Journal of Personality and Social Psychology
2010, Vol. 98, No. 4, 659–667

0022-3514/10/$12.00 DOI: 10.1037/a0018770

A Meta-Analysis of Normal and Disordered Personality Across the Life Span

Christopher J. Ferguson
Texas A&M International University

Debate continues about whether personality, both normal and disordered, can change significantly or is mainly stable across the life span. One issue that receives little attention is the degree to which personality stability coefficients may be influenced by attenuation due to measurement error. The current meta-analysis examines the data from recent research on personality stability, reporting both uncorrected and corrected stability coefficients. Attenuation due to measurement error was found to cause a significant reduction in personality stability coefficients, raising the possibility that some studies may conflate personality change with measurement error. Overall, corrected stability coefficients suggested that the stability of personality across adulthood is high, with only modest change. By contrast, personality during childhood is significantly more changeable. Both normal personality and personality disorders were highly stable across the life span, and patients in therapy experienced no more personality change than did nonpatients. Cross-cultural comparisons suggested relatively similar levels of personality stability cross-culturally, although personality stability among people in South Pacific nations is slightly lower than among those in the United States, Canada, or European nations.

Keywords: personality traits, personality development, personality disorders, life span, personality change

Debate about the stability of both normal personality and personality disorders across the life span continues to be an issue of considerable interest. Some scholars have concluded that personality is largely stable and unchanging, particularly after approximately age thirty years (Costa & McCrae, 2006; McCrae & Costa, 1982), whereas others have emphasized that personality may continue to experience at least moderate change throughout the life span (Ardelt, 2000; Clark, 2009; Roberts, Walton, & Viechtbauer, 2006). One issue that has not been fully discussed is the degree to which stability estimates are conflated with measurement error. The current article seeks to build upon the work of others by considering personality stability across the life span, across personality traits (including general, nongeneral, and disordered traits), and across nations. The influence of correcting for attenuation due to measurement error on estimates of personality stability is also examined.

The issue of personality stability is one that is important beyond the scientific and public debates. Many decisions made in therapy as well as by persons in normal life may hinge upon this issue. For example, patients and therapists may set personality change as a goal of therapy. If personality is highly stable, this may not be a realistic goal. Similarly, many personal decisions are made in hopes that individuals—ranging from failed employees to abusive spouses—can "change." Elucidating the degree to which change is possible may be valuable for individuals making decisions based upon the hope for personality change.

Correspondence concerning this article should be addressed to Christopher J. Ferguson, Department of Behavioral, Applied Sciences and Criminal Justice, Texas A&M International University, 5201 University Boulevard, Laredo, TX 78041. E-mail: cjferguson1111@aol.com

Three Paradigms for Personality Stability

Historically, personality scholars have occupied one of three broad camps in relation to personality stability. The first of these is the *radical contextual perspective* (Caspi, Roberts, & Shiner, 2005). This perspective holds that personality traits are highly prone to change over time and that stability coefficients should be universally low. The radical contextual perspective appears to have fewer current adherents and is considered by some to be refuted (see Roberts et al., 2006).

The *biological essentialist perspective* (Caspi et al., 2005) asserts that personality traits are highly immutable and stable over time and are most likely the product of genetic rather than environmental influences (Costa & McCrae, 2006; McCrae & Costa, 1982). This perspective suggests that stability coefficients should be universally high, particularly during adulthood, although some modest personality change may occur throughout the life span.

The *compromise perspective* takes a middle ground between the other two. A moderate degree of personality stability is acknowledged. However, adherents of this perspective suggest also that some significant degree of personality change continues throughout the life span (Ardelt, 2000; Roberts & DelVecchio, 2000; Roberts & Mroczek, 2008).

Debates about personality stability occur not only in relation to normal personality traits, but also in regard to disorders of personality. Some scholars argue for the relative stability of such disorders (Lenzenweger, 1999), as does the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR*; American Psychiatric Association, 2000) itself, whereas others argue that personality disorders may be less stable than previously believed (Clark, 2009).

How Stable Is Stable?

Few scholars today would assert that the personality is either entirely changing ($r = .00$ or near to it) or entirely stable ($r = 1.00$). Rank-order stability refers essentially to test–retest results (data at Time 1 are correlated to data at Time 2). In one meta-analysis of rank-order stability, Roberts and DelVecchio (2000) found that uncorrected personality stability coefficients increased across the life span from approximately .35 in childhood to .75 in later adulthood. Ardelt (2000) found fairly similar results, with a mean stability coefficient of .55. Interestingly, Ardelt also found that the type of personality construct (general personality vs. nongeneral traits) had an influence on stability coefficients, with general personality traits (i.e., the Big Five traits) showing greater stability than do nongeneral traits (i.e., traits ostensibly not part of the Big Five). Nongeneral personality traits are typically personality traits that have not fit well into the Big Five or similar models and are usually more limited in scope. Examples may include, but are not limited to, aggressiveness, religiosity, and realism/naïveté.

