Stability and Change of Personality Traits, Self-Esteem, and Well-Being: Introducing the Meta-Analytic Stability and Change Model of Retest Correlations

Ivana Anusic
Michigan State University

Ulrich Schimmack
University of Toronto Mississauga

The stability of individual differences is a fundamental issue in personality psychology. Although accumulating evidence suggests that many psychological attributes are both stable and change over time, existing research rarely takes advantage of theoretical models that capture both stability and change. In this article, we present the Meta-Analytic Stability and Change model (MASC), a novel meta-analytic model for synthesizing data from longitudinal studies. MASC is based on trait–state models that can separate influences of stable and changing factors from unreliable variance (Kenny & Zautra, 1995). We used MASC to evaluate the extent to which personality traits, life satisfaction, affect, and self-esteem are influenced by these different factors. The results showed that the majority of reliable variance in personality traits is attributable to stable influences (83%). Changing factors had a greater influence on reliable variance in life satisfaction, self-esteem, and affect than in personality (42%–56% vs. 17%). In addition, changing influences on well-being were more stable than changing influences on personality traits, suggesting that different changing factors contribute to personality and well-being. Measures of affect were less reliable than measures of the other 3 constructs, reflecting influences of transient factors, such as mood on affective judgments. After accounting for differences in reliability, stability of affect did not differ from other well-being variables. Consistent with previous research, we found that stability of individual differences increases with age.

Keywords: meta-analysis, stability, personality, life satisfaction, self-esteem

Supplemental materials: http://dx.doi.org/10.1037/pspp0000066.supp

The stability and change of personality and well-being has been a controversial issue in personality psychology, just like the nature–nurture debate has been a source of heated debate. The nature–nurture debate has quieted down because quantitative behavioral genetics studies provide evidence that most personality characteristics are influenced by genetic factors and environmental factors. Moreover, behavioral genetics models have been used to quantify the contribution of genes and environment to various personality characteristics, and the question is no longer whether genes or environment matter but how much genes and environment matter.

We propose that controversies about stability and change of personality would equally benefit from quantitative models of stability and change. Ample evidence shows that personality characteristics are neither fixed nor rapidly changing from moment to moment (Caspi & Roberts, 2001; Conley, 1984; Ferguson, 2010; Roberts & DelVecchio, 2000; Terracciano, McCrae, & Costa, 2010; Trzesniewski, Donnellan, & Robins, 2003). Thus, personality traits like the Big Five and personality characteristics like well-being are stable, and they change over time (Roberts, Wood, & Caspi, 2008; Specht et al., 2014). To move research on stability and change forward, it is necessary to quantify the degree of stability and change.

This focus on quantifying the extent to which personality characteristics change or remain stable over time is consistent with recent calls in psychology to move from testing of the null hypothesis to parameter estimation (Cumming, 2014). The null hypothesis is particularly uninformative for questions about variance components. The probability that personality has zero stability or never changes is very small, and empirical data can never prove that it is zero. Thus, it is more productive to quantify the amount of stability and change in personality traits and other individual differences rather than simply test whether change happens.

Stability and change are most commonly quantified by test–retest correlations. Interpretation of retest correlations is made difficult by two methodological problems. First, retest correlations are attenuated by random measurement error. Thus, observed retest correlations underestimate stability and overestimate change. One solution to this problem is to use internal consistency as an estimate of reliability. However, internal consistency can be in-
flated by systematic measurement error and attenuated by item heterogeneity. Furthermore, it is not possible to calculate internal consistency for single-item measures. A more direct way to estimate random measurement error is to use retest correlations over shorter time periods. However, retest correlations can overestimate random measurement error if the construct actually changes during the retest interval (e.g., well-being may actually change in a 1-month period).

The second obstacle is that the amount of change that can be observed depends on the time interval between repeated measurements. Change is unlikely to occur during a short time interval of a few weeks, but the likelihood of personality change increases with time (e.g., an extravert is likely to be an extravert a month later, but may become an introvert in 20 years). As a result, a simple retest correlation does not quantify stability or change because it is dependent on the time interval between retests. Thus, a single retest correlation provides insufficient information about the amount of stability or change of a personality characteristic.

To estimate stability and change, it is necessary to create a model that takes into account both stability and change, fits observed retest correlations to the model, and interpret the model parameters. In a seminal yet often neglected article, Conley (1984) made a first attempt to quantify stability and change of intelligence tests, personality traits, and self-evaluations (self-esteem, life satisfaction). He conducted a meta-analysis of retest correlations and plotted retest correlations as a function of the retest interval. This resulted in a nonlinear decay function where retest correlations became smaller as retest intervals lengthened. He then fitted Heise’s (1969) autoregressive state model to the retest correlations and found that intelligence was more stable than personality traits (extraversion and neuroticism), which were more stable than self-evaluations (self-esteem and life satisfaction). The rate of change of variance at the between-person level for intelligence was less than 2% per year. The rate of change for extraversion and neuroticism was 4% per year, and the rate of change for self-evaluations was 12%. As change accumulates, these differences lead to much larger differences over longer time periods. For example, only 18% of the between-person variance in intelligence changes over a 10-year period. For extraversion and neuroticism, 33% of the variance changes over a 10-year interval, and 71% of the variance in self-evaluations changes over a 10-year period.

Conley’s (1984) findings provided the first empirical evidence that some personality characteristics are more stable than others. The results also provided the first evidence that personality traits and well-being are not merely fluctuating around a fixed set point, but can change for extended periods of time. Unfortunately, this important finding was neglected in personality theories of well-being that exaggerated the importance of stable dispositions and ignored the importance of environmental factors (see Lucas, 2007, for a review).

