A Reliability Generalization Meta-Analysis of Self-Report Measures of Adult Attachment

James M. Graham & Marta S. Unterschute

Department of Psychology, Western Washington University

Published online: 25 Jun 2014.

To cite this article: James M. Graham & Marta S. Unterschute (2014): A Reliability Generalization Meta-Analysis of Self-Report Measures of Adult Attachment, Journal of Personality Assessment, DOI: 10.1080/00223891.2014.927768

To link to this article: http://dx.doi.org/10.1080/00223891.2014.927768

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
A Reliability Generalization Meta-Analysis of Self-Report Measures of Adult Attachment

JAMES M. GRAHAM AND MARTA S. UNTERSCHUTE

Department of Psychology, Western Washington University

This study is a reliability generalization meta-analysis that reviews 5 of the most frequently used continuous measures of adult attachment security: the Adult Attachment Scale, Revised Adult Attachment Scale, Adult Attachment Questionnaire, Experiences in Close Relationships, and Experiences in Close Relationships—Revised. A total of 313,462 individuals from 564 studies provided 1,629 internal consistency reliability estimates for this meta-analysis. We present the average internal consistency reliability of scores for each measure and test the consistency of score reliabilities across a wide variety of sample characteristics. In light of this, we highlight several issues in the measurement of adult attachment security and make concrete recommendations for researchers seeking to measure adult attachment.

Adult attachment theory is based largely on the work of Bowlby (1969/1982, 1973, 1980), who focused on the idea of a “working model,” or an individual’s internal representations of the world and of what to expect from relationships with others. Since that initial work, the concept of attachment has been applied to a variety of relationships beyond the infant–caregiver relationship. Most notably, attachment theory has been applied to adult romantic relationships. Although adult attachments tend to be more complex and have a greater capacity for reciprocity than infant–caregiver relationships (Weiss, 1982), the same basic concepts apply. A large body of research supports the prototype hypothesis in adult romantic relationships, which posits that our internal representations of past experiences shape our expectations and behavior about future relationships (Simpson & Rholes, 2010).

Adult attachment security is perhaps one of the most frequently studied topics in the current romantic relationship literature. A vast body of research suggests that adult attachment security is related to a wide variety of outcomes in romantic relationships. Individuals with avoidant attachment styles provide less support to their romantic partners (Westmaas & Silver, 2001). Individuals with high levels of attachment security enjoy more satisfying romantic relationships (e.g., Collins & Read, 1990) and experience better mental health (Galynter et al., 2012; Marganska, Gallagher, & Miranda, 2013) and well-being (Karremann & Vingerhoets, 2012) than those who are less securely attached.

As with any well-established research domain, a host of tools for measuring adult attachment have emerged. Attempts to measure adult attachment security have included interviews (Hesse, 2008), single-item categorical measures (Hazan & Shaver, 1987), behavioral observation (Wampler, Rigg, & Kimball, 2004), the coding of narratives (George & West, 2001), and multi-item self-report inventories (Bartholomew & Horowitz, 1991). Although attachment researchers have used a wide variety of measurement approaches, multi-item self-report inventories seem to have emerged as the most common approach, likely due to the ease and lack of expense in obtaining self-report survey data.

Whereas early attempts at measuring attachment involved attempting to classify individuals into specific categories (e.g., Hazan & Shaver, 1987), subsequent research has suggested that adult attachment does not fit a taxonic model (Fraley & Waller, 1998) and has highlighted the problems in power, discriminatory ability, and the misrepresentation of data that accompany categorizing continuous data. Continuous measures of attachment allow for finer distinctions to be made among research participants, and allow research questions to address not only types of attachment, but also levels of attachment (Fairchild & Finney, 2006).

Perhaps in part due to the wide scope of research on adult attachment security, a single, universally used measure has not emerged (Fairchild & Finney, 2006). The diversity of measures used in studying adult attachment security provides advantages and disadvantages for researchers. Multiple measurement techniques benefit researchers by helping to ensure that findings are not simply an artifact of a specific measurement instrument. Conversely, the use of diverse measures can make it difficult to connect findings across research groups. Attachment research occurs in many countries and with widely diverse populations; as such, the ability to draw parallels across these areas of research is important. Because the psychometric properties of scores produced by these instruments can vary extensively across different populations, it is essential to find a measure that reduces the effect of sample characteristics on the produced scores. Furthermore, high-quality substantive research requires access to scores of high psychometric quality and of sufficient reliability.