One issue that has gone largely unmentioned in the literature on personality stability, at least of late, is that of measurement error and the attenuation of stability coefficients. Nunnally and Bernstein (1994) provided some nice discussion about the issue of measurement error, unreliability, attenuation, and correction. Measurement error, or the unreliability of responses to a measure, can attenuate resultant effect sizes. One example of a correction for attenuation formula is $r'_{xy} = r_{xy}/(r_{xx}\, r_{yy})^{1/2}$. Nunnally and Bernstein (1994) specifically recommended the use of correction formulas in personality trait research, stating:

> One such use is in personality research to estimate the correlation between two traits from imperfect indicators of these traits. Determining the correlation between traits is typically essential in this area of research, but if the relevant measures are only modestly reliable, the observed correlation will underestimate the correlations among the traits. (p. 257)

Although Nunnally and Bernstein speak here about two traits, certainly correlating the same trait at two time points would be conceptually similar. Personality measures vary widely in regard to reliability estimates for different samples. Reliability estimates in the .70 range, and sometimes significantly below, are not uncommon.

Despite Nunnally and Bernstein's (1994) recommendations, few recent articles on personality stability bring up the issue of attenuation due to unreliability or measurement error. None of the recent meta-analyses provide any corrected stability estimates. Ardelt (2000) did briefly note that stability estimates corrected for unreliability do tend to be higher, but she nonetheless used only uncorrected estimates in her meta-analysis. Although a few individual studies do provide corrected estimates (e.g., Johnson, McGue, & Krueger, 2005; Laptook, Klein, & Dougherty, 2006), most do not.

This issue is raised by Watson (2004), who noted that personality researchers have virtually ignored the issue of measurement error. Watson identified several sources of measurement error—including wording choice and response format—that can attenuate stability coefficients in long-term personality stability research. Similar issues were also discussed by Cattell (1964a, 1964b), who suggested using short-term dependability studies (essentially test–retest reliability) to establish the degree to which stability coefficients are due to real change or measurement error.

It is unclear which reliability estimates are best used for correction due to attenuation. Coefficient alpha has the advantage of being more conceptually different from the stability coefficients of interest than is test–retest reliability. In a footnote, Roberts and DelVecchio (2000, p. 15) appear to endorse the use of coefficient alpha and/or kappa reliabilities, particularly because these estimates are available for meta-analysis in more reports than are test–retest reliabilities. Nunnally and Bernstein (1994) noted that it is generally desirable for the coefficient alpha to be high, whereas the value of temporal stability may vary by construct. They also suggested that coefficient alpha as an estimate of measurement error is superior to other forms of reliability, stating: "Even though coefficient alpha ignores certain potentially important sources of measurement error, it is surprising how little difference these sources of measurement error usually make" (p. 252). By contrast, they advised against the use of retest reliability except under certain conditions (p. 255).

Other Issues in Personality Stability

There are other issues related to personality stability that warrant attention. Much discussion exists regarding the degree to which personality dimensions and stability are influenced by culture (McCrae, 2000). For example, the Big Five model has been found to be consistent across cultures (McCrae & Costa, 1997). Most longitudinal examinations of personality stability have taken place with U.S. samples, although studies outside the United States are increasing in number. Examining the stability as well as the structure of personality may help inform science about the universality or variation of personality across cultures. Not doing so risks applying an overly European or American model of personality and personality stability to cultures for which it may not apply.

Debates remain regarding the life course pattern of personality, with conflicting results on peak years for personality stability. McCrae and Costa (1982) found that personality stability peaked and remained stable at age 30, whereas Ardelt (2000) has suggested that adults around age 50 demonstrate declines in personality stability from a peak in young adulthood. Roberts and DelVecchio (2000) instead found that personality stability did not peak until age 50. Although all studies agree that personality is relatively stable throughout adulthood, more data are needed to examine the course of stability throughout adulthood.

Previous studies have suggested that gender differences in normal personality stability are minimal (Ardelt, 2000; Roberts & DelVecchio, 2000). However, this issue has not been examined as clearly with personality disorders, and more research would be welcome in this area.

It has been observed that some patients in therapy set personality change as one of their goals for therapeutic outcome (de Maat, de Jonghe, Schoevers, & Dekker, 2009). However, if personality is relatively stable across the life span, setting personality stability as a goal for treatment may be unrealistic. Indeed, current research indicates that treatment effectiveness for personality change is much lower than for treatment of individual symptoms (de Maat et al., 2009). Few studies have compared personality stability among patients in therapy with that among individuals in the general population. Such a comparison may elucidate whether

setting personality change as a therapeutic goal is a worthwhile endeavor.

The Present Study

The past decade has seen several meta-analyses on rank-order stability (e.g., Ardelt, 2000; Roberts & DelVecchio, 2000) and one more recent analysis of mean-level stability (Roberts et al., 2006). The present study aims to expand upon these previous works in the following ways:

1. None of the previous analyses provided corrected stability estimates. The current meta-analysis will provide both uncorrected and corrected stability coefficients in order to gauge the potential influence of measurement error on the discussion of personality stability.

