward in exchange of one's own efforts or character, whereas entitlement typically reflects the expectation of a reward as a result of a social contract", W. K. Campbell et al., 2004), the authors of the PES used the terms interchangeably in the 9 PES items (e.g., "I honestly feel I'm just more deserving than others", "I feel entitled to more of everything"). Recall that to evaluate whether one is deserving vs. psychologically entitled to an outcome I asked the question *why*. A quick glance through the PES items (see Appendix) highlights that one is unable to answer this question, and therefore cannot determine whether the items are measuring entitlement or deservingness. For example, why am I more deserving than others? Am I more "deserving" because I actually put forth more effort than others, or simply because I am special? Different answers to the question "why" will result in a different interpretation of a strong endorsement of this item. That is, if a respondent believes that s/he is more deserving than others because s/he put forth more effort than others, then her or his strong endorsement of the item would indicate a high level of *deservingness.* Alternatively, if one endorses the item due to a feeling of being special, then the strong endorsement would be interpreted as a high level of *psychological entitlement*. Despite the likely confound of psychological entitlement and deservingness inherent in the measure, the PES remains the best available stand-alone measure of general psychological entitlement. Since its publication in 2004, the PES has been used in over 20 published studies.

Although much research has been conducted on psychological entitlement in the past few years and a good deal of this work employed the PES, there remains a question regarding the stability of the construct over time. Research efforts in the domain of psychological entitlement have largely focused on expanding entitlement's nomological net (i.e., investigating how entitlement relates to other constructs). Although such research is of great importance, it would also be useful to know if psychological entitlement increases, decreases, or remains approximately the same over time. Authors of the PES claim psychological entitlement is stable: "we consider psychological entitlement to reflect a chronic or stable disposition rather than a response to a specific social situation" (W. K. Campbell et al., 2004, p. 35). It appears from this description that psychological entitlement could be considered a rather steady personality characteristic (trait). In order to assess this argument, next I discuss the distinction between trait and state-line personality attributes.

Trait vs. State

When studying a construct, it is important that one understands how likely it is for the construct to change over a long time period. With respect to changeability, constructs are often categorized as trait, state, or both. *Traits* constitute attributes in individuals that are relatively stable across occasions (Hertzog & Nesselroade, 1987). Spielberger (1972, 1975, 1983) described personality traits as enduring individual differences in people or tendencies to perceive the world in a certain way and dispositions to react or behave in a particular fashion with predictable regularity. Trait anxiety, for example, falls in the category of traits, because it is characterized by constant worry and uneasiness.

States comprise attributes in individuals that are relatively changeable over time or occasions (Hertzog & Nesselroade, 1987). Spielberger (1972, 1975, 1983) defined personality states as transitory conditions that exist at a given moment at a certain intensity level. As such, states are often thought to be situation-specific. State anxiety, for example, only occurs in certain conditions (e.g., test anxiety occurs in the context of completing a test).

Given that traits and states are very different in nature (i.e., traits are stable, whereas states are transitory, changeable attributes), examining longitudinal stability can provide some insight into whether a construct is a trait or a state (Conley, 1984; Meyer & Shack, 1989; Nesselroade, 1986). Given that traits are stable, enduring attributes, a trait construct is expected to demonstrate high rank-order consistency and little mean-change over time (except when traits are measured during a developmental stage; Ozar & Gjerde, 1989). On the other hand, given that states are attributes transitory in nature, such constructs are expected to demonstrate less stability when measured at different times or occasions (Usala & Hertzog, 1991).

The importance of studying the temporal stability of constructs has been shown by many researchers. Veenhoven (1994), for example, investigated whether happiness was a fixed trait or a variable state, and found that over a short period of time happiness was stable, but over longer periods of time it was not. This finding is important because it demonstrates that constructs are not always easy to categorize as either trait or state, and often empirical evidence from longitudinal studies determines how the construct is conceptualized over the course of human development.

Existing Claims and Evidence Regarding the Stability of Psychological Entitlement

Although psychological entitlement has been the subject of much study, little research has been devoted to assessing the temporal stability of the construct. For this reason, all available research on entitlement's stability over time will be reviewed. Moreover, stability research on the construct that is next of kin to entitlement—narcissism—will also be reviewed. That is, because entitlement was first defined as a component of narcissism and later studied in the context of narcissism, examining the stability of narcissism should provide additional insight into the stability of entitlement across time.

Research on the stability of entitlement across time. The study of entitlement as an independent construct was largely influenced by the development of the Psychological Entitlement Scale (W. K. Campbell et al., 2004). Recall that the authors defined psychological entitlement as "a stable and pervasive sense that one deserves more and is entitled to more than others" (p. 35). The researchers further state: "we consider psychological entitlement to reflect a chronic or stable disposition rather than a response to a specific social situation" (p.35). It appears that the authors conceptualize entitlement as a personality trait; however, no theoretical explanation is provided as to why entitlement should be considered a stable trait. Empirically, there was some evidence of relative stability of entitlement across time. Specifically, two different samples were administered the PES at two time points, 1 month apart for one sample, and 2 months apart for the other sample. The results yielded rank-order consistency coefficients of .72 and .70, respectively, leading the authors to the conclusion that "the PES is stable over time" (W. K. Campbell et al., 2004). With respect to mean-level change in entitlement across time, PES scores at baseline were 28.8 and 27.7 and increased to 30.3 (1 month later) and 30.6 (2 months later) upon subsequent testing (scale ranges from 7 to 63). No interpretations of these mean-level changes were provided by the researchers.

Zitek and colleagues (2010) proposed that psychological entitlement is not just a chronic disposition, but can also be a dynamic mindset. They suggested that "an individual can also vary in the extent to which s/he feels entitled in the course of any given day, depending on what past experiences are salient in the individual's mind when the opportunity for selfish behavior presents itself" (Zitek, Jordan, Monin, & Leach, 2010, p. 246). The authors supported this view of psychological entitlement through a series of experiments in a repeated measures design, where participants expressed higher levels of psychological entitlement (represented by several adapted PES items) when they were wronged, or even reminded of a time when they were wronged. Thus, these studies provide some evidence that psychological entitlement may not be as "pervasive and stable" as defined by W. K. Campbell and colleagues (2004) but rather may be situation-specific, or state-like.

Research on the stability of narcissism across time. Given the limited empirical evidence regarding the stability of psychological entitlement, the stability of narcissism is reviewed next as a source of evidence for the temporal stability of a construct close to entitlement in the nomological net. Given that narcissism was included in the DSM-III-R and DSM-IV-TR as a personality disorder, and all personality disorders stem from maladaptive personality traits (DSM-III-R, 1980), one would expect narcissism to be relatively stable over time. Consistent with the initial interest in narcissism in clinical populations, the first instance of studying narcissism over time was with patients diagnosed with clinical narcissistic personality disorder (NPD). Ronningstam, Gunderson, and Lyons (1995) studied 20 NPD patients over the course of 3 years, and found that for 12 of the patients, levels of pathological narcissism as measured by the NPD criteria in the DSM-III-R and DSM-IV decreased by over 50%, and remained unchanged for the other 8 patients. Moreover, the qualitative research accompanying the study revealed that corrective achievements (e.g., graduations, promotions, recognitions, acceptance to sought-after schools, programs, or positions, etc.) were the most common type of event that contributed to the decrease in pathological entitlement (Ronningstam et al., 1995).

Another stability study of narcissism was conducted a decade later using the Narcissistic Personality Inventory (NPI; Raskin & Terry, 1988). Del Rosario and White (2005) administered the NPI to a college population at baseline, and again 13 weeks later. The authors reported rank-order consistency coefficients for the full 40-item NPI scale, as well as for each of the seven subscales. The entire NPI scale demonstrated high test-retest stability (r = .81), and test-retest stability coefficients for most of the subscales ranged from high moderate (r = .70-.80) to low moderate (r = .60-.70), with entitlement being the component of narcissism with lowest stability (r = .57). The researchers concluded that the NPI scale and its subscales demonstrated satisfactory stability, indicating that narcissism and its components remained more or less stable over time, as expected (del Rosario & White, 2005). It should be noted, however, that one could question the utility and interpretability of an overall NPI stability coefficient, considering the wide range of test-retest stability coefficients among the subscales.

In a study examining the psychometric properties of the Italian version of the NPI, Fossati, Borroni, and Maffei (2008) reported test-retest stability for the full NPI scale for a clinical and a non-clinical population over the course of three months: the stability coefficient for the non-clinical sample was higher (r = .87) than it was in the clinical sample (r = .72). Nonetheless, this study lends further support that narcissism, as measured by the NPI, may be relatively stable over time. However, as mentioned above, it seems it would be more meaningful to present and interpret stability coefficients for the NPI subscales individually.

In sum, there are few studies assessing the stability of entitlement and narcissism over time. Moreover, the few empirical findings do not appear to align with the commonly accepted theoretical conceptualization of psychological entitlement. Specifically, W. K. Campbell and colleagues (2004) defined entitlement as “stable” but provided no theoretical justification for this classification. Furthermore, the empirical evidence they presented is insufficient to support their claim for several reasons. First, they did not assess measurement invariance across time. Thus, any rank-order consistency coefficients or mean-level change estimates may be uninterpretable. Second, the authors deemed test-retest consistency coefficients to be sufficient evidence that the construct is stable, without interpreting the mean-level change in entitlement from baseline to 1 and 2 months, respectively, in both samples.

Similar limitations are associated with the stability studies using the NPI (del Rosario & White, 2005; Fossati et al., 2008). First, there have been no studies of longitudinal measurement invariance of the NPI. Second, del Rosario and White (2005) reported high moderate (*r* = .70 - .80) to low moderate (*r* = .60 - .70) test-retest consistency coefficients for six of the subscales, and the lowest coefficient (*r* = .57) for the entitlement subscale. Because of the discrepancies of the coefficient values across subscales, the high overall stability coefficient for the full scale is difficult to interpret. Moreover, del Rosario and White did not interpret mean-level change. Such isolated results make it difficult to determine whether the entitlement construct is stable or not. Clearly, further study is needed to address these issues and provide more empirical evidence.

Is psychological entitlement stable over long periods of time? Both empirical findings and theory suggest that some individual differences may be highly stable over short periods of time but less stable as the time interval between testing occasions increases (Fraley & Roberts, 2005; Veenhoven, 1994). Recall that when studied over 1 or 2 month periods, entitlement's test-retest consistency was .72 and .70, respectively. In addition, when studied as a component of narcissism and assessed over a 13-week time period, entitlement was associated with the lowest stability coefficient (*r* =. 57). These results may suggest that over short periods of time, entitlement may be indeed stable; however, over long periods of time (>1 year), entitlement may not be as stable. Given the limited empirical evidence and theoretical background on psychological entitlement, I believe that entitlement will not be stable over a longer period of time (e.g., 1.5 or 2 years).