Reliability, or the proportion of nonerror variance in a set of scores, is of central importance in substantive research. Reliability serves as an upper limit for any related effect size, with unreliable measurement resulting in reduced power to detect meaningful effects (Baugh, 2002; Johnson, 1944). As such, unreliable scores undermine researchers’ ability to answer substantive questions. Many researchers erroneously believe that reliability is a property of a measure (Thompson, 2002).
Although it is true that characteristics of a measure can influence reliability, the characteristics of the sample completing the measure as well as the setting in which the measure was completed could also influence reliability. As such, the scores produced by a measure are reliable, not the measure itself. A measure might obtain highly reliable scores from one sample, but obtain unreliable scores from a different sample. Such differences in reliability can affect researchers’ conclusions regarding their study. For example, when making cross-group comparisons, unequal reliability across groups might give the impression that group membership is moderating the relation between two variables, when in fact it is not. Given that the scope of adult attachment research has expanded well beyond the populations in which the measures were initially developed, a thorough understanding of the effect of sample characteristics on score reliability is essential.

In this study, we present a reliability generalization (RG) meta-analysis of five of the most commonly used self-report measures of adult attachment security. Specifically, we focus on the Adult Attachment Scale (AAS; Collins & Read, 1990), Revised Adult Attachment Scale (RAAS; Collins, 1996), Adult Attachment Questionnaire (AAQ; Simpson, Rholes, & Nelligan, 1992), Experiences in Close Relationships (ECR; Brennan, Clark, & Shaver, 1998), and the Experiences in Close Relationships–Revised (ECR–R; Fraley, Waller, & Brennan, 2000). RG meta-analysis allows researchers to characterize the average reliability of scores obtained by a measure across a wide variety of samples and settings (Vacha-Haase, 1998). We describe the average reliabilities of scores produced by these measures and the degree to which the reliability of these scores varies from study to study. Furthermore, we describe the sample and study characteristics that explain the variance in score reliabilities. In this way, we can assess these attachment measures for potential bias, and ascertain which measures tend to produce the most reliable scores for which type of people.

**AAS**

The AAS (Collins & Read, 1990) is an 18-item scale based on Hazan and Shaver’s (1987) adult attachment descriptors. Participants respond to items on a 5-point scale ranging from 1 (not at all characteristic of me) to 5 (very characteristic of me). The measure consists of three subscales, measuring fear and anxiety over being unloved and abandoned (Anxiety; e.g., “I often worry that my partner does not really love me”), comfort with closeness and intimacy (Close; e.g., “I find it relatively easy to get close to others”), and the extent to which one can trust and depend on others (Depend; e.g., “I know that others will be there when I need them”). The original sample used to develop the measure was comprised entirely of undergraduate students at a large U.S. research university. Subsequent research has shown that AAS scores correlate as expected with other measures of attachment (Sperling, Foelsch, & Grace, 1996), and that the AAS appears to produce appropriately valid scores. Collins and Read (1990) found that, in addition to the three attachment styles that were used to guide the creation of the measure, a cluster analysis also supported dividing individuals into four groups based on their AAS scores. Specifically, they divided the “anxiously attached” group into those who were anxious/secure and those who were anxious/avoidant. These four categories are reflective of the two-dimensional approach to measuring attachments that appeared in later measures.

**RAAS**

Collins (1996) later revised the AAS in an attempt to improve the reliability of scores produced by the measure and to correct several poorly worded items. The resulting RAAS is an 18-item test with 6 items each measuring attachment across the three dimensions of the AAS: Close, Depend, and Anxiety. Like the AAS, participants responded to items on a 5-point scale ranging from 1 (not at all characteristic of me) to 5 (very characteristic of me). The revisions for this scale were made using a sample comprised of U.S. undergraduate students. The Close and Depend subscales were highly correlated ($r = .53$) and Anxiety was negatively correlated with both Close and Depend ($r = -.34$ and $r = -.46$, respectively; Collins, 1996). The original and revised versions of the AAS were highly correlated ($r = .98$).