Although Conley made an important first step toward a quantitative model of stability and change, his model had some limitations. The most serious limitation was the choice of a simple autoregressive function to model change. This model assumes that all factors that contribute to variation in personality are changing. As change accumulates, retest correlations will eventually asymptote toward zero. For example, the finding that retest correlations for self-evaluations changed at the rate of 12% per year (i.e., 88% is stable year-to-year) suggests that after 20 years only 8% (.88 \(^{20}\)) of the between-person variance in self-evaluations is retained, whereas 92% of variance (100 – 8) has changed. This prediction is inconsistent with actual retest correlations of life satisfaction in more recent panel studies that remain above .30 over retest intervals of 20 years or more (Lucas & Donnellan, 2007; Lucas & Donnellan, 2012; Schimmack, Krause, Wagner, & Schupp, 2010; Schimmack & Lucas, 2010).

The high retest stability of individual differences over retest intervals longer than two decades suggests that some factors that produce variation in personality across individuals are stable. Similar conclusions can be drawn from findings of longitudinal behavioral genetics studies that have found substantial genetic influences on personality traits and well-being across the life span (Briley & Tucker-Drob, 2014; Kandler et al., 2010; Nes, Røysamb, Tambs, Harris, & Reichborn-Kjennerud, 2006). However, stable influences are not solely based in our biology. Evidence is accumulating that certain features of the environment can also produce stability in personality (Bleidorn, Kandler, & Caspi, 2014; Briley & Tucker-Drob, 2014).

To allow for the possibility that stable influences on personality exist, Kenny and Zautra (1995) introduced the trait-state-error model (TSE). The state and error factors in the TSE model are equivalent to Conley’s (1984) model, with state factor reflecting changing influences on personality. The major advantage of the TSE model is that it includes a stable trait factor. The stable trait factor in the model essentially allows for an above-zero asymptote in retest correlations. For example, if retest correlations over periods of 20 years no longer decrease but hover around \(r = .30\), the data suggest that 30% of between-person variance in a personality characteristic is influenced by a stable factor that never changes. Because this factor has a persistent influence on personality at every time point, all measures share a minimum of 30% of the variance independent of the time interval between measures.

At present, only relatively few studies have used the TSE model to quantify stability and change of personality characteristics. The main reason is that the model requires a minimum of four occasions of measurement, relatively long time periods, and large samples to provide meaningful parameter estimates; that is, parameter estimates with relatively narrow confidence intervals (CIs) around them. For example, Anusic, Lucas, and Donnellan (2012) conducted a short-term panel study that included measures of the Big Five and life satisfaction. A total of 237 students completed eight waves of measurement over a 2-month interval. The study revealed that personality traits were more stable than life satisfaction judgments, which were more stable than pleasant and unpleasant affect. However, the 2-month interval was too short to reliably estimate the trait factor.

So far, the model mainly has been used for panel studies of life satisfaction where it could be fit to raw data (Lucas & Donnellan, 2007; Lucas & Donnellan, 2012; Schimmack & Lucas, 2010; Schimmack, Krause, Wagner, & Schupp, 2010). The main finding is that somewhere between 30% and 50% of the between-person variation in life satisfaction is attributable to a stable trait factor. Donnellan, Kenny, Trzesniewski, Lucas, and Conger (2012) and Kuster and Orth (2013) obtained similar results for self-esteem. Based on Conley’s finding that personality traits are more stable than self-evaluations (see also Anusic et al., 2012), it is likely that a stable factor accounts for even more variation in personality traits that exists at the between-person level. However, this pre-
diction has not been tested because there are no equivalent panel studies of personality.

In sum, the existing evidence suggests that personality traits are more stable than self-concepts and self-evaluative judgments such as judgments of self-esteem and life satisfaction. However, the evidence regarding the relative contribution of stable and changing influences on these constructs is inconclusive because previous studies have been limited by their design. The main limitation is that original studies often do not include all measures, sufficient measurement points, or a sufficient time period to observe significant change in individual differences.

In the present study, we were able to compare the stability and change of personality, well-being, and self-esteem measures by developing a Meta-Analytic Stability and Change model (MASC) for retest correlations. MASC can be regarded as an extension of Conley’s meta-analytic model and as a meta-analytic version of Kenny and Zautra’s (1995) TSE model for original data. The main contribution of our article is to introduce MASC as a model for meta-analyses of retest correlations and to apply the model to retest correlations of personality traits (Roberts & DelVecchio, 2000), life satisfaction ratings (Schimmack & Oishi, 2005), self-esteem ratings (Trzesniewski, Donnellan, & Robins, 2003), and ratings of affect. This model can be simultaneously applied to various individual differences and provide quantitative tests of differences in stability and change among them.

### Statistical Model

MASC can be considered a meta-analytic version of Kenny and Zautra’s (1995) TSE model for raw data. TSE is a structural equation model, and structural equation models are most meaningful when they are interpreted as realistic models of a set of causal processes that could have produced an observed pattern of data (Borsboom, Mellenbergh, & Van Heerden, 2003). Our model of stability and change is illustrated in Figure 1, where the boxes represent observed variances in a measure of individual differences (e.g., IQ tests, trait ratings, life satisfaction ratings). There are two boxes because the measure was administered twice. The model parameters predict the observed retest correlation between the two observed measures.