2. The previous analyses considered normal personality only. The current meta-analysis will extend this further to consider long-term stability coefficients of not only normal personality but also disorders of personality.

3. Related to the previous point, the current analysis will contrast general and clinical samples, which has not previously been done.

4. Researchers have called for greater investigation of cross-cultural issues in personality (McCrae, 2000; Wong, Lee, Ang, Oei, & Ng, 2009). A fair number of personality stability studies have emerged from non-U.S. westernized countries including Canada, the South Pacific Basin (Australia and New Zealand), and Europe (particularly Sweden, Finland, and the Netherlands but also including Croatia, Norway, Germany, Spain, and Estonia). Although it would be ideal to have studies from non-Western nations, there are enough studies from the United States, Canada, Europe, and the South Pacific Basin to allow for comparisons across these regions, and this will be undertaken in the current article.

5. Previous rank-order analyses (Ardelt, 2000; Roberts & DelVecchio, 2000) have come to differing conclusions regarding the stability of personality, particularly in later life. The current study will replicate the life-span approach of these previous studies to provide more data to help resolve this issue.

6. The current study will examine gender differences in personality stability.

It is worth considering how large effects need to be to support either the biological essentialist or compromise perspective. How large a difference between time intervals/groups/ages is necessary to support the compromise position? How large do personality stability coefficients need to be to support the biological essentialist position? There are no clear criteria for answering these questions. Cohen's (1992) highly influential standards for interpreting effect sizes probably do not apply here. If they did, then personality stability coefficients, particularly in adulthood, would be universally large by his recommendations, but this is not the position of the compromise perspective. Although not exactly intended for this purpose, Ferguson's (2009) recommendations may provide some guidelines. Ferguson recommended an effect size of $r = .20$ as a minimum to indicate practical significance. Particularly if one uses the absolute value of the correlation to represent the coefficient of determination (so $r = .20$ represents 20% shared variance rather than 4%) as is recommended by Ozer (1985), then this standard can be applied to examine whether the differences between corrected and uncorrected coefficients are large enough to warrant concern. Similarly, Cohen's recommendation of $r = .5$ for large effects was meant to be relativistic to social science data, whereas Ferguson's similar recommendation of $r = .8$ for large effects is not relativistic. This $r = .8$ standard may be more applicable regarding judging the merits of the biological essentialist position. Even for corrected personality stability coefficients, $r = .80$ is a high bar to reach and, if reached, would provide strong support in favor of the biological essentialist position while still acknowledging the presence of some moderate personality change across the life span.

Method

Literature Search and Study Inclusion

The primary method of searching the literature was searching the PsycINFO data base. The following search terms were used: (*personality* OR *temperament*) AND (*stability* OR *consistency* OR *longitudinal* OR *change*). Also, the reference sections of several recent articles on the issues of personality stability (e.g., Roberts & Mroczek, 2008; Watson, 2004) and personality disorders (Clark, 2009) were examined for any studies that may have been missed, as were the reference sections of articles identified during the PsycINFO search. So as not to overlap with previous meta-analyses, and to examine more recent and potentially methodologically updated work, I included only work published in the last 10 years (1999–2008; studies included in Roberts & DelVecchio, 2000, end at approximately 1999). Only peer-reviewed published studies were included in the current analyses. Although it is not uncommon for meta-analyses to include unpublished studies, some scholars have indicated that attempting to do so may increase rather than decrease bias in the sample of included studies (Baumeister, DeWall, & Vohs, 2009). The quality of unpublished studies may be different from that of published studies. Because no repository of unpublished studies exists, attempting to recruit unpublished studies from known authors may ultimately result in a highly biased sample of studies, because unknown authors and studies are not sampled. Consistent with the concerns described by Baumeister et al. (2009), the current analysis includes only published studies, in hopes of maintaining a more homogeneous sample of studies in regard to quality. Included studies had to meet the following criteria:

1. Each study had to include some measure of personality trait, whether general, nongeneral, or disorder. Measures of other variables such as affect, self-esteem, and cognition were not included (Roberts & DelVecchio, 2000).

2. Studies that examined only short-term test–retest reliability were not included. A 3-month minimum longi-

tudinal period was required for study inclusion. This minimum longitudinal period was used to eliminate studies for which test–retest reliability rather than personality stability was the primary objective. Although any time cutoff is arbitrary, personality change in less than a 3-month period is unlikely.

3. Each study had to provide information on whether the sample was clinical or general, the age of the sample, and the study time interval, as well as provide some data of reliability or measurement error of the measures used.

4. For studies involving personality disorders, only studies examining the stability of personality disorders over time were included. Studies that were specifically treatment outcome studies for personality disorders were not included, because the goal of such studies is quite different.

Forty-seven studies that met these criteria were found.[1] The total number of participants was 30,990.