Chapter III

Methods

Data Collection Procedure

Responses to the PES were collected from college students at a mid-sized southeastern university as part of a large-scale testing session for institutional accountability purposes (U.S. Department of Education, 2006). Specifically, in order to examine the "value-added" due to attending the university, a sample of students was assessed on a battery of both knowledge (e.g., quantitative reasoning) and attitudinal measures (e.g., PES) twice during their college career: once as incoming freshmen (Time 1), and again in the spring semester of their sophomore year (1.5 years after they started college) if they had completed between 45 and 70 credit hours (Time 2). Most incoming freshmen were assessed at Time 1. However, it is important to note that not all students who were administered the PES at Time 1 also completed the measure at Time 2. Some students, who completed the PES at Time 1, earned more than 70 credits by the end of their third semester, and therefore were not required to complete the assessments at Time 2. Likewise, not all students who were administered the PES at Time 2 completed the measure at Time 1. For example, students who transferred to the university and had earned enough credit hours to complete the assessments at Time 2 would not have PES scores at Time 1. The group of participants that is of interest for the current study is comprised of students who completed the PES both at Time 1 and Time 2.

It is important to note that although these assessments are mandatory and high-stakes to the university, they are of low-stakes to the students. That is, although the university makes important decisions based on the assessment results, for students the results bare no serious personal consequences. Therefore, some participants may not have been very motivated to put forth their best effort while completing the assessments, which in turn may affect the reliability of scores, as well as the validity of inferences made based on scores. Fortunately, the PES is a short measure, thus participants are not likely to get fatigued. Moreover, as described below, extensive data screening was conducted to identify and remove any participants who were likely not providing valid responses to the PES.

Measure

The Psychological Entitlement Scale (PES; W. K. Campbell et al., 2004) was used to measure participants' levels of psychological entitlement at Time 1 and Time 2 (see Appendix for PES items). The PES consists of 9 Likert items with response options ranging from 1 ("strong disagreement") to 7 ("strong agreement"). Item 5 ("I do not necessarily deserve special treatment") is negatively-phrased, and thus Item 5 was reverse-scored prior to data analysis. Higher PES scores indicate a greater level of psychological entitlement. Importantly, the reverse-scored Item 5 was used to identify respondents who provided invalid scores (i.e., engaged in a response set), as is discussed below.

Participants

Data from Time 1: Incoming college freshmen. At Time 1, responses were gathered from 3749 incoming college freshmen. Only 35 participants (0.9%) had missing data on one or more of the PES items. These 35 students were excluded from the sample, resulting in 3714 freshmen with complete data from the first testing occasion. Further, Mahalanobis distance values were examined to identify multivariate outliers. Six participants were deemed multivariate outliers due to their nonsensical response patterns (e.g., 1, 7, 1, 7,…). Moreover, 29 participants provided what appeared to be response sets (e.g., 7, 7, 7, 7, 1, 7, 7, 7, 7). Recall Item 5 is negatively-phrased and was reverse-scored; thus, these 29 participants appear to not have attended to the content of the items and simply recorded the same response (e.g., "7") for every item. Thus, an additional 35 participants were removed, leaving a sample of 3679 at Time 1.

Data from Time 2: Second-semester sophomores. At Time 2, responses to the PES were gathered from 3346 participants. It is important to note that this sample of 3346 participants consisted only of students with 45-70 credits. Thus, not all students who were tested at Time 1 were tested at Time 2; some students tested at Time 1 had less than 45 credits or more than 70 credits and were not eligible to be tested at Time 2. In addition, this Time 2 sample contains participants who were not tested at Time 1 (transfer students).

Only 33 participants had missing data on some PES items (1.0%), and thus were excluded from the sample, resulting in 3313 participants with complete data from the second testing occasion. Further, Mahalanobis distance values were examined, and six participants were identified as multivariate outliers with nonsensical response patterns (e.g., 2, 7, 1, 7,…). Moreover, 31 participants were identified as providing a response set (e.g., 2, 2, 2, 2, 6, 2, 2, 2, 2), thus resulting in a sample of 3276 at Time 2.

Matched data (Time 1 and Time 2). For the purposes of the analyses in the current study, datasets from Time 1 and Time 2 were matched by participant ID, resulting in a sample of 2195 participants with complete data from both Time 1 (incoming freshman) and Time 2 (fourth semester college student). A summary of sample sizes across testing occasions before the removal of outliers and response sets is presented in Table 1. The noticeable decrease in the number of participants in the matched sample compared to the individual samples at Time 1 (N = 3679) and Time 2 (N = 3276) is due to the fact that some participants were not tested at either occasion for one of two reasons. As mentioned above, some students who were tested at Time 1 were not tested at Time 2 because they were not eligible to be tested; in fact only 65% of those tested at Time 1 (or 2442 of the 3749 students tested at Time 1) were within the required range of 45-70 credit hours in order to be tested at Time 2. In addition, 12% of the students who were tested at Time 2 (or 402 of the 3346 students tested at Time 2) had no data from Time 1 because they were transfer students. Thus, these two groups of students explain the decreased sample size of the matched sample compared to the full sample at Time 1 or Time 2 (see Table 1).

The matched sample consisted of 83.1% Caucasian students, 5.2% Asian students, 2.7% Hispanic students, 2.1% Black students, 0.6% Pacific Islander students, 0.2% American Indian students, and 6.1% students who did not specify their ethnic background. The sample consisted of 66.7% female students. At Time 1, the participants had an average age of 18.43 years (SD = 0.39).

Data Analysis and Model Evaluation

The PES data from Time 1 and Time 2 were analyzed in three stages. First, an examination and interpretation of descriptive statistics was conducted. Next, measurement invariance was tested. Finally, given (at least partial) measurement invariance, latent variable stability testing was conducted.

When testing measurement invariance, model fit was assessed via a chi-square (χ^2) significance test of global fit, several indices of approximate fit, and residuals. The χ^2 significance test is an exact test of the model-data fit. That is, a significant χ^2 would indicate that elements of the reproduced covariance matrix are significantly different from elements of the original covariance matrix. Researchers have noted that the χ^2 test is very sensitive to sample size, such that when sample size increases, so does the likelihood of obtaining a significant χ^2 (Hu & Bentler, 1998). Because of this, and more importantly, because researchers are typically interested in measures of approximate fit rather than measures of exact fit, the χ^2 was supplemented by two types of approximate fit indices: absolute and incremental.

The standardized root mean square residual (SRMR) is an absolute fit index, which indicates the degree of global misfit between the observed and model-implied relationships among the modeled variables. The SRMR is especially sensitive to misspecified factor correlations. SRMR values range from 0 to 1, and smaller values are indicative of better fit. Specifically, researchers have recommended a cutoff value of $\leq$ .08 as an indicator of adequate global model-data fit (Hu & Bentler, 1998; 1999).

The root mean square error of approximation (RMSEA) is another absolute fit index indicative of global model-data fit. The RMSEA is sensitive to misspecified factor pattern coefficients (i.e., the relationships between the factor and the items). Values for the RMSEA range from 0 to 1, and smaller values are indicative of better fit. Specifically, values $\leq$ .06 have been recommended as indicators of good global model-data fit (Hu & Bentler, 1999).

The comparative fit index (CFI) is an incremental fit index, which compares the fit of the hypothesized model to a null model, where the relationships among all observed variables (items in the current study) are set to equal zero. That is, the CFI indicates the degree of model-data fit improvement of the current model from a null model. Similar to the RMSEA, the CFI is very sensitive to misspecified factor pattern coefficients (Hu & Bentler, 1998). CFI values range from 0 to 1, with larger values indicative of better fit. Hu and Bentler (1999) recommend using a cutoff of $\geq .95$ as a criterion for good global fit.

It is important to note that some researchers have challenged the appropriateness of the aforementioned fit index cutoff values and have recommended using them as guidelines rather than exact rules (Fan & Sivo, 2007; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004; Sharma, Mukherjee, Kumar, & Dillon, 2005). In fact, Hu and Bentler (1998) themselves advise that unlike the strict χ^2 statistic, which offers a dichotomous decision based on a significance test, fit indices can be used to "quantify the degree of fit along a continuum" (p. 426). Moreover, Vandenberg and Lance (2000) suggest using a range of cutoffs instead of strict cutoffs.

In addition to assessing global fit with the χ^2 test and approximate fit indices, local model-data fit was assessed by examining each correlation and mean residual. Correlation residuals are simply the difference values between elements of the observed correlation matrix and corresponding elements of the reproduced correlation matrix. Correlation residuals greater than $|.10|$ are indicative of specific item-pair relationships not reproduced well by the model (Kline, 2011). Mean residuals, which are pertinent when testing scalar invariance, are the difference values between the observed item means and the reproduced item means for each item at a given time point. Unlike the correlation residuals, there is no specific cutoff value that could serve as a guideline in establishing scalar measurement invariance. Rather, the mean residuals should be interpreted in the context of the scale being used (e.g., residuals of |.50| on a 4-point scale would be considered large).

Chapter IV

Results

Descriptive statistics

Descriptive statistics for the 9 PES items are summarized in Table 2. The item means and standard deviations were examined to determine how respondents scored on average and the variability in scores around these means. Item means varied between 2.4 and 3.7 at Time 1, and between 2.5 and 3.6 at Time 2, indicating respondents tended to slightly or moderately disagree with the entitlement statements. All standard deviations were greater than 1.0, indicating there was sufficient variability associated with each item (i.e., no floor or ceiling effects). In addition, frequency distributions indicated all response options (i.e., 1 through 7) were utilized by respondents at both time points.