The model in Figure 1 distinguishes two sources of variance in responses at the between-person level: reliable variance and random error variance. The assumption that random measurement error influences personality measures is not controversial. It is common practice to estimate the amount of reliable variance in a personality measure and to report this information. However, random measurement error is often ignored in the interpretation of research findings (Schmidt & Hunter, 1996). For the meta-analytic integration of retest correlations, it is essential to distinguish between random measurement error and true change (Conley, 1984; Charles, 2005). The reason is that differences in reliability can distort comparisons of the true stability and change of the construct that is being measured. For example, a single life satisfaction rating is less reliable than a multiple item life satisfaction scale or a 20-item personality scale (Schimmack & Oishi, 2005). Without correction for unreliability, it is impossible to compare the stability of personality traits and life satisfaction. Thus, MASC corrects for unreliability and examines the stability and change of the reliable factor in Figure 1.

There has been some debate about the ability of the TSE to distinguish between random measurement error and true change (Lucas & Donnellan, 2012; Schimmack et al., 2010). The reason is that the estimate of the error parameter depends on the length of the shortest retest interval. To illustrate, in the German Socio-Economic Panel (e.g., see Lucas & Donnellan, 2012), life satisfaction is measured in annual intervals. A 1-year retest interval implies that some events can produce changes in life satisfaction that are only observed on a single occasion. For example, Petra may rate her life satisfaction a 7 in February 2001. She loses her job in October 2001. Her life satisfaction rating in February 2002 is a 4. However, in December 2002 she gets a new job, and in February 2003 she rates her life satisfaction a 7. The TSE model cannot distinguish the temporary shift in the life satisfaction rating in 2002 from random measurement error (e.g., Petra was in a bad mood when she rated her life satisfaction in 2002), although an actual life event produced a real change in life satisfaction that lasted from October 2001 to December 2002, a period of >1 year.

This is not a problem for MASC because a meta-analysis includes studies with different retest intervals. For example, retest intervals for life satisfaction judgments range from a few minutes when the same item was administered twice in a single survey to more than 20 years (Schimmack & Oishi, 2005). By including retest correlations over shorter time periods, the model does a better job of distinguishing random measurement error from true change. Random measurement error in this case also includes transient influences, such as mood effects and other factors that may influence response styles but do not necessarily reflect changes in one’s true standing on a psychological characteristic (Chmielewski & Watson, 2009).

Figure 1 shows that two factors contribute to reliable variation across individuals at any moment in time. One factor represents stable causes. Kenny and Zautra (1995) called this factor trait (i.e., the T in the TSE). We think that this terminology can create confusion because the term trait is also used to refer to personality constructs like extraversion. We use the term trait to refer to the Big Five and similar constructs, and the term stable factor to refer to stable causes of individual differences. For example, a study

---

**Figure 1.** The Meta-Analytic Stability and Change model (MASC), shown fitted to a variable measured at two occasions.
might report retest correlations of weight in an adult sample. One stable factor that produces individual differences in weight is height. Thus, height would be one stable cause of individual differences in weight. If height would explain 25% of the between-person variance in weight, the path coefficient from height to weight on any single occasion would be .5 ($S^2 = .25$ explained variance). The 25% shared variance between height and weight at two occasions would produce a retest correlation of $r = .25$ in weight measures. The retest correlation could no longer asymptote to zero because the stable factor height will produce a minimum correlation of $r = .25$. The square of the coefficient $S$ in Figure 1 can be called the stability coefficient ($S^2$), and it reflects the extent to which individual differences are caused by stable factors, just like the heritability coefficient in twin studies refers to the amount of variation across individuals that can be attributed to the effect of genetic variation.

In MASC, the nature of stable factors that contribute to retest correlations is unknown. MASC can only estimate the proportion of reliable variance on a single occasion that can be attributed to influences of stable factors. This includes direct stable influences (e.g., genes, stable aspects of the environment) but also influences of events that remain even after the event itself ends. For example, an adult person is no longer in her childhood environment, but her childhood environment may have long lasting influences on her thoughts, feelings, and behaviors, and thus would be considered a stable cause of adult personality (i.e., one’s childhood environment cannot change in adulthood).

The other factor that contributes to retest correlation is the change factor. Kenny and Zautra (1995) referred to this factor as state. We think it is confusing to refer to this factor as state in a model of stability and change of personality traits because state typically refers to occasion-specific states (e.g., mood). For this reason, we use the term change factor. As with the stable factor, the change factor reflects combined effects of several influences. These influences are grouped together because they share the property that they can change over time. As such, changing influences and stable influences are mutually exclusive. To use the example of weight again, changing factors that can contribute to weight are diet and exercise. When individuals change their diet or exercise habits, their weight changes, and these changes will influence the rank-order of individuals. For example, Michael may weigh 200 pounds and lose 50 pounds; Natasha may weigh 160 pounds and gain 10 pounds. As a result of these changes, Michael and Natasha switch places in the rank-order. These changes will lower retest correlations. Over longer periods of time, more changes can accumulate resulting in smaller retest correlations over longer retest intervals than over shorter retest intervals.

Figure 1 illustrates that the stable factor and the changing factor are two mutually exclusive factors that determine the reliable component of measured individual differences. This implies that the total reliable between-person variance is the sum of the amount of between-person variance explained by the stable factor ($S^2$) and the amount of between-person variance explained by the changing factor ($C^2$). Because retest correlations are based on standardized variables with variances of 1, it follows that

$$S^2 + C^2 = 1 \text{ and that } S^2 = 1 - C^2$$

Thus, although $S^2$ and $C^2$ are conceptually distinct factors, they are captured by a single parameter in a model with standardized coefficients. As a result, they reflect relative proportions of stable and changing variance. Original data would be needed to estimate the absolute amount of stable variance when the same measure is used. Once more, this is also true for behavioral genetic models, where environmental factors account for all of the variance not explained by genetic factors.