Correction for Attenuation Due to Measurement Error

As indicated in the previous section, both uncorrected and corrected stability coefficients were calculated and will be reported here. As noted earlier, debate continues about whether correction formulas should be used for stability coefficients and, if so, what form of measurement error estimate should be used in the correction. In the current analysis, coefficient alpha (internal consistency) was identified as the most desirable for several theoretical and practical reasons. First, as simply a practical issue, coefficient alpha was reported in the majority of studies. Relying on short-term test–retest data would have resulted in the disqualification of many studies. There were no studies that reported only short-term test–retest data. A few studies employing behavioral ratings or standardized diagnostic interviews reported kappa reliabilities rather than coefficient alpha, and kappa was used in these cases. These studies were mainly among those examining the consistency of personality disorders. It should be noted that the Standards for Educational and Psychological Testing (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) caution against the use of differing reliability estimates such as coefficient alpha and kappa interchangeably. However, because the current study sought to examine the relative stability of both normal (which was mainly reported with alpha) and disordered (which was mainly reported with kappa) personality, studies reporting both kinds of reliability were included with care. Results from a point-biserial correlation between the use of kappa or coefficient alpha and the size of the attenuation adjustment for stability coefficients revealed that there was no relationship ($r = .01$). Thus, in this situation, coefficient alpha and kappa seemed to perform similarly, although this cautionary note should be considered in interpreting the data. Nunnally and Bernstein (1994) recommended correcting for measurement error as an ideal; however, precise information on measurement error is not always available. Their discussions on coefficient alpha and attenuation (see pages 241 and 255) suggest that coefficient alpha functions well as an estimate of measurement error for this purpose of correcting for attenuation. As such, the decision to use coefficient alpha appears to be on reasonably firm ground. Ultimately when multiple estimates of measurement error were provided in a study, their magnitudes tended to be similar.

Study Variables

Rank-order consistency. As with two previous reviews a decade earlier (Ardelt, 2000; Roberts & DelVecchio, 2000), the current analysis examined rank-order stability. All reports found during the literature search reported rank-order stability coefficients. Some also reported other forms of stability estimates, but not all did. As such, rank-order appears to allow for the broadest possible sample of studies.

Age. The current analysis followed Roberts and DelVecchio's (2000) recommendations for evaluating the age variable. Age at Time 1 was coded. For studies that reported age ranges, the mean age was used as an estimate. As in Roberts and DelVecchio, age blocks of 6 years (e.g., 0–6, 7–13, 14–20) were used to evaluate stability across the life span.

Time interval. The time interval from Time 1 to the final longitudinal evaluation was coded in number of months. For studies in which multiple evaluations were conducted, the longest time interval was taken as the unit of analysis. This allowed for analysis of personality stability across maximal time periods.

Trait categories. Roberts and DelVecchio (2000) organized traits under a limited number of discrete categories, such as the Big Five categories as well as several others. A slightly different approach was taken here in order to examine personality traits from an alternate perspective. As in Ardelt (2000), studies were grouped into those that examined global or general traits such as the Big Five and similar global personality models (e.g., the Big Three, the 16 primary factors of Cattell, 1957) or those that considered nongeneral traits, which were much more specific and limited and did not clearly fit within the various Big X or other general models. The current analysis adds a third category—that of personality disorders. Personality disorders represent deviant personality traits that are pervasive and have historically been considered as stable if not more rigidly so than are normal personality traits (Clark, 2009; Lenzenweger, 1999). Arguably, deviant personality traits can be as pervasive and impactful as normal personality traits. To date, they have not been considered side-by-side in a meta-analytic review of longitudinal studies.

Clinical status. The clinical status of participants (patients or nonpatients) was recorded. Any sample recruited from an outpatient clinic, psychiatric hospital, or similar therapeutic environment was considered to be a patient sample. All other samples—whether university, general, community, or other—were considered nonpatients. It is important to note that not all patient samples were examining the stability of personality disorders, nor were all personality disorder stability studies conducted on patient samples, although there naturally was considerable overlap.

Nationality. The nationality of the sample used in the analyses was coded. No samples crossed nationalities (that is to say, no samples were drawn from two separate countries). The majority of

[1] A table of included studies with sample characteristics is available from Christopher J. Ferguson upon request.

samples were from the United States, although a fair number of samples came from Canada, the South Pacific Basin (Australia and New Zealand), and Europe, allowing for at least some general cross-cultural comparisons. One reasonable concern is the heterogeneity of cultures in Europe, which may make a European comparison somewhat tenuous (e.g., combining data from Spain and Estonia, at the geographic extremes). With that in mind, one comparison looked at Europe as a whole, and a second examined only a cluster of Germanic and Scandinavian countries (Norway, Sweden, Finland, Germany, and the Netherlands), which might be assumed to be somewhat more culturally homogeneous, to examine whether the degree of homogeneity in European countries is an issue.

Gender. The current study examined gender differences in personality stability. Unfortunately most studies collapse personality stability coefficients across gender. Thus, in order to gain a reasonable sample of studies for which gender differences could be assessed, the authors of all identified studies not supplying gender data were contacted with a request that data on stability coefficients broken down by gender be supplied for the purposes of this meta-analysis only. Some study authors could not be located; others did not reply or no longer had possession of the original data. Ultimately data from 27 studies were included.