Univariate normality was assessed via comparison of skewness and kurtosis values for the 9 PES items against suggested cutoffs of |2| and |7|, respectively (West, Finch, & Curran, 1995). All 9 items displayed univariate normal distributions; univariate skewness and kurtosis values at both time points were less than |1.0|. Multivariate normality was assessed via Mardia's normalized multivariate kurtosis statistic. Although there is no agreed-upon cutoff value, the Mardia's value is typically reported in covariance structures analysis. Mardia's normalized multivariate kurtosis for the sample at Time 1 was 35.55, and 43.30 at Time 2, suggesting multivariate non-normality (Finney & DiStefano, 2006). Because of the possible effects of multivariate non-normality, the Satorra-Bentler (S-B) scaling method (Satorra & Bentler, 1994) was employed in conjunction with maximum likelihood (ML) estimation. The S-B scaling adjusts the ML χ^2, ML standard errors, and ML fit indices for multivariate kurtosis, thus producing less biased estimates of model fit and standard errors of the estimated parameters (Finney & DiStefano, 2006). Given the univariate normality of the data, CFAs were also conducted using the ML estimator without the S-B adjustment, in order to determine whether the use of the S-B scaling method would produce substantively different results. Previous studies have reported that when the data are fairly normally distributed univariately and have at least five response options, bias due to multivariate non-normality is negligible (Finney & DiStefano, 2006; West, Finch, & Curran, 1995). Indeed, the χ^2 statistic, standard errors, and fit indices did not differ substantially when the unadjusted ML estimator and the S-B scaling method were employed. Therefore, only the (unadjusted) ML estimation results are presented here.

Multicollinearity was assessed within time points by examining the correlation matrix for highly related items ($r > .80$). Within each time point, the PES items were moderately correlated at both time points, with the largest correlations being .59 at Time 1 and .61 at Time 2, indicating no extreme multicollinearity.

Recall that Item 5 is negatively-phrased, and as such it correlated less strongly with the other items. Ignoring Item 5, the inter-item correlations were of approximately the same magnitude (ranging from .36 to .59 for Time 1 and from .36 to .61 for Time 2), suggesting that a unidimensional model might represent the data well, with a weak relationship between Item 5 and the factor at both time points. This hypothesized model was formally assessed when testing configural invariance.

Measurement Invariance

LMACS analysis was used to test configural, metric, and scalar invariance. Researchers generally agree that when both rank-order *and* latent mean change are evaluated, configural, metric, and scalar invariance need to be established (Bontempo & Hofer, 2006; Sass, 2011; Steenkamp & Baumgartner, 1998; Thompson & Green, 2006, Vandenberg & Lance, 2000; Widaman, Ferrer, & Conger, 2010). Several additional tests of invariance can be performed (e.g., test of equivalent error variances, test of equal factor variances); however, these tests are not necessary when the researcher's purpose is to compare *latent* means (Byrne & Stewart, 2006; Schmitt & Kuljanin, 2008; Widaman & Reise, 1997).

Indices used to assess measurement invariance. Testing for measurement invariance involved a multi-step approach, with less constrained models being supported before more constrained models were tested. Moreover, because the more constrained models are nested within the less constrained models, differences in model fit could be evaluated. Specifically, if the fit of a more constrained model was not statistically or practically worse than the fit of a less constrained model, invariance at the level of the more constrained model was established (Dimitrov, 2006; Steenkamp & Baumgartner, 1998). Both a non-significant chi-square difference ($\Delta\chi^2$) test and CFI difference $\leq .01$ were used as criteria in evaluating measurement invariance models. Specifically, some researchers contend that similar to the χ^2 statistic, the $\Delta\chi^2$ is also an exact test of fit, and therefore it should be supplemented with change in the values of approximate fit indices, such as the CFI (Steenkamp & Baumgartner, 1998). Cheung and Rensvold (2002) have suggested that the CFI is especially effective when examining the difference in fit across nested models, and have recommended change in CFI (ΔCFI) $\leq .01$ as an indicator of invariance. Other researchers, however, have challenged this approach, noting that the CFI lacks power and may not signal non-invariance when the $\Delta\chi^2$ does, and therefore only the $\Delta\chi^2$ test should be used (French & Finch, 2006). Still other researchers have argued that a relative fit (e.g., CFI) approach is more appropriate for the measurement part, whereas an exact fit ($\Delta\chi^2$) approach is more appropriate for the structural part (e.g., when estimating the latent mean difference) of invariance testing (Little, Card, Slegers, & Ledford, 2007). Given that the $\Delta\chi^2$ test and the ΔCFI often provide conflicting results, these two indices were supplemented by the correlation and mean residuals in order to determine whether invariance had been established.

Scaling the factor. There are several methods of setting the metric of the factor(s) being modeled and each method has different advantages.[1] For a single group CFA model, it is common practice to scale the factor by setting its variance to a value of one. In this case, the researcher is not interested in estimating or interpreting the factor variance, but rather the factor pattern coefficients. In testing measurement invariance over time, however, this method may not be appropriate for several reasons. First, fixing the factor variance to one at each time point assumes that factor variances are equal across time, which is overly restrictive (Marsh, 1994). Related to this point, when testing for metric invariance, one would be testing the invariance of factor variances in addition to the invariance of factor pattern coefficients (just the latter would be sufficient to establish metric invariance), thus making the model too restrictive. Second, it may of interest to the researcher to estimate the factor variances and interpret them in order to gauge the variability of latent scores across time.

Given these considerations, the factors in each model were scaled by fixing the factor pattern coefficient between the factor and Item 9 to a value of one (i.e., Item 9 served as the referent indicator; see Little et al., 2007). All other parameters were either freely estimated or constrained to a value appropriate for the level of invariance being tested. This constraint on Item 9 to set the metric of the factor assumes that Item 9 is invariant across time. To test the invariance of Item 9, each of the models were re-estimated using Item 1 as the referent indicator (see Rensvold & Cheung, 2001). Given that all of these re-estimated models resulted in the same fit as their counterparts when Item 9 was used as the referent item, the invariance of Item 9 was established.

Configural invariance. Configural invariance was tested in two steps. First, the covariance matrices from each time point were analyzed separately to test if the one-factor model fit the data at Time 1 and Time 2 independently. CFAs were conducted using LISREL 8.80 (Jöreskog & Sörbom, 2006).

The one-factor model fit the data well globally at both time points: χ^2 (27, N = 2195) = 434.45, CFI = .95, RMSEA = .08, SRMR = .04 at Time 1; χ^2 (27, N = 2195) = 541.19, CFI = .94, RMSEA = .09, SRMR = .04 at Time 2. Moreover, there were no localized areas of misfit at either time point; all correlation residuals were less than |.10|, with the largest being .08 at both time points, and 72% of the correlation residuals falling at or below |.05| at Time 1 and 67% at Time 2. Given adequate model-data fit, factor pattern coefficients could be interpreted.

Large standardized pattern coefficients suggest that items adequately represented the construct (Thompson, 1997). At both Time 1 and Time 2, all 9 PES items were associated with standardized pattern coefficients at or above .65 and .67, respectively, indicating that the factor explained at least 42% (Time 1) and 45% (Time 2) of each item's variance, except for Item 5 (8% and 14%, respectively; see Table 3). The weak relationship between the factor and Item 5 was not unexpected given the weak correlations between Item 5 and the remaining items on the PES. Importantly, four of the PES items at Time 1 and six at Time 2 had standardized pattern coefficients at or above .7 (i.e., the factor explained at least 50% of its items' variance). Except for Item 5, the items appeared to represent the factor well. Moreover, internal consistency, calculated using the CFA's unstandardized parameters (McDonald, 1999; Reuterberg & Gustafsson, 1992), was adequate (ω = .88 at Time 1 and .89 at Time 2).

The second step of testing for configural invariance involved fitting a unidimensional factor structure to the data from Time 1 and Time 2 simultaneously by analyzing the large covariance matrix that represents both time points (i.e., analyze the relationships represented in Table 2). This model, pictured in Figure 2, includes error covariances between corresponding items from Time 1 and Time 2 (autocorrelations) to account for the systematic error variance shared by corresponding pairs of items across testing occasions. As expected, given the well-fitting independent CFA models, this combined model fit the data from Time 1 and Time 2 well globally (see Table 4), with no localized areas of misfit (all correlation residuals were $< |.10|$, with 85% of the residuals being $< |.05|$). Both the unstandardized and standardized pattern coefficients remained essentially the same as those presented in Table 3. Given these results, configural invariance was established, and this model was used as the baseline model for testing metric invariance.

Metric invariance. In the metric invariance model, corresponding unstandardized factor pattern coefficients at Time 1 and Time 2 were constrained to be equal, whereas the remaining parameters were freely estimated. The metric invariance model fit the data well, and the global fit was not significantly worse than that of the baseline (configural invariance) model (see Table 4). Moreover, all correlation residuals were < |.10| (83% were < |.05|), indicating no localized areas of misfit. With respect to the PES, establishing metric invariance indicated that each of the nine items were equally salient to the factor across time. That is, the construct was equally well represented by the PES items at Time 1 and Time 2. Given metric invariance was established, this model served as the baseline model for testing scalar invariance.

Scalar invariance. Several scalar invariance models were tested (see Table 4). In each model, the intercept for Item 9 was fixed to zero at Time 1 and Time 2, in order to estimate the factor means. In the first (fully constrained) scalar invariance model, the intercepts of all corresponding items were constrained to be equal across time. This scalar model fit the data well globally; however, the $\Delta\chi^2$ test comparing the fit of this model to the fit of the metric invariance model was significant, signaling partial scalar non-invariance.

Examination of the residuals for the means indicated somewhat large residuals for the means of Items 4, 2, and 3: |.11|, |.10|, and |.05|, respectively, on a 7-point scale. It is important to note that due to the equality constraints placed on the item intercepts, when an item mean is not fully driven by the latent mean across time, there will be a positive residual associated with that item mean at one point, and a negative residual for the item mean at the other time point. For example, the value of -.11 indicated that the mean for Item 4 was overestimated by the model by about a tenth of a point at Time 1 and accordingly underestimated by the same amount at Time 2 (i.e., a positive residual of .11). That is, the model implied means for Item 4 reflect solely the latent mean difference across time; thus, a value of -.11 at Time 1 and a value of .11 at Time 2 indicates the estimated latent mean difference underestimated the observed mean difference across time for Item 4. Interestingly, the opposite was true for Item 2 (i.e., the item mean was underestimated by the model at Time 1 and overestimated at Time 2, respectively), indicating that the means for Items 2 and 4 were "biased" in opposite directions from Time 1 to Time 2. Again, these mean residuals reflect the restrictive equality constraints that were imposed on the item intercepts. That is, when specifying both metric and scalar invariance, the model-implied item means at Time 1 and Time 2 can *only* reflect the latent means at Time 1 and Time 2. In other words, something other than the factor is influencing the responses to those three items, leading to systematic over- (Item 4) and underestimation (Items 2 and 3) of the item means at Time 1.