**AAQ**

The AAQ (Simpson et al., 1992) consists of 17 items that are measured on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Whereas the AAQ was based on Hazan and Shaver’s (1987) three attachment descriptors, Simpson et al. (1992) reduced the measure to two dimensions, Avoidance and Ambivalence. The 8-item Avoidance subscale (e.g., “I’m not very comfortable having to depend on other people”) measures the degree to which one withdraws from closeness and intimacy, and is related to Close and Depend scales of the AAS. The 9-item Ambivalence subscale (e.g., “I often worry that my partner(s) don’t really love me”) reflects the degree to which one has conflicted feelings in close relationships, and is related to the Anxiety subscale of the AAS. The AAQ was developed using a sample of dating undergraduate students in a U.S. university. Although the AAQ was one of the first measures used to demonstrate that adult attachment might be best described as two orthogonal dimensions, the validity of AAQ scores has been called into question, as the dimensions of the AAQ appear to be unrelated to the results of the Adult Attachment Interview (Bouthillier, Julien, Dube, Belanger, & Hamelin, 2002).

**ECR**

The ECR (Brennan et al., 1998) is a 36-item self-report measure using a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Brennan and colleagues used responses from a U.S. undergraduate sample to create the ECR. They pooled 482 items obtained from existing measures and subjected them to a factor analysis. Two continuous dimensions emerged, and Brennan and colleagues selected the 18 highest loading items for each subscale. The items making up the subscales were consistent with avoidance, or the degree to which one is uncomfortable with closeness and avoids intimacy (e.g., “I prefer not to show a partner how I feel deep down”), and anxiety, or the degree to which one is anxious about being abandoned (e.g., “I worry about being
ADULT ATTACHMENT RELIABILITY GENERALIZATION

The authors reported a low correlation between the two subscales ($r = .11$), suggesting two independent factors, although subsequent research has found these two subscales to be more highly correlated (Cameron, Finnegan, & Morry, 2012). The ECR has been used extensively, and has been translated into a number of languages, including Chinese (Mallinckrodt & Wang, 2004), Dutch (Conradi, Gerlsma, van Duijn, & de Jonge, 2006), French (Lafontaine & Lussier, 2003), German (Neumann, Rohmann, & Bierhoff, 2007), Hebrew (Mikulincer & Florian, 2000), Japanese (Nakao & Kato, 2004), Norwegian (Olsson, Sorebo, & Dahl, 2010), Spanish (Alonso-Arbiol, Balluerka, & Shaver, 2007), and Turkish (Sümer, 2006). A shortened version of the ECR has also been developed (Wei, Russell, Mallinckrodt, & Vogel, 2007).

ECR–R

Fraley and colleagues (2000) undertook a revision of the ECR, and created the ECR–R. The researchers used item-response theory (IRT) to analyze 323 items from four commonly used adult attachment measures and selected 18 items for inclusion in each subscale. Like the ECR, the ECR–R uses a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). The ECR–R contains the same Avoidance (e.g., ‘I prefer not to show a partner how I feel deep down’) and Anxiety (e.g., ‘I’m afraid that I will lose my partner’s love’) dimensions as the ECR. Confirmatory factor analysis has continued to support a two-factor model over treating attachment as a single, unidimensional construct (Sibley & Liu, 2004). ECR–R scores have been shown to be stable over a 3-week (Sibley, Fischer, & Liu, 2005) and a 6-week period (Sibley & Liu, 2004). Avoidance correlated positively with touch avoidance and loneliness and correlated negatively with affectionate proximity and social support (Fairchild & Finney, 2006). The Anxiety subscale was positively correlated with loneliness and worry and was negatively correlated with social support (Fairchild & Finney, 2006). The ECR–R has been found to provide more precise estimates for individuals with insecure attachment (high scores on both Anxiety and Avoidance subscales) rather than secure attachment (low scores on both Anxiety and Avoidance subscales; Fraley et al., 2000). The ECR–R has been translated into a variety of languages, including Arabic (Hijazi, 2004), Dutch (Buysse & Dewitte, 2004), German (Ehrenthal, Dinger, Lamla, Funken, & Schauenburg, 2009), Greek (Tsagarakis, Kafetsios, & Stalikas, 2007), Korean (Lee, Grossman, & Krishnan, 2008), Serbian (Hanak & Dimitrijevic, 2013), Thai (Wongpakaran, Wongpakaran, & Wannarit, 2011), and Turkish (Selcuk, Gunaydin, Sumer, & Uysal, 2005).