Calculating Effect Size Estimates

Pearson's *r*, a flexible and easily interpreted index of effect size, was used as the effect size estimate in this study. Correlation coefficients were transformed to Fisher's *z*, weighted, averaged, and transformed back to a pooled *r*. The Comprehensive Meta-Analysis (Version 2) software program was used to fit both random and fixed effects models. Hunter and Schmidt (2004) also argued that random effects models are appropriate when population parameters may vary across studies, as is likely here. As such, only random effects models are presented.

Results

Overall Results

Of the 47 studies in the current analysis, two provided only corrected stability coefficients, five provided both corrected and uncorrected stability coefficients, and the remaining 40 provided only uncorrected stability coefficients. As such, we can see that the use of corrected stability coefficients remains relatively rare. Thus, the impact of measurement error on results presented in this field may be rather pervasive.

The average longitudinal time span of studies in the current analysis was 80 months (SD = 105 months), with a range of 3 months through 504 months (42 years). The descriptive breakdown on age and nationality is presented in Table 1. Twenty-four of the included studies examined general personality, such as the Big Five model. Ten studies examined nongeneral, limited personality traits. Thirteen studies examined the stability of personality disorders. Fourteen studies examined psychiatric patients, whereas 33 examined samples of individuals from the general population.

Table 1 presents the meta-analytic results for corrected and uncorrected personality stability estimates for all studies and across moderator variables. Results for all included studies reveal that the corrected personality stability coefficient across all studies is r = .79. The uncorrected coefficient is r = .60. There are two ways of interpreting the difference between the two. If we interpret *r* directly as the coefficient of determination, then this suggests that measurement error may be causing a 19% reduction in shared variance. Alternatively, if r^2 is used as the coefficient of determination, then the difference between the two coefficients of determination can be taken (62% – 36% = 26%). Thus, measurement error may be causing attenuation of between 19% and 26% in shared variance. These values match or exceed Ferguson's (2009) recommendations for minimum practical significance.

Age Trends Over the Life Span

In examining age trends over the life span, only analyses including general or nongeneral traits were considered. This was out of concern that analyses of personality disorders may bias the results of some age groups but not others, causing misleading results (there were few analyses of personality disorders for children, not surprisingly). One study was left out of this analysis because it did not provide specific information on mean age. Results are presented in Table 1 as well as in Figure 1.

Age-related trends in personality disorders were also examined. These data are presented in Table 1 as well as in Figure 2. As with the previous analysis, two studies were not included due to imprecise age information.

In both cases, data suggest that personality is relatively changing during youth, yet quickly becomes stable by early adulthood. Personality remains generally stable throughout adulthood, particularly after the late 20s, and remains stable throughout later adulthood.

Time Interval

In order to examine the influence of longitudinal time span, a hierarchical multiple regression was used. Stability coefficient effect size was the outcome variable. Participant age and type of personality measure (general, nongeneral, personality disorder) were entered in the first step of the regression in order to eliminate extraneous variance due to these variables. Time interval was entered in the second step. The overall regression model was significant (R = .59, adjusted R^2 = .30), $F(3, 43) = 7.45$, $p < .001$. Age was clearly the most influential variable in regard to effect size (β = .56). Personality variable type was of little practical significance (β = –.07). Time span (β = –.18) was a smaller predictor, although approaching Ferguson's (2009) recommendations for practical significance. Thus, longer time intervals do produce smaller stability coefficients, although the effect overall is small. This is consistent with Fraley and Roberts (2005), who found that although stability coefficients decline with longer time intervals, they do not reach zero but gradually form a nonzero asymptote. In other words, the pattern of stability coefficients stabilizes and levels at a certain time interval, although the time interval at which stability coefficients stabilize may differ from one construct to another.

Trait Categories

The previous analysis suggests that personality variable (general, nongeneral, or personality disorder) may have little influence

Table 1
Meta-Analytic Results for Main Analysis and Moderator Variables, Including Uncorrected and Corrected Effect Size Estimates