The mean residuals associated with these three items were not large given the 7-point scale. In order to assess the potential effect of scalar non-invariance (or uniform differential item functioning, DIF) for Items 2, 3, and 4, three additional scalar invariance models were estimated, the fit of which was compared back to the baseline (metric invariance) model. These results are presented in Table 4. The equal-intercept constraint for the three potentially non-invariant items was released in each model one at a time (i.e., first for Item 4, then for Item 2, and finally for Item 3). The tests were conducted in this particular order based on the absolute size of the residuals associated with these items (largest to smallest). The $\Delta\chi^2$ tests between the baseline (metric invariance) model and the first two models were significant, indicating that even when the intercepts associated with Items 4 and 2 were freely estimated across time, these two models still fit significantly worse than the baseline metric invariance model. That is, some additional items, beyond those already freely estimated, appeared scalar non-invariant. Specifically, each of these two models had one mean residual greater than |.05| indicating the intercept for that item should not be constrained to be equal across time, in order to obtain accurate mean estimates. For the last scalar invariance model (freely estimated intercepts for Items 4, 2, and 3) however, the $\Delta\chi^2$ test was not significant, indicating this partial scalar invariance model fit the data no worse than the baseline metric invariance model. Moreover, no mean residual was greater than |.03| (a trivial difference, given the 7-point scale), indicating this partial scalar invariance model reproduced the means most accurately compared to the other scalar invariance models that were tested. Given these results, this third partial scalar invariance model was deemed to best represent the data from Time 1 and Time 2, with Items 1, 5, 6, 7, 8, and 9 demonstrating full measurement invariance, and Items 2, 3, and 4 having configural and metric, but no scalar invariance. However, this conclusion may be conservative, as the mean residuals were not overly large for any item. Thus, to more fully evaluate the impact of the scalar non-invariant items and assess its practical significance, latent mean difference tests were computed with and without placing equality constraints on the intercepts associated with Items 2, 3, and 4.

Latent Variable Stability

Given partial scalar invariance, a test of the latent mean difference was conducted both with and without placing equality constraints on the scalar non-invariant items (see Table 5). It is important to note that when the intercepts for some items are freely estimated across time, those items do not contribute to the latent mean at either time point. Regardless of whether the intercepts for the non-invariant items were constrained to be equal or freely estimated across time, the latent mean difference was not significant, and the effect size was small (.012 and .041, respectively). The high similarity in latent mean difference results based on all PES items (i.e., the intercepts for all corresponding items were constrained to be equal across time) vs. only scalar invariant items (i.e., the intercepts of scalar non-invariant items were freely estimated across time) was most likely due to items having opposite patterns of uniform DIF. That is, the mean for Item 4 was overestimated at Time 1, whereas the means for Items 2 and 3 were underestimated at Time 1. The opposite pattern occurred for Time 2. Thus, the latent mean difference was not affected by the presence of this DIF in the fully scalar invariant model. Practically the same results were obtained by computing a mean difference at the observed level. The similarity of results at the latent and the observed level was due to the high internal consistency of the observed PES scores at Time 1 and Time 2 (ω = .88 at Time 1 and .89 at Time 2).

Given the equivalence of factor structures and unstandardized pattern coefficients (i.e., full configural and metric invariance), latent rank-order consistency across time could be interpreted. The correlation coefficient between the psychological entitlement factors at Time 1 and Time 2 was .61, $p < .05$ (factor variances at Time 1 and Time 2 were 1.16 and 1.32, respectively), indicating that the rank-order of respondents did change to some degree. Although attenuated due to measurement error, the correlation of observed scores at Time 1 and Time 2 was practically of the same value ($r = .58$, $p < .05$). Again, the highly similar latent and observed score correlations are due to the high internal consistency of the PES scores at both time points.

Given that rank-order consistency was only moderate, it appears that some individuals increased on entitlement, whereas others decreased across time, leading to minimal, practically insignificant mean-level change over 1.5 years. Moreover, this differential change was demonstrated by plotting the change trajectories of a random sample of 5% of the participants (see Figure 4). The graph clearly shows that some participants increased on entitlement from Time 1 to Time 2, others decreased across time, and still others remained the same over time

Chapter V

Discussion

The purpose of the current study was two-fold. First, a review and comparison of commonly used methods of assessing construct stability over two time points and a modern latent variable approach (LMACS analysis) was provided. Second, the application of LMACS analysis to assess stability claims about psychological entitlement was demonstrated. Below is an in-depth discussion of the importance of assessing longitudinal measurement invariance and the effects of partial non-invariance (i.e., items displaying DIF), followed by the implications of the current study for the psychological entitlement domain, and ending with some highlights of the advantages of using a unified LMACS analysis approach to assessing invariance and construct stability over two time points.

Establishing Measurement Invariance for the PES and Assessing the Effects of Partial Non-Invariance

The current study demonstrated how to empirically test the assumption of measurement invariance in longitudinal research. In order to do so, several models with various levels of equality constraints were specified and fit to the PES data at Time 1 and Time 2. First, a hypothesized one-factor model was fit to the data at each time point independently and then with the data combined (configural invariance) to establish a comparable factor structure across time. Both the independent one-factor models and the combined data (correlated single-factor) model fit the PES data well, meaning that participants defined the construct represented by the PES items similarly across time. If this were not the case, none of the subsequent invariance testing, nor the latent stability tests, would have been meaningful because one would be comparing scores representing conceptually different things. Fortunately, the PES demonstrated stable factor structure across time; thus, the stability of measurement parameters could be assessed.

Subsequent invariance testing involved constraining the magnitude of factor pattern coefficients (metric invariance) and intercepts (scalar invariance) of corresponding items to be equal across time. The model with constrained factor pattern coefficients resulted in adequate fit, indicating that corresponding PES items were equally representative of the factor across time. That is, each item was as salient to the factor at Time 1 as it was at Time 2. However, if this metric model did not fit the data well, one would be observing non-uniform DIF (revealed by large correlation residuals for metric non-invariant items across time). In a longitudinal design, non-uniform DIF occurs when an item's observed score differs from one testing occasion to the next, whereas the latent score remains the same across time. That is, the item could be highly related to the factor at Time 1, but the relationship with the factor becomes weaker at Time 2, or vice versa; thus biasing the total score at either time point. When non-uniform DIF is present, one should no longer interpret the total scores from Time 1 and Time 2 because they are no longer comparable and could lead the researcher to erroneous conclusions. Fortunately, the model with equality constraints on the factor pattern coefficients of corresponding PES items fit the data just as well as the unconstrained configural invariance model, thus establishing full metric invariance for the PES and allowing subsequent invariance and rank-order stability testing.

The final set of parameters for which invariance needed to be established before being able to analyze full construct stability (rank-order *and* mean-level stability) involved testing a model with equality constraints on the intercepts of corresponding items across time. Unfortunately, this model resulted in a significantly worse fit than the metric invariance model described above, and the relatively small mean residuals for three items (Items 2, 3, and 4) indicated that these items may be exhibiting uniform DIF. That is, the items were equally representative of the factor across time (they were metric invariant), but their intercepts (item means when the factor equals zero) were systematically biased from one time point to the next. Specifically, the mean for Item 4 was overestimated at Time 1, whereas the means for Items 2 and 3 were underestimated. Recall that DIF occurs when extraneous variables, other than the construct of interest (psychological entitlement), are affecting the responses to non-invariant items. Both non-uniform and uniform DIF affect the responses but do so differently (i.e., when the DIF is non-uniform vs. when it is uniform). With non-uniform DIF, the extraneous variables affect the relationship between the metric non-invariant item and the factor such that this relationship is stronger at one time point and weaker at another time point, thus resulting in different observed scores for the item with the same latent score across time. With uniform DIF, the extraneous variables affect responses to the item such that that the extraneous variables systematically "bump" the intercept of a scalar non-invariant item at a given time point when the intercept (the item mean when the factor is zero) actually remains the same across time. Given that only uniform DIF was present for three of the PES items, below I discuss several potential variables that may be responsible for the uniform DIF displayed by those three item (Items 2, 3, and 4).

The first thing to notice regarding the three scalar non-invariant items is their pattern of change from Time 1 to Time 2 (see Table 2). Unlike the means of the rest of the PES items, which either increase or remain roughly the same from Time 1 to Time 2, the means for two of the scalar non-invariant items (Items 2 and 3) *decrease* over time. Clearly, Items 2 and 3 show a different pattern of change over time compared to the rest of the PES items; thus, Items 2 and 3 do not reflect the small *positive* increase of PES scores across time (i.e., the means of Items 2 and 3 decrease over time). This differential pattern of change could well be the key to uncovering the aforementioned possible extraneous variables affecting the responses for the three items displaying uniform DIF.

By far the most plausible explanation for the systematic upward or downward bias in Items 2, 3, and 4 could be the wording of the items, coupled with the effect of maturation after 1.5 years of college. That is, Item 2 ("Great things should come to me") and Item 3 ("If I were on the Titanic, I would deserve to be on the *first* lifeboat!") appear to be more extreme in nature and language than the rest of the items. Thus, Items 2 and 3 may have been endorsed more at Time 1 than the model predicted and less at Time 2 than the model predicted because of the expectations college freshmen typically have at the very beginning of their college career vs. later in their college career. However, this attitudinal change only affected the responses to the two most extreme items (Items 2 and 3). That is, 1.5 years later students may not feel as "special" as they did when they were freshmen, and thus endorsed these two more extreme items less strongly at Time 2 vs. Time 1 relative to the endorsement of the other less extreme items at Time 2 vs. Time 1 (which actually increased or remained approximately the same across time). Thus, because of the extremeness of these two items and the perceived feeling of being special typical for students at the beginning of college, responses to Items 2 and 3 were systematically biased across time (i.e., underestimated at Time 1 and overestimated at Time 2).

A similar logic could be applied to the uniform DIF of Item 4 (“I demand the best because I am worth it”). Unlike the rest of the PES items, Item 4 contains an explanatory clause as to why the respondent should endorse the item. That is, the respondent must possess some sort of “worthiness” in order to “demand the best”. Thus, Item 4 may be affected by the extraneous variable of deservingness. When students completed the measure as incoming freshmen, they may have not perceived themselves as worthy of demanding the best, and thus did not endorse Item 4 very strongly. However, after 1.5 years of college, students likely experienced some hard work and/or struggle, and therefore agreed with Item 4 more strongly than would be expected based solely on their level of the latent construct psychological entitlement (an increase in the extraneous variable deservingness is causing this unexpected increase in the item mean across time).