The final parameter in MASC models the rate of change of changing factors. Figure 1 shows that changing factors can change during the time interval between two retests. It is important that the amount of change that occurs during a retest interval depends on the length of the retest interval. Over a very brief interval, there is very little time for any changing factor to change. Over a very long interval, many changing factors will have changed and after infinitely long interval all changing factors will have changed. Factors that never change, and factors whose influence never changes even if the factor itself changes (e.g., a life event that continues to exert a constant influence on personality even after the event is over), are by definition stable factors and are captured in the stable factor.

In structural equation modeling, it is common to estimate the rate of change with the path coefficient from changing factors at Time 1 to changing factors at Time 2. This approach is useful in studies with equally spaced retest intervals (e.g., daily ratings). However, this approach is less useful for comparisons of retest intervals that vary in length because coefficients vary as a function of the retest interval. One solution to this problem is to convert coefficients for different time intervals into a common time interval. For example, Conley (1984) expressed the rate of change in terms of the retest correlation over a 1-year period. We decided to follow the same procedure and estimate a 1-year stability of the changing factors. In Figure 1, the quantity of change and stability of changing factors is quantified by the amount of variance in the changing factor at Time 2 that has not yet changed versus the amount of variance in changing factors that has changed.

To use a concrete example, we plotted a hypothetical pattern of retest correlations in Figure 2. The intercept in this figure is .80. This means that without any real changes, retest correlations are only $r = .80$. The remaining 20% of the variance is measurement error. Thus, the reliability of the hypothetical test scores in this study is 80%. Figure 2 also shows that retest correlations decrease over time, but then hit an asymptote at $r = .50$. This finding shows that 50% of the variance in test scores never changes. Thus, 50% of the total variance is determined by stable factors. Adjusted for unreliability, this implies that 62.5% ($0.50/0.80 = 0.625$) of the reliable variance is caused by stable factors ($S^2 = 0.625$). Changing factors account for the remaining 30% of the total variance in test scores (80% reliable variance – 50% stable variance = 30% changing variance). Changes in these factors lead to decreasing correlations over longer retest intervals. For example, as the retest interval increases from zero to 1 year, retest correlations decrease from .80 to .65. This drop implies that the stability coefficient in Figure 1 is .50. The reason is that changing factors contribute 30% of the variance, and stability of .50 implies that the total effect of changing factors is .15 over a 1-year retest interval because .30 * .50 = .15. Adding to this the effect of stable factors, $r = .50$, produces the 1-year retest correlation of $r = .65$ that is observed for a retest interval of one year. A stability coefficient of .50 over a 1-year interval implies that changing factors at Time 1 predict only 25% of the variation in changing factors at Time 2. As Figure 2 shows, a stability coefficient of .50 leads to a rather rapid decay
in retest correlations and predictability over time. After 5 years, 99% of between-person variance in the changing factors has changed, whereas only 1% is left unchanged.

MASC uses nonlinear regression analyses to estimate three regression coefficients that maximize predictions of retest correlations ($r_{t1,t2}$) from the length of the retest interval ($\text{TIME}$). The nonlinear equation is

$$ r_{t1,t2} = \text{REL}^2 \times \left(1 - C^2\right) + \left(C^2 \times \text{StCh}^{\text{TIME}}\right) $$

(2)

where \text{REL}^2 is the amount of reliable variance that exists at the between-person level, $C^2$ is the proportion of this reliable variance explained by the change factor, $(1 - C^2)$ is the proportion of the reliable variance that is explained by the stable factor, and \text{StCh} is the stability of the changing factors over 1 unit of \text{TIME} (e.g., 1 year).

### Current Study

Equation 2 is the basic form of MASC. It can be extended as necessary to evaluate the impact of moderator variables on different components of the model. Moderator variables are specified as product terms of the basic parameters and moderator variables. We include these product terms to examine whether model parameters differ significantly between studies of personality traits, self-esteem, life satisfaction, and affect. Based on existing literature, we expect the largest influence of stable factors on personality traits (Conley, 1984; Roberts & DelVecchio, 2000). Life satisfaction and self-esteem should show less stability over longer periods of time, yet they should show evidence of notable influence of stable factors as well (Anusic et al., 2012; Donnellan et al., 2012; Kuster & Orth, 2013; Lucas & Donnellan, 2012; Schimmack & Lucas, 2010).

It is more difficult to make predictions about where affect would fit on this continuum. Existing longitudinal studies of affect are less prevalent than those that focus on other psychological constructs. Some studies have suggested that affect is influenced by stable personality traits to a greater degree than life satisfaction (e.g., Schimmack, Radhakrishnan, Oishi, Dzokoto, & Ahadi, 2002), suggesting that affect would show greater influences of stable factors. On the other hand, there is some longitudinal evidence that, at least in the short-term, ratings of affect are less stable than those of life satisfaction (Anusic et al., 2012). One possible explanation for these inconsistent findings is that affect is more strongly influenced by stable factors than life satisfaction judgments, but that changing factors in affect decay more quickly. This would lead to lower stability for affect measures in the short-term and higher stability of affect measures in the long-term. Our study provides the first test of this hypothesis.

We also evaluate additional moderators of stability and change. The most prominent moderator in the literature is age. There is widespread consensus that stability of personality traits increases from childhood into adulthood (Roberts & DelVecchio, 2000; Specht et al., 2011; Wortman, Lucas, & Donnellan, 2012). Recent evidence suggests that stability decreases again in old age. A study by Lucas and Donnellan (2007) found that stability in life satisfaction also increases throughout the life span, with greatest stability occurring in the sixth decade of life, and a similar trend has been found for self-esteem (Donnellan et al., 2012; Kuster & Orth, 2013; Trzesniewski et al., 2003). However, the processes that lead to higher stability over the life span are still unclear. In this study, we evaluate for the first time to our knowledge the extent to which increasing stability is attributable to increasing reliability, increasing influence of stable factors, or increasing stability of changing factors. For example, if personality traits become increasingly more influenced by stable factors, and personality traits influence well-being (Diener & Lucas, 2000), we would expect that well-being also becomes more stable with age because of the greater influence of stable factors. However, it is also possible that well-being becomes more stable if changing factors such as income and social relationships become more stable with age.