Variable	*k*	r_c	r_u	95% CI	Homogeneity test
All observations	47[a]	.79	.60	[.72, .85]	$\chi^2(46) = 9{,}802.30, p < .001$
Age in years					
General personality					
0–6	3	.56	.43	[.46, .65]	$\chi^2(2) = 0.90, p > .05$
7–13	3	.44	.35	[.10, .69]	$\chi^2(2) = 119.51, p < .001$
14–20	9	.72	.57	[.68, .77]	$\chi^2(8) = 81.32, p < .001$
21–27	3	.70	.56	[.57, .80]	$\chi^2(2) = 26.68, p < .001$
28–34	3	.94	.66	[.60, .99]	$\chi^2(2) = 550.98, p < .001$
35–41	7	.87	.72	[.65, .95]	$\chi^2(6) = 3{,}922.15, p < .001$
42–62[b]	3	.82	.73	[.71, .88]	$\chi^2(2) = 45.40, p < .001$
63–83[b]	2	.94	.73	[.87, .97]	$\chi^2(1) = 29.29, p < .001$
Personality disorder					
0–20[b]	5	.66	.42	[.40, .82]	$\chi^2(4) = 249.49, p < .001$
21–34[b]	5	.85	.59	[.39, .97]	$\chi^2(4) = 1{,}009.01, p < .001$
35–41[c]	1	.85	.68	[.79, .90]	
Trait category					
General	24	.76	.59	[.66, .84]	$\chi^2(23) = 3{,}742.22, p < .001$
Nongeneral	10	.86	.68	[.65, .95]	$\chi^2(9) = 3{,}349.39, p < .001$
Personality disorder	13	.79	.55	[.59, .90]	$\chi^2(12) = 2{,}002.10, p < .001$
Clinical status					
Patient	14	.81	.62	[.63, .91]	$\chi^2(13) = 1{,}307.48, p < .001$
Nonpatient	33	.78	.59	[.70, .85]	$\chi^2(32) = 7{,}968.34, p < .001$
Nationality					
United States	28	.79	.60	[.68, .86]	$\chi^2(27) = 4{,}958.41, p < .001$
Canada	4	.82	.69	[.54, .94]	$\chi^2(3) = 272.19, p < .001$
South Pacific	3	.71	.59	[.51, .83]	$\chi^2(2) = 74.66, p < .001$
Europe	12	.81	.58	[.64, .91]	$\chi^2(11) = 2{,}363.99, p < .001$
Scand/Germ	9	.85	.59	[.65, .94]	$\chi^2(8) = 2{,}115.64, p < .001$

Note. CIs and homogeneity tests reported are for the corrected estimates. *k* = number of independent studies; r_c = corrected personality stability estimate; r_u = uncorrected personality stability estimate; CI = confidence interval; Scand/Germ = Scandinavian and Germanic countries only.
[a] For uncorrected estimates, *k* = 45 because two studies did not provide uncorrected estimates. [b] Age ranges were collapsed due to fewer studies. [c] Only one study was available in this age range, so data should be interpreted with caution.

on effect size. These nonetheless were examined more closely through meta-analysis, and results are presented in Table 1. Results suggested little difference between types of personality measure. Personality disorders, general personality, and nongeneral traits were roughly equal in stability.

Clinical Status

The difference in stability coefficients between psychiatric patients and nonpatients was examined. Results are presented in Table 1. Results indicated little difference between patients and nonpatients in regard to personality stability.

Nationality

Differences between nationalities (United States, Canada, Europe, South Pacific Basin) were examined, with results presented in Table 1. Germanic and Scandinavian countries were also examined alone in a separate analysis out of concern that the Euro-

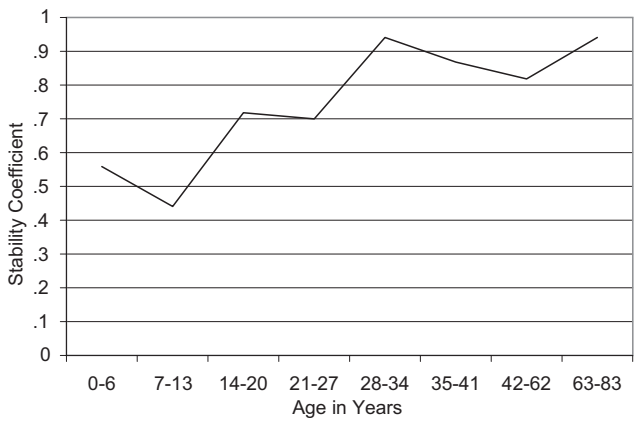

Figure 1. Age trends in normal personality stability.

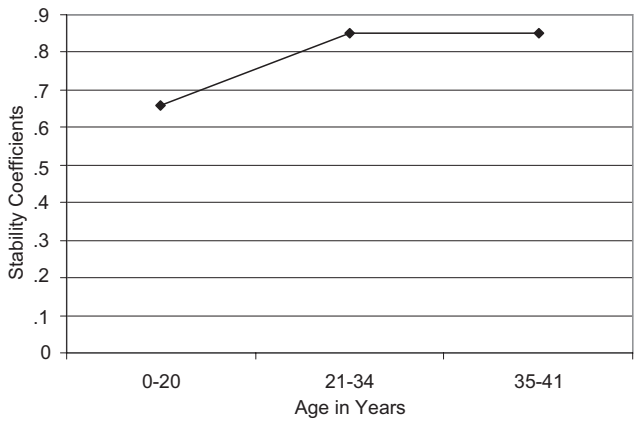

Figure 2. Age trends in personality disorders.

pean group may have been too culturally heterogeneous. Results indicated that cultural differences were small. However, the lowest stability coefficients were found for the South Pacific Basin, and the highest coefficients were found for Scandinavian and Germanic countries.

Gender Differences

Results from the current analysis revealed that female participants (r = .81) demonstrated slightly higher personality stability than did male participants (r = .74), although this difference (r = .07) was far below the threshold set for significance. These basic results held for general personality (male = .71, female = .74), nongeneral traits (male = .84, female = .91) and personality disorders (male = .72, female = .80). Ultimately, these results suggest that gender differences in personality stability are small.