It is important to note that wording of the items combined with maturation of the students is one of many possibilities why Items 2, 3, and 4 are exhibiting uniform DIF. The fact that the bias among these items is bidirectional (i.e., positive for Item 4, but negative for Items 2 and 3 at Time 1) is interesting in and of itself, and so is its effect on the mean difference testing. Two things are important to note with respect to the bidirectional uniform DIF exhibited by Items 2, 3, and 4 and its effect on the mean difference analysis. Recall that the test of latent mean difference was conducted in the full scalar invariance model (where all 9 items contributed to the factor mean) and also in a partial scalar invariance model (where only the six scalar invariant items contributed to the factor mean). First, a greater latent difference between Time 1 and Time 2 factor scores was found when only the scalar invariant items contributed to the mean. The reason for this greater effect is that, on average, latent scores tended to increase from Time 1 to Time 2, whereas the means for two of the non-invariant items (Items 2 and 3) decreased. Because these two items would have contributed to the latent mean difference in the opposite direction of many of the other items, the latent mean difference when these non-invariant items did *not* contribute to the estimation of this difference was greater than if they contributed to the estimation of the latent mean difference. The same effect of the bidirectional change in Items 2, 3, and 4 was present when observed means were analyzed. In fact, the combined decrease in the observed means of Items 2 and 3 from Time 1 to Time 2 equaled exactly the increase in the observed mean of Item 4 across time. This bidirectional change completely masked the effect of the scalar non-invariant Items 2, 3, and 4, resulting in nearly the same observed mean difference (.02 vs. 03) regardless of whether these items contributed to the total mean scores for Time 1 and Time 2.

Fortunately, the uniform DIF discussed above was present only in three of the PES items, and the mean residuals were not very large considering the PES response scale. If uniform DIF were present in more items and the mean residuals were large, neither the latent, nor the observed mean differences would have been meaningful. On the one hand, when computing the latent mean difference test, only scalar invariant items contribute to the mean. This may appear as a protection against the bias of scalar non-invariant items on the factor mean, but if many of the items on a given measure are scalar non-invariant, the factor mean may no longer represent the construct well enough for any substantive conclusions to be made. On the other hand, when observed scores are analyzed, the effects of large uniform DIF could seriously distort the results, and potentially lead the researcher to a wrong conclusion. For example, an increase in observed scores from Time 1 to Time 2 could be attributed to the latent construct under investigation when the effect could be due simply to uniform DIF of a large number of items in the same direction. In addition, it is important to note that uniform DIF can only be detected when some, not all, items display uniform DIF. One would be unable to distinguish true change on the construct (alpha change) from the effects of an extraneous variable, if all items were equally affected by this extraneous variable.

Implications for the PES and the Psychological Entitlement Domain

The current study has several important implications for the domain of psychological entitlement and the use of the PES to measure psychological entitlement over time. First, I empirically tested whether a widely used measure of psychological entitlement (PES; W. K. Campbell et al., 2004) displayed longitudinal measurement invariance. Assessing the invariance of the PES was important because often it is assumed that respondents conceptualize the construct the same way across time and that the measure functions the same way across time, when neither may be true. Given that no prior research has studied the PES in terms of measurement invariance, the current study provided further construct validity evidence for the PES by establishing full configural, metric, and partial scalar invariance for the PES with practically insignificant bias of the three scalar non-invariance items over 1.5 years. Thus, the PES can be used as a reliable measure of psychological entitlement in longitudinal research with college student populations, in that the validity of inferences regarding levels of entitlement across time based on scores collected with the PES on such a sample has been empirically supported.

After establishing measurement invariance for the PES, the stability of psychological entitlement was assessed by examination of both rank-order and mean-level consistency over two time points. It is essential to note that this is the first study to assess the stability of psychological entitlement over a long period of time (1.5 years). In this regard, the study provides empirical support to previous anecdotal evidence that found, *on average,* entitlement to be relatively stable over time (e.g., W. K. Campbell et al., 2004). Importantly, this study lends evidence that this is the case not only for short periods of time (e.g., one to two months) but also for longer periods of time (1.5 years). However, as mentioned previously, a construct's stability cannot be accurately and completely determined by examining only mean-level change *or* rank-order consistency. Both indicators need to be considered in assessing the presence of change. This necessity was clearly illustrated in the case of psychological entitlement, in that no change was found on average, but individual change *was* present across time (as indicated by a moderate rank-order consistency coefficient). Thus, only when both mean-level change and rank-order consistency are examined does one fully understand the type of change taking place and the extent to which the construct under study remains stable over time. Put simply, when assessing the change in psychological entitlement for college students, it appeared that psychological entitlement was stable over time based on the lack of change on average. Examination of individual change, however, through the moderate rank-order consistency and especially by plotting the change trajectories for a sample of participants over time revealed the *differential* change in psychological entitlement taking place among students. Some students increased, others decreased, and still others retained approximately the same levels of psychological entitlement over 1.5 years. This is a very important finding because it challenges the claims in the literature stating that psychological entitlement is stable over time, which categorizes the construct as a trait. The results discussed above, however, clearly indicate that while psychological entitlement may be stable *on average,* individuals change *differentially* on the construct over time.

Related to this point, the current study opens a whole new arena for future research on the stability of entitlement. If this finding is replicated, it would be interesting to examine possible explanations as to how and why levels of psychological entitlement change for some people but not for others. That is, what predictors could be added to the model to explain why some students increase in psychological entitlement, whereas others decrease, or remain fairly consistent? Moreover, why is it that over short periods of time (e.g., one and two months) test-retest consistency is high (.72 and .70; W. K. Campbell et al., 2004), but over long periods of time (1.5 years in the current study) it is only moderate (.61). Is psychological entitlement one of the individual differences that Fraley and Roberts (2005) and Veenhoven (1994) described as highly consistent over short periods of time but less stable over longer periods of time? Obviously, answering this question would shed light on possible interventions to decrease entitlement.

LMACS Analysis as a Unified Approach to Assessing Construct Stability

Given the importance of empirically testing the assumption of measurement invariance in longitudinal research and the examination of both average and individual change over time, the above didactic demonstration hopefully highlights the advantages of LMACS analysis over more traditional approaches to assessing construct stability over two time points. First, unlike traditional statistical techniques, such as *t*-test or ANOVA, which *assume* measurement invariance, I demonstrated how LMACS analysis allows the researcher to empirically test this assumption through the multi-step process of invariance testing. As mentioned previously, establishing measurement invariance for a given instrument ensures that respondents conceptualize consistently the construct of interest across testing occasions and also that the instrument measures the construct the same way across time. By testing for measurement invariance, the researcher collects evidence for the presence of these highly desired properties of scores in longitudinal research.

Second, once measurement invariance is established, I demonstrated how LMACS analysis allows for the tests of stability to be conducted at the latent level. Conducting these tests at the latent level has several advantages, the most important being that a latent modeling approach may be a more appropriate method of modeling unobservable constructs. This advantage is most apparent when analyzing items that are complex in nature (i.e., items representing multiple constructs). If items are truly multidimensional, analyzing the data at the latent level allows the researcher to "disentangle" the variance due to the construct of interest from the variance due to other constructs that may or may not be of interest. Analyzing multidimensional items using observed variable techniques would prevent the researcher from better understanding the structure of the scores and could lead to erroneous conclusions regarding the construct(s) of interest. Alternatively, if the items are truly unidimensional, the data could easily be analyzed using observed variable techniques. However, even with simple structure, analyzing the data at the latent level is advantageous. A latent analytical approach accounts for both measurement error in the indicators and any systematic error variance in corresponding items across time, thus yielding more accurate parameter estimates and a more powerful analysis overall.

Finally, LMACS analysis allows all of the analyses described above to be conducted within a single analytical framework. Once configural and metric invariance have been established, the researcher can assess rank-order consistency by examination of the factor correlation across time, which is readily available in the output. Similarly, once scalar invariance is established, one can compute and interpret the latent mean difference test using the latent means, variances and covariances across time from the output. As such, LMACS analysis represents a convenient unified framework for conducting sound construct stability research with numerous advantages over traditional statistical techniques.

Footnotes

[1] Regardless which method is used to scale the factor, conclusions regarding measurement invariance remain the same. Different scaling methods offer different interpretation of parameter estimates, but model fit and latent mean difference effect size are the same across scaling methods (see Little et al., 2007).

[2] The formula used to compute the latent mean difference effect size is:

$$\text{Latent}\, d = \frac{(\kappa_{2j} - \kappa_{1j})}{\sqrt{\psi_{\text{pooled}}}}\text{, where } \sqrt{\psi_{\text{pooled}}} = \sqrt{\frac{n_1\psi_{1jj} + n_2\psi_{2jj}}{(n_1 + n_2)}} \times \sqrt{2(1 - r_{12})}$$

[3] The formula used to compute the observed mean difference effect size is:

$$\text{Observed}\, d = \frac{(\mu_{2j} - \mu_{1j})}{SD_{pooled}}\text{, where } SD_{pooled} = \sqrt{\frac{(n_1 - 1)(s_1^{\;2}) + (n_2 - 1)(s_2^{\;2})}{(n_1 + n_2 - 2)\sqrt{2(1 - r_{12})}}}$$

Appendix

Please respond to the following items using the number that best reflects your own beliefs. Please use the following 7-point scale:

1 = strong disagreement.

2 = moderate disagreement.

3 = slight disagreement.

4 = neither agreement nor disagreement.

5 = slight agreement.

6 = moderate agreement.

7 = strong agreement.

1. I honestly feel I'm just more deserving than others.
2. Great things should come to me.
3. If I were on the Titanic, I would deserve to be on the *first* lifeboat!
4. I demand the best because I'm worth it.
5. I do not necessarily deserve special treatment.
6. I deserve more things in my life.
7. People like me deserve an extra break now and then.
8. Things should go my way.
9. I feel entitled to more of everything.

References

Adams, J. S. (1965). Inequity in Social Exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (pp. 267-299). New York, NY: Academic Press.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders (3rd ed., text revision).* Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mentaldisorders (4th ed., text revision).* Washington, DC: American Psychiatric Association.

Baranik, L. E., Barron, K. E., & Finney, S. J. (2010). Examining specific versus general measures of achievement goals. *Human Performance*, *23,* 155-172.

Bontempo, D. E., & Hofer, S. M. (2006). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 153-175). New York: Oxford University Press.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order factor structure: A walk through the process. *Structural Equation Modeling*, *13,* 287-321.

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83,* 29-45.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis

(LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421-483.

Chan, D. (2003). Data analysis and modeling longitudinal processes. *Group & Organization Management*, *28,* 341-365.

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

Chowning, K., & Campbell, N. J. (2009). Development and validation of a measure of academic entitlement: Individual differences in students' externalized responsibility and entitled expectations. *Journal of Educational Psychology, 101,* 982-997.