### Method

#### Data Selection

We relied on Trzesniewski et al.’s (2003) meta-analysis of self-esteem to obtain studies that reported retest correlations for self-esteem. We used Roberts and DelVecchio’s (2000) meta-analysis to obtain a list of studies containing retest correlations for personality traits. We used Schimmack and Oishi’s (2005) meta-analysis to obtain studies containing retest correlations for life satisfaction. These sources provided the bulk of our data. We conducted additional searches using PsycInfo and WebOfScience databases with keywords for individual differences (i.e., personality, affect, feelings, mood, life satisfaction, happiness) paired with each of the following keywords: consistency, stability, longitudinal, trait, change. We defined each of the four types of individual differences (personality traits, life satisfaction, affect, and self-esteem) broadly and retained any studies that explicitly referenced one of these four psycholog-
We computed additional retest correlations from several publicly available panel studies. We obtained life satisfaction data from the British Household Panel Study, the Swiss Household Panel, the German Socioeconomic Panel, the National Survey of Families and Households, the Americans’ Changing Lives study, and the National Health and Nutrition Examination Survey. The German Socioeconomic Panel also provided retest correlations for personality for years 2005–2009. Additional self-esteem data were obtained from the Americans’ Changing Lives study, the National Educational Longitudinal Study, and the National Longitudinal Study. For these datasets, we obtained separate retest correlations over all possible retest intervals separately for both genders and for the following age groups (age was coded at the first wave of data collection): under 15 years, 15 to 30 years, 30 to 60 years, and over 60 years.

For each retest correlation reported in the articles or computed from panel studies, we recorded the following information: construct (personality traits, life satisfaction, affect, or self-esteem), age, gender, scale length, retest interval, and retest correlation. For age, we coded the mean age at the start of the retest interval as reported in the article. If age range was reported instead of the mean, we took the mean of the two endpoints of the range. Some articles also reported information on sample sizes of specific age groups (e.g., 10 participants aged 12–15 years, 48 participants aged 16–20 years)—in these cases we estimated age as the weighted mean. We centered the age variable around the overall mean (32.4 years), and transformed it by dividing it by 10. Thus, the moderating effects of age can be interpreted as change over a 10-year period. We coded gender as the proportion of female participants in the sample. Scale length was coded as the number of items in the scale for which the retest correlation was reported. If this information was not explicitly given in the article, we tried to find the information about the scale elsewhere (e.g., other articles that used the same scale). Because the association between scale length and reliability is not linear (e.g., the reliability difference between a single-item and a five-item scale is larger than the difference between 25- and 30-item measures), for our analyses we transformed the scale length variable by taking the natural logarithm of the actual number of items. We recorded retest interval in years. Finally, raw retest correlations were recorded as reported in the article or computed from the panel data. We included only retest correlations of self-reported measures, and omitted any retest correlations based on informant reports, observer ratings, or projective tests.

Some articles did not provide information about age, gender distribution, or scale length. In such cases, we imputed values for those variables using the mean for that construct. For example, if age information was missing in a study that reported a retest correlation for life satisfaction, we used the average age across all life satisfaction studies (40.83 years) as the age for that study. In total, 13 cases were missing age information (2 from personality, 4 from affect, and 7 from life satisfaction subsamples), 116 were missing information on gender distribution in their sample (30 from personality, 32 from affect, 22 from self-esteem, and 32 for life satisfaction subsamples), and 15 cases were missing information about scale length (5 from personality, 5 from affect, 1 from self-esteem, and 4 from life satisfaction subsamples).

We next limited our overall sample to include only retest correlations taken over retest intervals of 15 years or less. The reason for doing so was that retest correlations over longer retest intervals simply were not available for certain constructs (i.e., affect and self-esteem), which complicates comparisons of different constructs. As an important goal of our study was to compare stability and change of different individual difference constructs, we included only the range of retest intervals that was common across the constructs.

Next, we aggregated all information across cases that were obtained from the same sample over the same retest interval length. For example, a three-wave study in which the data were collected annually could provide two retest correlations over a 1-year retest interval (Time 1 to Time 2, and Time 2 to Time 3). In this case, we averaged all information (age, gender composition, retest correlation, etc.) across these two retest periods in order to obtain only a single retest correlation for each retest interval length in each sample. If an article reported retest correlations for multiple measures or multiple domains (e.g., different personality subscales or domains), we averaged these in the same manner. The data used in our meta-analysis are available online as supplemental material.

**MASC**

We used the *nls* function of the *stats* package of the R Statistical Software for our analyses ([R Development Core Team, 2010]). We fit three models to the data. The first model estimated overall reliability, relative effects of the stable and the changing factors, and the stability of the changing factors from all available retest correlations. The second model included moderating effects of construct. This model was used to test whether there were differences in the model parameters across personality, life satisfaction, self-esteem, and affect. This was estimated by adding product terms between the three dummy variables (that distinguished between the four individual differences of interest) and the model parameters. The third
model included all moderating effects. This allowed us to test whether, in addition to differences between construct, there was any evidence of the effects of scale length, gender, and age (and age squared) on model parameters. The complete script for our analyses can be found in the Online Supplement.