The current consensus among the research community appears to be that adult attachment security is best conceptualized across two dimensions: attachment anxiety and attachment avoidance (Mikulincer, Shaver, & Pereg, 2003). This two-factor conceptualization is used in a variety of measures (Stein et al., 2002), and explains both the two-factor structures of the ECR and the AAQ, as well as the three-factor structure of the AAS, for which closeness and dependency could be combined into a single avoidance subscale. Although some researchers assume that anxiety and avoidance should be orthogonal dimensions (see Mikulincer, Shaver, Sapir-Lavid, & Avihou-Kanza, 2009), others argue that anxiety and avoidance are distinct, but correlated (Baldwin, Keelan, Fehr, Enns, & Kohn-Rangarajo, 1996; Bartholomew, 1990). A previous meta-analysis of attachment measures suggests that the correlation between anxiety and avoidance across studies is approximately .20 (Cameron et al., 2012). This correlation, however, also appears to differ by measure, with correlations between the ECR–R subscales being higher than the correlations between the ECR subscales.

This Study

This study uses RG meta-analysis to examine self-report measures of adult attachment security. This meta-analysis has several goals. First, we seek to characterize the typical reliability of scores from five of the most commonly used measures of adult attachment security. Although there are multiple types of reliability, we chose to focus exclusively on Cronbach’s alpha, as it is the most widely used measure of reliability for multi-item self-report surveys. Reliability is of central importance for any research, and information about the typical reliability of scores produced by a measure can guide researchers when making decisions about measurement selection. Second, we seek to characterize the variability in reliability across studies for each of the measures in question. Finally, we examine the influence of a variety of sample characteristics on score reliability. This will allow us to not simply identify which measures produce the most reliable scores, but which measures produce the most reliable scores with which type of respondents.

Method

We identified studies for potential inclusion in the meta-analysis through a variety of means. We used a cited reference search to identify every article in PsycINFO and Web of Science that contained a reference to the work in which one of the five measures under consideration first appeared. We conducted additional searches through both search engines using a variety of search terms, including adult attachment style, adult attachment, and romantic attachment. We restricted these searches to only articles published later than the year in which the earliest measure under consideration, the AAS, was published. After deleting duplicates, we identified 2,001 articles for potential inclusion in the analysis.

We made attempts to obtain each of the 2,001 articles. We could not obtain 11 of the articles through normal interlibrary loan channels. We classified and coded the articles by consensus. The study authors or at least two trained undergraduate research assistants independently examined each of the obtained articles. A third reviewer addressed any coding discrepancies. Approximately 7% of studies were reexamined and corrected in this manner because of coding discrepancies. Seventy-five articles were written in languages other than English and were discarded, as the study authors and research assistants did not possess fluency in languages other than English sufficient for coding the articles. Inclusion in the present meta-analysis required that the articles include original...
Table 1.—Frequency (and percentage of row) of reliability reporting practices by adult attachment measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>No Mention</th>
<th>No Values</th>
<th>Values</th>
<th>Unusable</th>
<th>Usable</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAQ</td>
<td>9 (8.9%)</td>
<td>25 (4.0%)</td>
<td>11 (1.8%)</td>
<td>9 (5.2%)</td>
<td>503 (81.3%)</td>
</tr>
<tr>
<td>ECR</td>
<td>55 (8.9%)</td>
<td>25 (4.0%)</td>
<td>11 (1.8%)</td>
<td>25 (4.0%)</td>
<td>503 (81.3%)</td>
</tr>
<tr>
<td>ECR–R</td>
<td>9 (5.2%)</td>
<td>2 (1.2%)</td>
<td>2 (1.2%)</td>
<td>9 (5.2%)</td>
<td>150 (87.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>112 (14.0%)</td>
<td>41 (5.1%)</td>
<td>19 (2.4%)</td>
<td>62 (7.8%)</td>
<td>564 (70.7%)</td>
</tr>
</tbody>
</table>