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. *Personality and Individual Differences, 5*, 11-25.

del Rosario, P. M., & White, R. M. (2005). The Narcissistic Personality Inventory: Test-retest stability and internal consistency. *Personality and Individual Differences, 39*, 1075-1081.

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods*, *3,* 412-423.

Dimitrov, D. M. (2006). Comparing groups on latent variables: A structural equation modeling approach. *WORK: A Journal of Prevention, Assessment & Rehabilitation*, *26,* 429-436.

Elliott, J. C. (2010). *Humility: Development and analysis of a scale* (Doctoral dissertation). Retrieved from http://trace.tennessee.edu/cgi/viewcontent.cgi?article=1887&context=utk_graddiss

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*, 509-529.

Feather, N. T. (2003). Distinguishing between deservingness and entitlement: Earned outcomes versus lawful outcomes. *European Journal of Social Psychology, 33,* 367-385.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.

Fossati, A., Borroni, S., & Maffei, C. (2008). Psychometric properties of the Italian version of the Narcissistic Personality Inventory. *Rivista di Psicologia Clinica, 1,* 94-112.

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112,* 60-74.

Freud, S. (1916). Einige Charaktertypen aus der psychoanalytischen Arbeit. Imago. *Zeitschrift für Anwendung der Psychoanalyse auf die Geisteswissenschaften, 4*, 317-336.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *The Journal of Applied Behavioral Science, 12,* 133-157.

Greenberger, E., Lessard, J., Chen, C., & Farruggia, S. P. (2008). Self-entitled college students: Contributions of personality, parenting, and motivational factors. *Journal of Youth and Adolescence, 37,* 1193-1204.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development, 30,* 91-105.

Harvey, P., & Harris, J. K. (2010). Frustration-based outcomes of entitlement and the influence of supervisor communication. *Human Relations, 63,* 1639-1660.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait-state distinction for the structural modeling of developmental change. *Child Development, 58,* 93-109.

Horney, K. (1950). *Neurosis and human growth: The struggle toward self-realization.* New York, NY: Harper.

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

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6, 1-55.

Huseman, R. C., Hatfield, J. D., & Miles, E. W. (1987). A new perspective in equity theory: The equity sensitivity construct. *Academy of Management Review, 12,* 222-234.

Jacobson, E. (1959). The “exceptions”: An elaboration of Freud’s character study. *The Psychoanalytic Study of the Child, 14*, 135-154.

Jöreskog, K. G., & Sörbom, D. G. (2006). LISREL 8.80 [Computer software]. Lincolnwood, IL: Scientific Software.

Kline, R. B. (2011). *Principles and practice of structural equation modeling, third edition.* New York: Guilford Press.

Kopp, J. P., Zinn, T. E., Finney, S. J., & Jurich, D. P. (2011). The development and evaluation of the Academic Entitlement Questionnaire. *Measurement and Evaluation in Counseling and Development, 44*(2), 105-129.

LexisNexis. Harley Davidson, Inc. (2010). Retrieved 2011, July 23 from LexisNexis Academic database.

Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal analyses of human development* (pp. 121-147). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1*, 5-34.

Marsh, H. W., Hau, K. T., & Greyson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McCardle (Eds.), *Psychometrics. A Festschrift to Roderick P. McDonald* (pp. 275-340). Mahwah, NJ: Erlbaum.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

McCrae, R. R., Terracciano, A., & Khoury, B. (2007). *Dolce far niente*: The positive psychology of personality stability and invariance. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 176-188). New York: Oxford University Press.

McDonald, R. P. (1999). *Test theory: A unified treatment.* New Jersey: Lawrence Erlbaum Associates, Inc.

Meyer, G. J., & Shack, J. R. (1989). Structural convergence of mood and personality: Evidence for old and new direction. *Journal of Personality and Social Psychology*, *57*, 691-706.

Miles, E. W., Hatfield, J. D., & Huseman, R. C. (1989). The equity sensitivity construct: Potential implications for worker performance. *Journal of Management, 15,* 581-588.

Millsap, R. E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In R. Cudeck & R. MacCallum (Eds.) *Factor analysis at 100* (pp. 131-152) Mahwah, NJ: Lawrence Erlbaum.

Nesselroade, J. R. (1986). Some implications of the trait-state distinction for the study of development over the life span: The case of personality. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (pp.163-189). Hillsdale, NJ: Erlbaum.

Ozar, D. J., & Gjerde, P. F. (1989). Patterns of personality consistency and change from childhood through adolescence. *Journal of Personality, 57*, 483-507.

Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods, 7,* 27-65.

Pryor, L. R., Miller, J. D., & Gaughan, E. T. (2008). A comparison of the Psychological Entitlement Scale and the Narcissistic Personality Inventory's Entitlement Scale: Relations with general personality traits and personality disorders. *Journal of Personality Assessment, 90,* 517-520.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*, 890-902.

Rensvold, R. B., & Cheung, G. W. (2001). Testing for metric invariance using structural equation models: Solving the standardization problem. In C. A. Schriesheim & L. L. Neider (Eds.), *Research in managemnent: Vol. 1. Equivalence in measurement* (pp. 25-50). Greenwich, CT: Information Age.

Reuterberg, S., & Gustafsson, J. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement, 52*, 795-811.

Riordan, C. M., Richardson, H. A., Schaffer, B. S., & Vandenberg, R. J. (2001). Alpha, beta, and gamma change: A review of past research with recommendations for new directions. In C. A. Schriesheim & L. L. Neider (Eds.), *Research Management* (pp. 51-98). Greenwich, CT: Information Age.

Ronningstam, E., Gunderson, J., & Lyons, M. (1995). Changes in pathological narcissism. *American Journal of Psychiatry*, *152*, 253-257.

Sass, D. A. (2011). Testing measurement invriance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29,* 347-363.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.). *Latent variable analysis* (pp. 299-419). Thousand Oaks, CA: Sage.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and limitations. *Human Resource Management Review, 18,* 210-222.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58,* 935-943.

Spielberger, C. D. (1972). Conceptual and methodological issues in anxiety research. In C. D. Spielberger (Ed.). *Anxiety: Current trends in theory and research* (Vol. 2, pp. 481-493). New York: Academic Press.

Spielberger, C. D. (1975). Anxiety: State-trait process. In C. D. Spielberger & I. G. Sarason (Eds.). *Stress and anxiety* (Vol. 1, pp. 115-143). Washington, DC: Hemisphere.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Consulting Psychologist Press.

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

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57,* 5-19.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *A Second Course in Structural Equation Modeling* (pp. 119-169). Greenwich, CT: Information Age.

Usala, P. D., & Hertzog, C. (1991). Evidence of differential stability of state and trait anxiety in adults. *Journal of Personality and Social Psychology, 60,* 471-479.

U.S. Department of Education. (2006). *A Test of Leadership: Charting the Future of American Higher Education.* Report of the Commission Appointed by Secretary of Education Margaret Spellings. Washington, DC: Author.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Veenhoven, R. (1994). Is happiness a trait? Tests of the theoty that a better society does not make people any happier. *Social Indicators Research, 32,* 101-160.

Watson, P. J., & Biderman, M. D. (1993). Narcissistic Personality Inventory factors, splitting, and self-consciousness. *Journal of Personality Assessment, 61,* 41-57.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.) *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*, 10-18.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention* (pp. 281-324). Washington, DC: American Psychological Association.

Zitek, E. M., Jordan, A. H., Monin, B., & Leach, F. R. (2010). Victim entitlement to behave selfishly. *Journal of Personality and Social Psychology, 98*, 245-255.

Table 1

Sample Size across Testing Occasions

	Number of students administered the PES	Missing Data	Complete Data	Outliers
Time 1	3749	35	3714	29
Time 2	3346	33	3313	31
Match	NA	NA	2267	2195

Note. Of the 3749 students tested at Time 1, only 2442 students (65%) were eligible to be tested at Time 2. In addition, 402 (12%) of the 3346 students tested at Time 2 were not tested at Time 1 because they were transfer students. Thus, the 2267 reflects only those students who completed between 45 and 70 credit hours by the second semester of their sophomore year.

Table 2

Correlations and Descriptive Statistics for the PES Items at Time 1 and Time 2

Item	1_{T1}	2_{T1}	3_{T1}	4_{T1}	5_{T1}	6_{T1}	7_{T1}	8_{T1}	9_{T1}	1_{T2}	2_{T2}	3_{T2}	4_{T2}	5_{T2}	6_{T2}	7_{T2}	8_{T2}	9_{T2}
1_{T1}	--																	
2_{T1}	.526	--																
3_{T1}	.497	.493	--															
4_{T1}	.435	.485	.525	--														
5_{T1}	**.228**	**.181**	**.261**	**.201**	--													
6_{T1}	.482	.477	.475	.497	**.171**	--												
7_{T1}	.442	.417	.409	.364	**.130**	.550	--											
8_{T1}	.451	.515	.458	.467	**.164**	.523	.563	--										
9_{T1}	.566	.470	.570	.481	**.277**	.574	.533	.586	--									
1_{T2}	.468	.344	.340	.300	.182	.345	.310	.336	.394	--								
2_{T2}	.313	.479	.334	.340	.152	.357	.288	.373	.352	.584	--							
3_{T2}	.295	.285	.451	.312	.172	.316	.249	.299	.352	.537	.519	--						
4_{T2}	.259	.291	.310	.479	.130	.322	.239	.304	.289	.467	.532	.540	--					
5_{T2}	.183	.137	.189	.133	.228	.160	.100	.135	.192	**.297**	**.245**	**.312**	**.250**	--				
6_{T2}	.300	.328	.306	.294	.147	.439	.317	.339	.342	.502	.547	.491	.492	**.226**	--			
7_{T2}	.290	.263	.276	.243	.152	.307	.416	.320	.312	.462	.432	.456	.401	**.213**	.585	--		
8_{T2}	.288	.342	.283	.315	.126	.328	.301	.453	.348	.482	.574	.487	.518	**.244**	.581	.583	--	
9_{T2}	.348	.292	.337	.269	.189	.345	.291	.348	.421	.611	.503	.578	.491	**.355**	.606	.549	.593	--
Mean	2.70	3.72	2.57	3.19	2.73	3.19	3.21	3.23	2.37	2.76	3.56	2.50	3.42	2.71	3.18	3.26	3.33	2.39
SD	1.50	1.68	1.50	1.69	1.37	1.48	1.57	1.56	1.38	1.59	1.71	1.55	1.76	1.45	1.54	1.63	1.55	1.44
Skewness	0.45	-0.11	0.73	0.29	0.85	0.05	0.15	0.11	0.71	0.48	0.04	0.83	0.19	0.85	0.20	0.20	0.06	0.81
Kurtosis	-0.82	-0.90	-0.11	-0.91	0.52	-0.80	-0.85	-0.80	-0.36	-0.81	-0.093	-0.06	-0.95	0.36	-0.69	-0.83	-0.69	-0.02

Note. Responses range from 1 (strong disagreement) to 7 (strong agreement). Item 5 (*r*s in bold) was reverse coded prior to these statistics being computed.