Results

Data Characteristics

Our final dataset contained 984 retest correlations: 243 for personality (179 studies), 106 for affect (59 studies), 206 for self-esteem (100 studies), and 429 for life satisfaction (104 studies; see the Online Supplement for details). The modal number of retest correlations per study was 1. Multiwave studies provided multiple retest correlations. For example, a study with three measurement occasions provides three retest correlations over the interval from T1 to T2, T2 to T3, and T1 to T3. The relatively high number of retest correlations for life satisfaction stems is because of the inclusion of a few panel studies with many retest correlations. Table 1 provides descriptive statistics about the samples. As can be seen Table 1, there was substantial variability in scale length, age, and retest interval among the studies.

Overall Model

We first fit a model without moderators. In this model, retest interval explained 28.8% of the variance in retest correlations. The next model added three dummy variables that distinguished between retest correlations of personality traits and self-esteem (Dummy 1), life satisfaction (Dummy 2), and affect (Dummy 3). Personality traits were used as the reference group. The amount of explained variance increased substantially to 48.3%, $\Delta R^2 = .195$, $F(9, 972) = 40.8$, $p < .0001$. We then added gender, age, age$^2$, and scale length as moderators. Scale length was only included as a moderator of the reliability parameter. Inclusion of these moderator variables further increased the amount of explained variance by 10% to 58.3%, $\Delta R^2 = .100$, $F(10, 962) = 23.0$, $p < .0001$.

Table 2 shows the estimates of reliability, relative influences of stable and changing factors, and 1-year and 10-year stabilities of the changing factors from the final model. Full model results that include effects of moderators and 95% CIs are available in the Online Supplement. The results are also depicted in Figure 3, which shows the observed retest correlations and the model estimates for personality traits, life satisfaction, self-esteem, and affect for different age groups.

The reliability estimate for personality traits was .72, 95% CI = (.65, .78). The results suggested that affect measures were less reliable than measures of other constructs (difference between measures of personality traits and affect was $-.14$, 95% CI = $(-.21, -.06)$). One plausible reason for this finding is that affect measures sometimes have shorter time frames (days, weeks). Thus, affect ratings are more likely to change so quickly that it becomes impossible to distinguish real changes from random measurement error. Unfortunately, we were not able to test this hypothesis because many studies failed to provide adequate information about the time frame of affect measures. However, for the purpose of comparisons with other constructs, the distinction between short-term fluctuations in mood and random measurement error is not important.

The second and third columns of Table 2 show the contribution of stable and changing factors to individual differences across different constructs. The results show that stable factors accounted for 83% of individual differences in personality traits; the remaining 17% of the reliable variance was explained by changing factors. As predicted, changing factors explained substantially more variance in affect (difference of 41%), life satisfaction (difference of 31%), and self-esteem (difference of 27%) than personality traits. The results for life satisfaction and self-esteem are very similar to each other, whereas affect appears to be even more strongly influenced by changing factors. However, the CI around this estimate is large and overlaps substantially with those for life satisfaction and self-esteem (see Table 1 in the online supplemental material). Before more data become available, our main conclusion is that personality traits are more stable than well-being and self-esteem.

The fourth column shows the estimates of the 1-year stability of the changing factors for personality traits, life satisfaction, self-esteem, and affect. We found significant but small stability of changing factors that influence personality (.25). With this small amount of stability, retest correlations reach the asymptote set by stable factors very quickly (Figure 3). Indeed, the 1-year stability of .25 implies that after 2 years only 0.004% of variance in changing factors remains ((.25$^2$)$^2$). We found strong evidence that the changing factors that influence well-being and self-esteem are considerably more stable. Our estimate of the 1-year stability of the changing factors for life satisfaction (.78) is consistent with previous estimates of about .80. Parameter estimates for affect and self-esteem are similar and CIs overlap. However, it is important to note that even 1-year stabilities as high as .80 would lead to large changes over longer periods of time because small changes accumulate over time. To illustrate

<table>
<thead>
<tr>
<th></th>
<th>Personality</th>
<th>Affect</th>
<th>Self-esteem</th>
<th>Life satisfaction</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>29.1 (16.5)</td>
<td>33.2 (18.6)</td>
<td>18.5 (12.5)</td>
<td>40.8 (21.4)</td>
<td>32.4 (20.2)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>46.5</td>
<td>54.9</td>
<td>48.7</td>
<td>50.7</td>
<td>49.7</td>
</tr>
<tr>
<td>Scale length</td>
<td>22.7 (17.0)</td>
<td>9.5 (5.8)</td>
<td>12.4 (16.4)</td>
<td>2.0 (2.9)</td>
<td>10.1 (14.3)</td>
</tr>
<tr>
<td>Retest interval</td>
<td>4.5 (3.6)</td>
<td>3.0 (3.2)</td>
<td>3.2 (3.1)</td>
<td>6.1 (4.1)</td>
<td>4.8 (3.9)</td>
</tr>
<tr>
<td>Retest correlations</td>
<td>243</td>
<td>106</td>
<td>206</td>
<td>429</td>
<td>984</td>
</tr>
</tbody>
</table>
this we also show 10-year stability estimates in the fifth column of Table 2. As this column shows, virtually no variance in changing factors of life satisfaction and self-esteem remains after 10-years (e.g., $0.09^2 = 0.01$).

In sum, the most important result was that the between-person variance in personality traits is notably more stable than variance in well-being and self-esteem, reflecting greater influences of stable factors on personality traits. Moreover, the changing factors that influence personality traits tend to change more rapidly than those for well-being and self-esteem. Because our analyses controlled for potential differences in reliability, this finding cannot be attributed to measurement artifacts.