Note. The total row numbers reflect the number of unique articles that used each reporting type. Because some articles used multiple measures, the total does not reflect the sum of the measures. AAS = Adult Attachment Scale; AAQ = Adult Attachment Questionnaire; ECR = Experiences in Close Relationships; ECR–R = Experiences in Close Relationships–Revised.

data collection using one or more of the study measures; as such, we discarded another 1,124 articles that failed to meet this criterion.1 We then examined the final sample of 791 articles in detail.

The study authors or at least two trained undergraduate research assistants classified the remaining 791 articles depending on whether and how the authors reported the reliability of the scores. Table 1 shows the results of these examinations by measure. Across all measures, 14.16% of the articles examined made no mention of reliability. The authors of some articles engaged in reliability induction, by either stating that the measure had been found to be reliable in the past (5%) or by reporting the reliability coefficient that was obtained by the original measure authors during the development of the measure (2%). Reliability induction involves making the potentially erroneous assumption that, because a measure had produced reliable scores at a single time in the past, it would likely do so in any future uses (Vacha-Haase, Kogan, & Thompson, 2000). Approximately 8% of articles reported the reliability for their data, but in a form that made it impossible to use in this meta-analysis (e.g., provided a range of estimates, or stated the alphas were higher than .8). Seventy-one percent of the final sample of articles reported exact values of Cronbach’s alpha for their own data, and could be used in this meta-analysis. This percentage of articles providing usable reliability coefficients is quite high compared to other RG studies, which have typical reporting averaging around 18.9% (Vacha-Haase & Thompson, 2011). Because articles might have included reliability coefficients for multiple studies or samples, the 564 articles reporting reliability for data in hand provided 1,627 reliability coefficients across 833 samples and 313,462 individuals.

The reviewers coded a variety of sample and study characteristics for those articles that reported usable reliability coefficients. In all cases, the reviewers aggregated the sample and study characteristics to the same level at which the reliability of scores was reported. For example, if the author of the study under review reported a different reliability coefficient for men and women, then the reviewers recorded the sample characteristics separately for men and women. The reviewers recorded the size of the sample, the average age of the sample, the gender of the sample (coded as percent male), an indication of whether the sample was obtained from an undergraduate convenience sample (coded as percent college), and an indication of how much of the sample was obtained from the United States versus another country (coded as percent international). The reviewers also noted the sexual orientation of the participants (coded as percent heterosexual), the relationship status of participants (coded as percent of sample in a current romantic relationship), the marital status of participants (coded as percent married), and the mean length of the participants’ relationships, for those currently in romantic relationships. Finally, reviewers also noted the year in which the study was published, to ascertain how the reliability of scores produced by a given measure might change over time.

RESULTS AND DISCUSSION

Analytic Strategy

To prepare the data for analysis, we first converted the reliability coefficients into appropriate form. Because a reliability coefficient is considered a variance-accounted-for ($r^2$) statistic, we took the square root of each reliability coefficient and converted it into a form amenable for further analysis with Fisher’s $r$-to-$z$ transformation. We weighted all analyses using the inverse variance weights of the reliability coefficients (in this case, $n = 3$). We converted all appropriate results back into the original form, so that mean reliabilities are reported in the same metric as Cronbach’s alpha. To analyze the data in this meta-analysis, we used Wilson’s (2005) meta-analysis macros for SPSS. These macros use maximum likelihood estimation to calculate random-effects estimates.