Discussion

Several important findings from the current study merit discussion. First, attenuation due to unreliability was found to reduce estimates of personality stability by approximately 19%–26%—a significant amount. The current study used coefficient alpha and kappa as estimates of reliability, due to their widespread reporting in articles. Had more precise estimates of measurement error been reported and utilized, it is likely that the corrected stability coefficients would have been even higher. Given the degree of debate regarding the stability of personality over time (e.g., Costa & McCrae, 2006; Roberts et al., 2006), this is not a trivial issue. Failing to correct for attenuation due to measurement error in instruments used may result in personality stability estimates that are biased and misleading. Subsequently, the scientific community may be unintentionally misinformed about the degree to which personality can change across the life span.

Examining stability trends over the life span revealed that personality is moderately stable during childhood, increases in stability in the teen years, peaks in stability around age 30, and remains stable from that point forward. These results were found both for normal personality traits and for disordered personality traits. These results are consistent with the view of McCrae and Costa (1982), who have suggested that personality stability peaks at approximately age 30 and remains steady thereafter. Previous meta-analyses of personality stability have generally agreed that personality stability is highest during adulthood, although with some mild variations. Ardelt (2000) suggested that stability may decline around age 50, perhaps due to social and family changes occurring about that time. By contrast, Roberts and DelVecchio (2000) found that stability coefficients peaked at age 50. The current results indicate a middle road, finding no evidence for either a decline or incline in stability coefficients at age 50. It may be that the differences discussed across all these meta-analyses are small enough that minor fluctuations may be statistical artifacts.

Type of personality construct (general, nongeneral, or personality disorder) had no influence on stability coefficients. All forms of personality traits remained equally stable across the life span. This may disappoint some scholars who have expressed optimism for change, particularly in the realm of personality disorders (e.g., Clark, 2009). However, the current results suggest that the prevailing view of personality disorders as essentially stable across the life course, as endorsed by the *DSM–IV–TR* (American Psychiatric Association, 2000), is correct. Relatedly, the current results found that patients in therapy were no more likely to see personality change than were nonpatients. These results suggest that therapeutic outcomes targeted toward changing fundamental personality structures in patients may be unrealistic. Treatment outcomes focused on specific behaviors rather than global personality change may see more positive outcomes.

A cross-cultural comparison of personality stability coefficients found that personality stability was generally similar in scale across most included cultural groups. Only individuals from the South Pacific demonstrated significantly smaller personality stability coefficients than did other cultures, although here, too, personality stability was generally high. Although cross-cultural comparisons of personality stability are an important issue, some limitations of cross-cultural comparisons should be noted. Several potential confounds could conceivably intrude upon cross-cultural comparisons, making interpretation of differences or similarities difficult. Measures of personality may display differential construct validity across cultures, the equivalence of measurement models cannot always be assumed, there may be differences in socially desirable responding, and errors in translation of instruments may also influence outcomes. Also, measurements of personality stability from developing nations—particularly those in Africa, Latin America, and Asia—remain in short supply.

Finally, the analyses seem to indicate that the personality stability of female participants is slightly higher than that of male participants across all personality constructs. However, this difference was small and below the threshold for significance established in this study. It seems that female and male individuals are more similar than different in regard to personality stability. Previous meta-analyses (Ardelt, 2000; Roberts & DelVecchio, 2000) have found similar results. As such, it is probably time to end speculation regarding gender differences in personality stability. The current analysis expands upon previous analyses by considering gender differences in personality disorder stability. Although the predominant view suggests that prevalence differences in personality disorders between the genders is well established (American Psychiatric Association, 2000), much controversy remains regarding the measurement and stability of personality disorders in male and female individuals (e.g., Grilo, 2002; Hynan, 2004). Current results suggest that, like the stability coefficients for general and nongeneral traits, personality disorder stability coefficients do not differ significantly by gender.

By and large, corrected stability coefficients reached the r = .80 threshold identified earlier in the article. Aside from those for children, no corrected stability coefficients were lower than .70. Similarly, psychiatric patients showed no greater likelihood of experiencing personality change than did nonpatients. Generally speaking, the results of the current study provide data in support of the biological essentialist perspective. Only the observation that time interval is related to decreased stability scores offered some support for the compromise perspective, although overall this effect was small (β = –.18). This result highlights that personality is not entirely stable over time. Nonetheless, the size of this effect is small and not inconsistent with Costa and McCrae's (2006) acknowledgment of modest changes in personality occurring throughout adulthood. As such, overall the data provided here support the general stability of personality, with modest change, across adulthood.