### Table 2

<table>
<thead>
<tr>
<th>Construct</th>
<th>Reliability (1 – Error)</th>
<th>Stability (1 – Change)</th>
<th>1-year stability of the Change component</th>
<th>10-year stability of the Change component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality</td>
<td>0.72</td>
<td>0.83</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>Affect</td>
<td>0.58</td>
<td>0.42</td>
<td>0.58</td>
<td>0.88</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>0.71</td>
<td>0.56</td>
<td>0.44</td>
<td>0.79</td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>0.67</td>
<td>0.52</td>
<td>0.48</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Note. See online supplemental material for complete results with 95% confidence intervals and estimates of moderating effects of scale length, gender, and age.

![Figure 3](image-url)

Figure 3. Observed retest correlations as a function of retest interval. x = under 20 years of age, dot = between 20 and 40 years of age, triangle = older than 60. Regression lines illustrate model predictions for average age (solid gray line), 15-year-olds (dotted line), 30-year-olds (solid black line), and 60-year-olds (dashed lines).
Additional Moderators

As predicted by psychometric theory, scale length was a significant moderator of reliability, estimate = .04, 95% CI = (.03, .05). It is interesting that life satisfaction judgments were as reliable as personality measures after controlling for the effect of scale length. This finding undermines the claims that life satisfaction judgments are highly unreliable and susceptible to influence of mood and priming effects (Schwarz & Strack, 1999). The results are rather consistent with the view that life satisfaction judgments are based on chronically accessible information and provide reliable information about individuals’ well-being (Schimmack, Diener, & Oishi, 2002; Schimmack & Oishi, 2005).

Gender was not a significant moderator of any of the model parameters. Meta-analyses have relatively low power to detect gender effects because there is relatively little variation in the gender composition of samples. However, original data also produced similar results for women and men (Schimmack & Lucas, 2010). Thus, our results are likely to generalize across genders.

Age had statistically significant linear and quadratic effects on the effect of stable versus changing factors. The coefficient estimates can be interpreted as differences in parameters between people who are 10 years apart in age. The linear estimate for the moderating effect of age on the extent to which stable factors influence personality is .10, 95% CI = (.07, .02), and the quadratic estimate is −.02, 95% CI = (−.03, −.02). This suggests that the influence of stable factors on personality increases with age, but the increase is greater early on than later in life. For example, according to the model, stable factors contribute 83% of reliable between-person variance in personality traits for 32-year-olds (average age), 55% for 12-year-olds, and 95% for 52-year-olds. This finding is consistent with the findings from Roberts and DelVecchio’s (2000) meta-analysis that personality stability increases with age, particularly during adolescence and early adulthood (Figure 3).

Discussion

The stability of individual difference constructs has been one of the most debated issues in personality psychology. In this article, we used a novel meta-analytic method, MASC, to estimate the contribution of stable influences to retest correlations of personality traits, life satisfaction, affect, and self-esteem. Our findings support the hypothesis that stable factors have a stronger influence on personality traits than on the other individual differences. This difference cannot be attributed to differences in the reliability of personality and well-being measures because our model controls for differences in reliability.

We also were able for the first time to compare the stability of the changing factors across different individual differences. We found that the changing factors that influence personality traits are not very stable. In contrast, we replicated the finding of high stability for the changing factors that influence life satisfaction. We also found high stability for the changing factors that influence self-esteem, consistent with previous findings by Donnellan et al. (2012) and Kuster and Orth (2013). Our finding that stability differs across constructs has important implications for identifying predictors of stability and change in these constructs. For example, if personality traits were influenced by the same environmental factors as well-being, the changing factors of personality should be as stable as the changing factors of well-being.

There have been few direct comparisons of environmental influences on personality and well-being. Schimmack, Schupp, and Wagner (2008) examined the influence of unemployment in a national representative sample of Germans. Consistent with longitudinal studies, unemployment predicted lower life satisfaction and a lower hedonic balance (lower positive affect, higher negative affect), but unemployment did not predict personality traits. This finding suggests that unemployment is one of the changing factors that can produce changes in well-being that can last several years (Lucas, Clark, Georgellis, & Diener, 2004) without producing similar changes in personality traits.

Our results are also consistent with a recent study of 14,000 participants that examined predictors of changes in personality over a 4-year period. Despite the large sample, the study found few significant effects of major life events such as marriage, childbirth, and widowhood on personality traits (Specht et al., 2011). It is important that the same sample has been used to demonstrate that these environmental factors influence life satisfaction (Lucas, 2007).

Nevertheless, our findings suggest that changing factors have an influence on personality and refute the “plaster model” that says that personality traits are fixed after age 30. Our model merely suggests that these factors account for a relatively smaller portion of the variance compared with stable factors. Indeed, individual life events may have rather small effects on personality, but taken together changing circumstances can contribute to changes in personality. Consistent with this idea, several studies have linked individual environmental events to personality changes (Bleichorn, 2012; Jackson, Thoemmes, Jonkmann, Lüdtke, & Trautwein, 2012; Zimmernmann & Neyer, 2013).

Our study provides additional support for the hypothesis that stability increases with age. It is interesting that our results indicated that this increase is not because of increasing stability of the changing factors, which would suggest that influences of the changing factors that contribute to personality become longer lasting with age (e.g., people may hold on to their jobs for longer in older adulthood than in early adulthood). Rather, we found relatively stronger influences of stable factors with age. One explanation for this finding is that internal biological factors or the environmental factors (or both) become more stable with age. For example, higher environmental stability later in life (e.g., stable income and housing) may contribute to increasing contribution of the stable factor. That is, the influence of increased income following graduation may have permanent effects on life satisfaction and would be considered a new stable influence. People may also actively shape their environments to fit their existing predispositions, leading to increased stability of personality over the life span (Fraley & Roberts, 2005). However, results based on retest correlations are difficult to interpret because standardized coefficients only provide information about the relative proportions of variance attributed to each model component. Thus, an alternative explanation is that stable factors remain constant throughout the life span, but there is less influence of changing factors later in life. For

\[ .83 + (-2) \times .10 + (-2)^2 \times (-.02) = .55; .83 + (2) \times .10 + (2)^2 \times (-.02) = .95. \]
this reason, it is important to examine stability and change of personality and well-being in original studies using structural equation modeling of covariances (Kenny & Zautra, 1995; Schimmack & Lucas, 2010). Moreover, more longitudinal twin studies are needed to examine the contributions of genes and environment to stability and change. Existing evidence suggests that stability increases with age because environmental factors increasingly contribute to stability of personality in adulthood (Bleidorn et al., 2014; Briley & Tucker-Drob, 2014).