T1 = Time 1. T2 = Time 2. N = 2195.

Table 3

Unstandardized (Standardized) Factor Pattern Coefficients, Error Variances, and Variance Explained for the 9 PES Items in the Independent One-factor CFA Models at Time 1 and Time 2

	Pattern Coefficients		Error Variance		R^2	
Item	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
1	0.96 (.69)	1.15 (.72)	1.01 (.52)	1.21 (.48)	.48	.52
2	1.05 (.68)	1.23 (.72)	1.08 (.54)	1.40 (.48)	.46	.52
3	0.97 (.70)	1.10 (.71)	0.96 (.51)	1.19 (.50)	.49	.50
4	1.02 (.65)	1.18 (.67)	1.04 (.57)	1.72 (.55)	.43	.45
5	0.37 (.29)	0.53 (.37)	0.47 (.92)	1.75 (.86)	.08	.14
6	0.99 (.73)	1.16 (.75)	1.02 (.47)	1.04 (.44)	.53	.56
7	0.97 (.67)	1.11 (.68)	1.98 (.56)	1.41 (.53)	.44	.47
8	1.04 (.72)	1.17 (.75)	1.03 (.47)	1.05 (.43)	.53	.57
9	1.00 (.79)	1.14 (.79)	1.00 (.38)	0.78 (.38)	.62	.62

Note. All unstandardized pattern coefficients were statistically significant (p <.05) except Item 9 which was fixed to 1 and could not be tested for significance; standardized pattern coefficients are in parentheses; standardized error variances (proportion of variance in the indicator not explained by the factor) are in parentheses; R^2 = proportion of variance in the indicator accounted for by the factor.

N = 2195.

Table 4

Fit Statistics for the Models Testing Measurement Invariance across Time 1 and Time 2

Model	MLχ^2	*df*	*p*-value	$\Delta\chi^2$	Δdf	Δp-value	SRMR	CFI	RMSEA
Configural Invariance	1055.24	125	< 0.001	--	--	--	0.032	0.95	0.06
Metric Invariance	1067.67	133	< 0.001	12.43	8	0.133	0.033	0.95	0.06
Scalar Invariance	1172.71	141	< 0.001	105.04	8	< 0.001	0.033	0.95	0.06
Equal Factor Means	1173.85	142	< 0.001	1.14	1	0.286	0.033	0.95	0.06
Scalar Invariance—i4 free	1123.76	140	< 0.001	56.09	7	< 0.001	0.033	0.95	0.06
Scalar Invariance—i2, i4 free	1091.64	139	< 0.001	23.97	6	0.001	0.033	0.95	0.06
Scalar Invariance—i2, i3, i4 free	1078.15	138	< 0.001	10.48	5	0.063	0.033	0.95	0.06
Equal Factor Means—i2, i3, i4 free	1081.01	139	< 0.001	2.86	1	0.091	0.033	0.95	0.06

Note. $\Delta\chi^2$ = chi-square difference; Δdf = degrees of freedom difference; Δp-value = probability value for the $\Delta\chi^2$ test; SRMR = standardized root mean square residual; CFI = comparative fit index; RMSEA = root mean square error of approximation. The $\Delta\chi^2$ tests were conducted to compare more constrained models to less constrained models in order to establish invariance for a set of parameters: the metric invariance model was compared to the configural invariance model, all scalar invariance models were compared to the metric invariance model; $\Delta\chi^2$ tests were also conducted to test the latent mean difference for significance by comparing the fit of both of the equal factor means models to the fit of the scalar model directly above it.

N = 2195.

Table 5

Mean Differences of Latent and Observed PES Scores Based on All vs. Only Scalar Invariant Items

Latent Estimates	All Items	Only Invariant Items
Latent mean difference	0.020	0.040
Latent mean difference effect size	0.012	0.041
Observed Estimates	All Items	Only Invariant Items
Time 1 observed mean	2.991	2.906
Time 2 observed mean	3.012	2.939
Observed mean difference	0.022	0.034
Observed mean difference effect size	0.021	0.032

Note. Invariant items are Items 1, 5, 6, 7, 8, and 9. Observed means range from 1 to 7. None of the mean differences were significant at $p < .05$. Latent mean difference effect size values indicate standard deviation units by which the latent mean at Time 2 is greater than that at Time 1, taking into account the correlation of scores across time.[2] Observed mean difference effect size values indicate standard deviation units by which the observed mean at Time 2 is greater than that at Time 1, taking into account the correlation of scores across time.[3]

$N = 2195$.

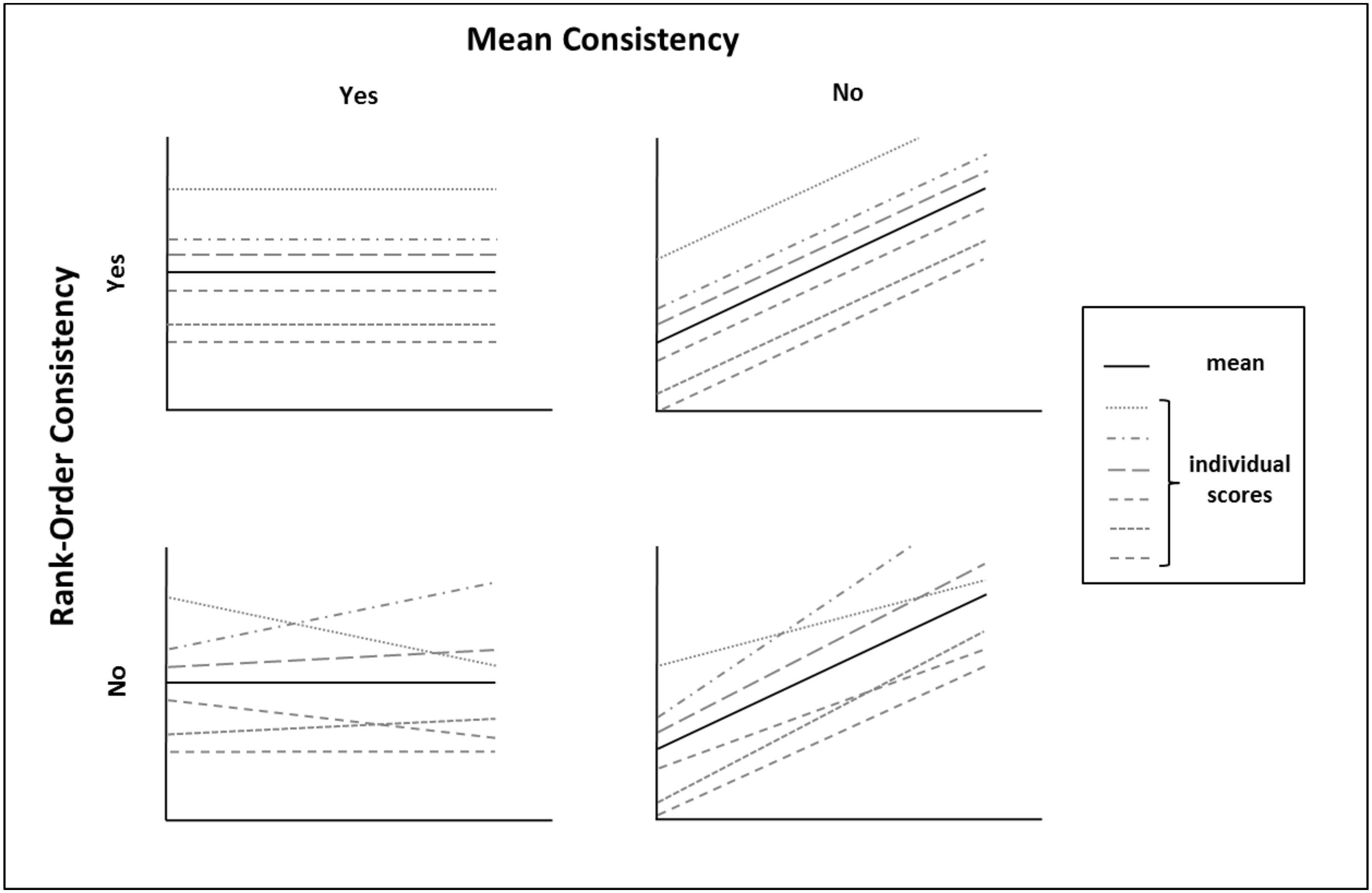

Figure 1. Rank-order consistency vs. mean consistency. The four scenarios above illustrate the results when crossing rank-order consistency (yes vs. no) with mean consistency (yes vs. no). Dashed lines represent individual change over time.