In one sense, the compromise perspective is certainly correct in stating that personality is neither entirely consistent nor entirely changing (and is probably the product of both biological and environmental forces). However, the biological essentialist perspective is better supported by the current data, because personality appears to be largely stable beginning in early adulthood, with only modest changes thereafter (Costa & McCrae, 2006). It is important to emphasize that the current data do not provide evidence either for or against the belief that personality is necessarily the product of genetics, only that whatever its origin, it remains largely stable in adulthood. This was true both for normal personality traits and for disorders of the personality. Other theoretical viewpoints may explain the consistency of personality in adulthood equally well or better than do genetic or biological essentialist positions. For example Bandura (2006) has suggested that individuals play an active role in shaping their environments and that these social environments in turn shape and influence their behavior and personalities. It may be that adults, having arguably greater control over their social environments, are more able to shape environments that are consistent with their existing personality, thereby essentially maintaining their personality more reliably over time. Examining the personality stability of individuals living in environments over which they hold minimal control (e.g., incarcerated inmates, monks or nuns living in religious communities) may shed some light on this issue.

In regard to the differences in results seen between Ardelt (2000) and Roberts and DelVecchio (2000) regarding stability trends in later adulthood, current results appear to support Roberts and DelVecchio (2000) more so than Ardelt (2000). It appears that personality in later years is equally stable to that seen in early adulthood.

The current study is the first meta-analysis to provide data examining cross-cultural trends in personality stability. Overall, results suggest that at least among Western cultures, personality stability is fairly consistent across nations. The South Pacific was a little lower in stability than were other groups but was still high overall. It is possible that more disparate groups (Asia, Africa, Latin America) might show different trends.

As with any study, this one is not without limitations. As noted earlier, although including studies reporting alpha as well as those reporting kappa did not seem to influence correction for attenuation, this situation is not ideal. In this study, this procedure was used to examine the relative stability of normal and disordered personality, although it would be useful to future studies to examine this issue with differing methodologies that allow for more precise direct comparisons. Also, data in the current meta-analysis remain focused on industrialized nations. Further long-term stability studies from other nations would be welcome.

In closing, the current analysis finds that the stability of personality over long periods, particularly during adulthood, is more stable than previously thought. It is likely that debates regarding personality stability will continue for some time, and it is hoped that this article will be a constructive addition to these dialogues.

References

American Education Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Ardelt, M. (2000). Still stable after all these years? Personality stability theory revisited. *Social Science Quarterly, 63,* 392–405.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1,* 164–180.

Baumeister, R., DeWall, N., & Vohs, K. (2009). Social rejection, control, numbness, and emotion: How not to be fooled by Gerber and Wheeler (2009). *Perspectives on Psychological Science, 4,* 489–493.

Caspi, A., Roberts, B., & Shiner, R. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453–484.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York, NY: World Book.

Cattell, R. B. (1964a). Beyond validity and reliability: Some further concepts and coefficients for evaluating tests. *Journal of Experimental Education, 33,* 133–143.

Cattell, R. B. (1964b). Validity and reliability: A proposed more basic set of concepts. *Journal of Educational Psychology, 55,* 1–22.

Clark, L. (2009). Stability and change in personality disorder. *Current Directions in Psychological Science, 18,* 27–31.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Costa, P., & McCrae, R. (2006). Age changes in personality and their origins: Comment on Roberts, Walton, and Viechtbauer (2006). *Psychological Bulletin, 132,* 28–30.

de Maat, S., de Jonghe, F., Schoevers, R., & Dekker, J. (2009). The effectiveness of long-term psychoanalytic therapy: A systematic review of empirical studies. *Harvard Review of Psychiatry, 17,* 1–23.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40,* 532–538.

Fraley, R., & Roberts, B. (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.

Grilo, C. (2002). Are there gender differences in *DSM–IV* personality disorders? *Comprehensive Psychiatry, 43,* 427–430.

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

Hynan, D. (2004). Unsupported gender differences on some personality disorder scales of the Millon Clinical Multiaxial Inventory–III. *Professional Psychology: Research and Practice, 35,* 105–110.

Johnson, W., McGue, M., & Krueger, R. (2005). Personality stability in late adulthood: A behavioral genetic analysis. *Journal of Personality, 73,* 523–551.

Laptook, R., Klein, D., & Dougherty, L. (2006). Ten-year stability of depressive personality disorder in depressed outpatients. *American Journal of Psychiatry, 163,* 865–871.

Lenzenweger, M. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry, 56,* 1009–1015.

McCrae, R. (2000). Trait psychology and the revival of personality and culture studies. *American Behavioral Scientist, 44,* 10–31.

McCrae, R., & Costa, P. (1982). Self-concept and the stability of personality: Cross-sectional comparisons of self-reports and ratings. *Journal of Personality and Social Psychology, 43,* 1282–1292.

McCrae, R., & Costa, P. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory.* New York, NY: McGraw–Hill.

Ozer, D. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

Roberts, B., & DelVecchio, W. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3–25.

Roberts, B., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science, 17,* 31–35.

Roberts, B., Walton, K., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25.

Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality, 38,* 319–350.

Wong, S., Lee, B., Ang, R., Oei, T., & Ng, A. (2009). Personality, health, and coping: A cross-national study. *Behavior Science Research, 43,* 251–279.

Received March 2, 2009
Revision received October 5, 2009
Accepted November 30, 2009 ■