Our results also provide new evidence regarding the influence of personality on well-being. Schimmack et al. (2002) proposed that personality traits have a direct influence on affect because the Big Five traits like extraversion and neuroticism influence individuals’ typical mood levels. In contrast, the influence of personality traits on life satisfaction is indirect; personality only influences life satisfaction because people rely on their typical mood levels to judge their lives. That is, independent of actual life events, people who are typically in a good mood will judge their lives more positively. This model predicts that personality traits are stronger predictors of hedonic balance (positive—negative affect) than of life satisfaction (Schimmack et al., 2008). If personality is more stable, the model also predicts that affect should be more stable than life satisfaction. Our results provided no evidence of greater stability of affect over life satisfaction. However, the wide CIs show that the existing data are inconclusive. In the future, it will be important to test this hypothesis in larger samples, using original panel data that measured personality, life satisfaction, and affect.

Another limitation of retest correlations is that correlation coefficients do not take mean level changes into account. Thus, our results do not contradict findings that average levels of personality change with age (e.g., Roberts et al., 2008). Mean level changes have been demonstrated and may explain some of the changing variance in our model. The reason is that individuals are unlikely to change at the very same moment to exactly the same degree. For example, the average level of conscientiousness increases considerably between ages 20 and 30. However, if Janice’s conscientiousness increases considerably from age 22 to 24 and Jason’s conscientiousness increases considerably from age 25 to age 28, our model would show some short-term changes in the rank-order of conscientiousness. To examine this issue further, it would be necessary to model mean-level and rank-order stability and change with original data.

The finding that the stability of personality traits is higher than the stability of other psychological constructs can have interesting implications for clinical applications. Specifically, personality traits in adulthood appear to be highly stable. Indeed, clinicians often report difficulties in treating conditions such as personality disorders. However, life satisfaction and affect are less stable. Thus, although it may be difficult to increase quality of life by changing one’s personality, it would be possible to design interventions to increase life satisfaction, affective experiences, and self-esteem. This being said, it is also important to recognize that the stable factors in our model are only stable given the naturally occurring circumstances in which the studies were conducted. Future research needs to uncover the actual stable factors, and a better understanding of these factors may suggest novel ways to change them. Still, we believe our results can give clinicians some idea about the constraints on change and the efforts that may be required to induce lasting changes in psychological characteristics.

### Limitations and Future Directions

In this article, we demonstrated how a novel meta-analytic approach can be applied to study the stability and change of individual differences. The main advantage of this approach is that it treats both stability and change of constructs as theoretically important. Furthermore, this method is able to separate transient influences that lead to changes over short periods of time (Chmielewski & Watson, 2009) from longer lasting changes that are more likely related to important changes in life circumstances.

An important limitation of this approach, as with other trait-state approaches, is that the retest correlations must asymptote during the duration of the study in order for the model parameters to be estimated correctly (e.g., Anusic et al., 2012). If the asymptote is not reached over the study interval, the model may have difficulty separating the influences of the stable factors from those of the changing factors. Another limitation particular to the nonlinear regression approach to modeling stability and change is that the approach presented in this paper can be used only for descriptive purposes (e.g., to estimate relative proportions of source variance). It does not offer an explanation of the particular mechanisms responsible for stability and change of any particular construct. However, it may be possible to extend this model to include bivariate correlations (i.e., correlations of one variable at one time with another variable at another time). This may provide useful insight in the dynamics of covariation between variables over time.

Finally, it is important to note that some studies in this meta-analysis provided multiple retest correlations. Ideally, nonlinear multilevel models should be used in this case to account for nonindependence among the data. However, nonlinear multilevel models tend to be difficult to estimate when within-study data are scarce. In this case there were many studies that provided only one retest correlation, which resulted in difficulties with model convergence. A reasonable approximation is to use a simple nonlinear model. As a result, CIs should be interpreted with caution and are likely to be even wider than suggested in this article. Thus, it is even more important to conduct longitudinal studies that can provide more conclusive evidence about stability and change of important individual differences. At present, our meta-analysis provides the best scientific evidence on this important issue.

Despite these limitations, we believe that the meta-analytic approach presented in this paper is advantageous to other models in literature. Such an approach provides more informative estimates of the sources of stability. Moreover, this approach can be used to analyze archival data for which only retest correlations are available. Further research needs to consider conditions under which this model can and cannot be identified in order to further facilitate feasibility of this useful statistical method. Another important direction for future research is to use MASC to examine the stability and change of other personality characteristics. For example, Conley (1984) proposed a hierarchy of consistency with intelligence as the most stable construct. MASC makes it possible to compare parameter estimates for different constructs and to build a comprehensive hierarchy of stability of individual differences.

### References


META-ANALYSIS OF STABILITY AND CHANGE


META-ANALYSIS OF STABILITY AND CHANGE

781


Received October 10, 2014
Revision received June 18, 2015
Accepted June 23, 2015

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.