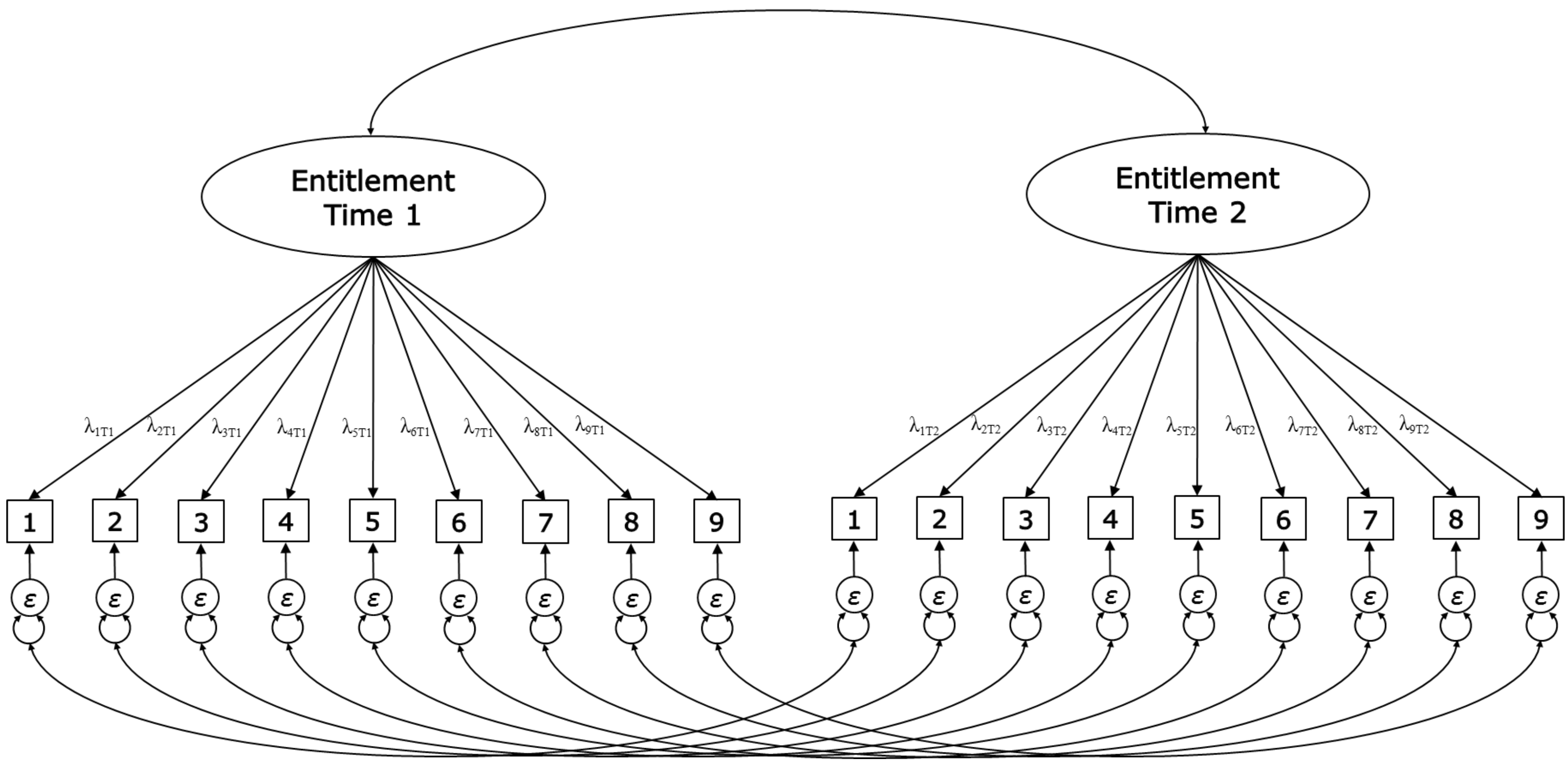

Figure 2. Configural invariance model with correlated factors at Time 1 and Time 2 and correlated error variances for pairs of items.

All parameters are freely estimated.

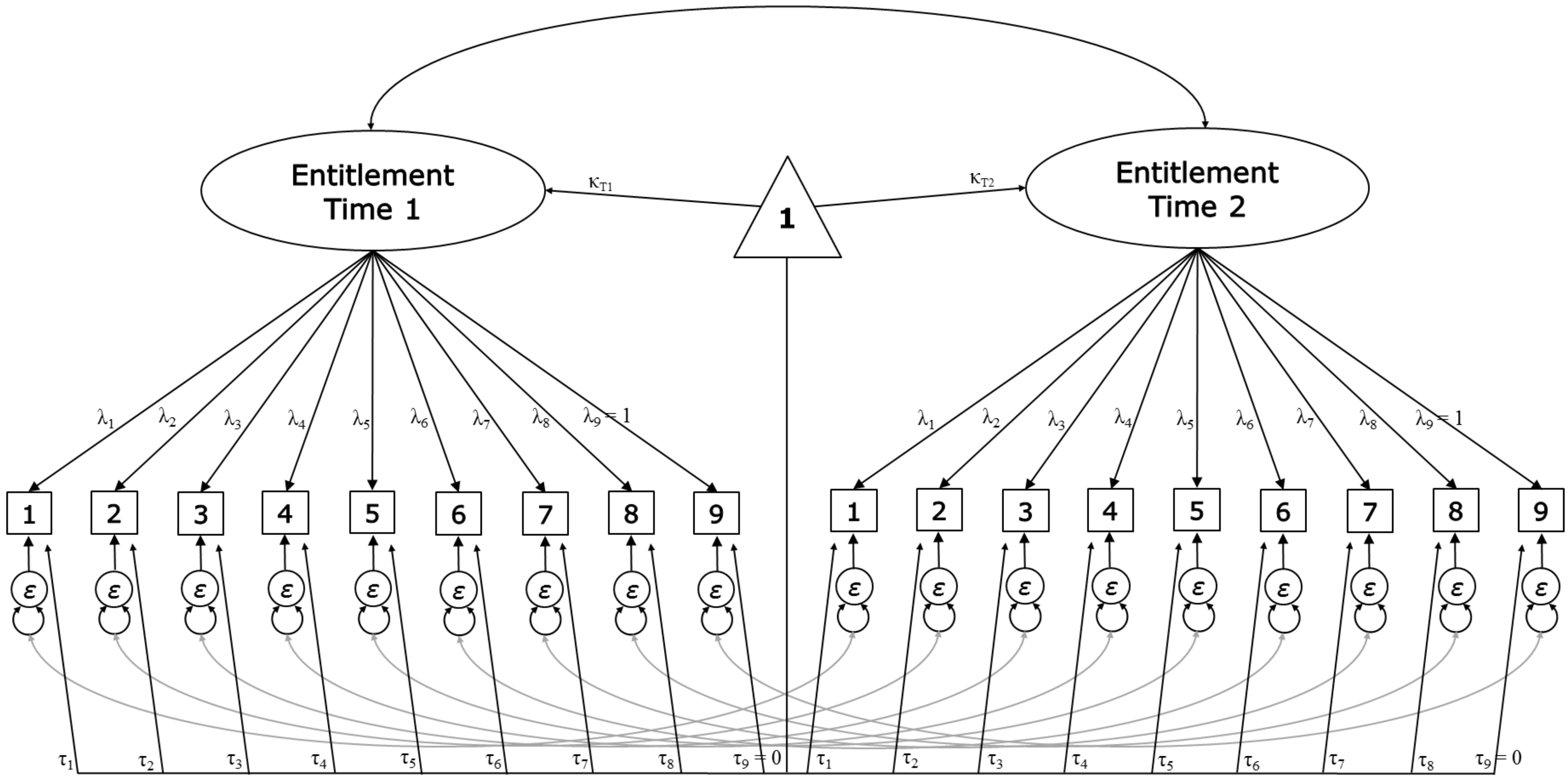

Figure 3. Metric and scalar invariance model, assuming configural invariance. Corresponding factor pattern coefficients (λ) and intercepts (τ) for items from Time 1 and Time 2 are constrained to be equal, thus the lack of time subscripts. Latent means (κ_{T1} and κ_{T2}) are estimated.

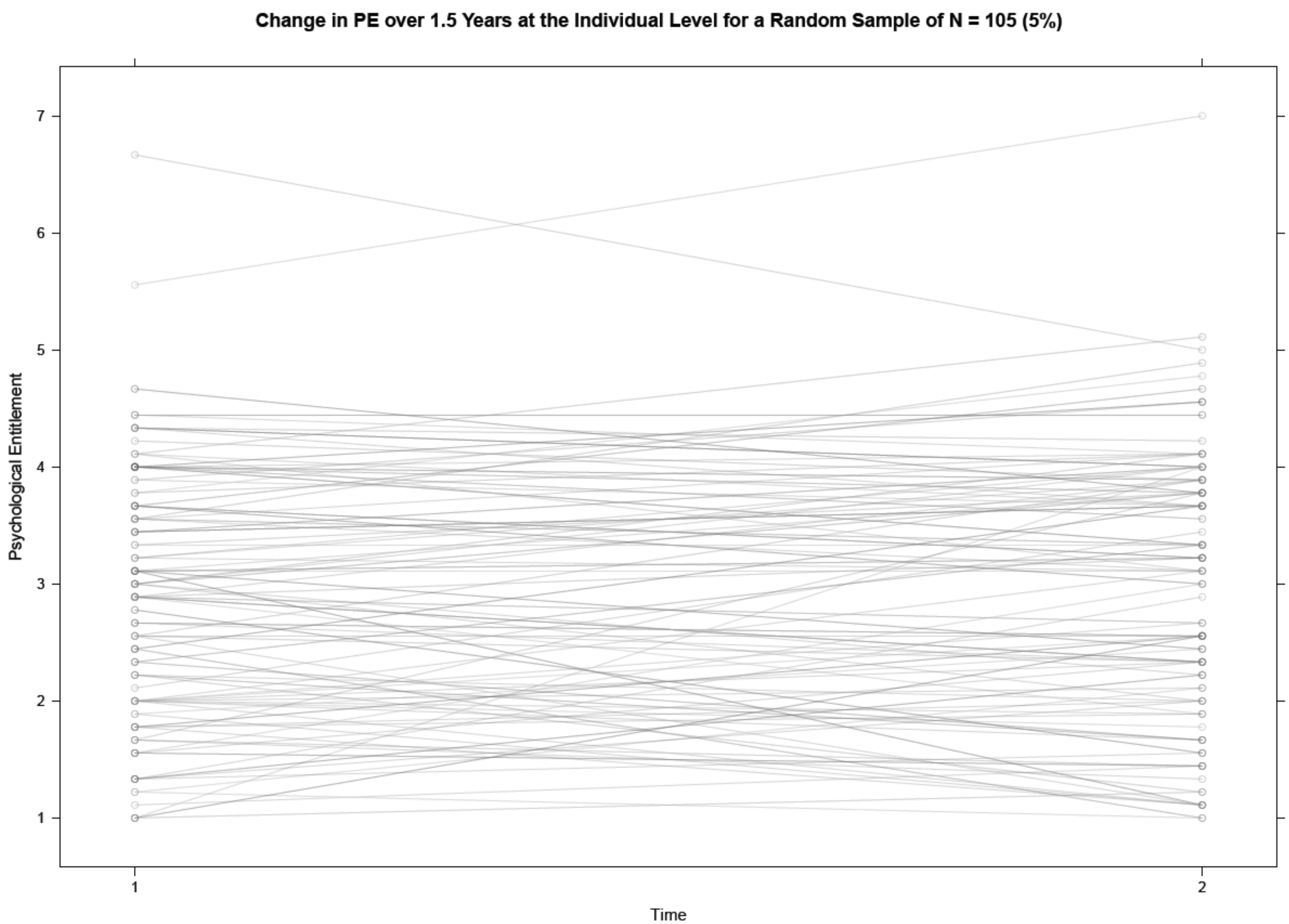

Figure 4. Individual change trajectories for a random sample of 105 participants (5%) from Time 1 to Time 2. The graph shows visually the differential change taking place at the individual level: some participants increase on entitlement, others decrease, and still others remain the same across time.