Table 2 shows the descriptive statistics for the reliability of scores produced by each of the measures and the number of studies contributing to the descriptive statistics ($k$). We describe these results by measure later. Using Wilson’s (2005) meta-analysis macros for SPSS, we calculated the random-effects average of Cronbach’s alpha for each subscale, and computed a 95% confidence interval about that mean. Given that these estimates are based on meta-analytic data, they represent the best available population estimates for the reliability of scores produced by these measures. We considered the reliability for each of these measures relative to Nunnally’s (1978) minimal acceptable reliability of .70. Although .70 is a somewhat arbitrary number, owing to the fact that the level of sufficient reliability depends on the manner in which the scores will be used, it represented the most commonly referenced minimal acceptable reliability for research purposes.

Because score reliability is influenced by both interitem correlations and the number of items in a measure (with higher interitem correlations and longer tests producing more reliable results), we also calculated the average interitem correlation from the average reliability coefficient. Table 2 shows these results. It is important to note that although the average interitem correlation is of some interest to test developers, it should not be used to equate tests of different lengths. Adding more items to a test will only increase the reliability test scores if the interitem correlations of the additional items are of sufficient strength, and taking items away from a longer test will

1Studies that failed to meet this criterion included a heterogeneous group of studies, ranging from theoretical articles, to qualitative studies, and to articles that only incidentally referenced the articles in which the measures were initially developed.
not necessarily result in less reliable scores (Niemi, Carmines, & McIver, 1986). As such, the average interitem correlations should not be used to assess what might happen if items were removed from or added to a test.

To examine the variance in reliability coefficients across studies and samples we used Q-tests of homogeneity and the $I^2$ measure of effect size heterogeneity. Cochran’s Q involves calculating the sum of the weighted squared differences between the individual study effects and the average pooled effect, and testing the degree to which that variance differs from zero with a chi-square distribution. Because the Q-test uses statistical significance testing, and because of the large sample sizes used in meta-analyses such as ours, the Q-test does not provide a good measure of the degree of variability among the reliability coefficients. As an alternative to the Q-test, we also report $I^2$, which measures the percentage of effect size variability that does not appear to be due to random chance, with higher values indicating more significant heterogeneity between reliability coefficients (Higgins & Thompson, 2002).

Because unreported reliability coefficients might adversely affect the results of the present meta-analysis, we used a derivative of Orwin’s (1983) fail-safe N to consider the effect of unreported reliability coefficients on this meta-analysis’s results. In typical meta-analyses, Orwin’s fail-safe N indicates the number of studies with null findings needed to make the average effect size statistically nonsignificant. Because reliability coefficients of zero would be highly unlikely, we calculated the number of studies reporting a reliability coefficient of .50 that would be required to bring the average reliability coefficient below .70. To do so, we converted all reliability coefficients to $z$ equivalents, and used the following derivation of Orwin’s equation:

$$N = k (E_{Sm} - E_s) / E_s - E_{Sa}$$  \hspace{1cm} (1)

where $k$ is the number of studies and $E_{Sm}$ is the average reliability across studies. It takes $N$ studies with a reliability of $E_s$ to create an average reliability of $E_{Sa}$.

Next, we considered the relations between the reliability coefficients and a variety of sample and study characteristics. We assessed which sample characteristics explain a meaningful amount of the variance in reliability coefficients. Although one would ideally assess the effect of multiple sample characteristics on score reliability simultaneously, the data needed to calculate these demographic variables are not consistently reported in the available literature. As such, assessing the simultaneous effect of these characteristics on score reliability would result in an unacceptable loss of data due to listwise deletion. We have taken two approaches to address this issue. First, we present the bivariate correlations between score reliabilities and sample characteristics in Table 3. Although these correlations make maximum use of the available data, they result in an inflated experiment-wise error rate. For the sake of completeness, we reported correlations calculated from fewer than 10 studies, but did not interpret these results as meaningful. Second, we selected five of the most commonly reported sample characteristics (mean age and percentage of the sample that is male, collected from a university subject pool, collected from a non-U.S. sample, and White) and predicted the reliability coefficients with random effects multiple regression. In cases where the inclusion of all five variables resulted in an unacceptable reduction of the available sample size, we used a smaller selection of these variables. In cases where the correlation between predictors was greater than .70, we removed the predictor that resulted in the smallest reduction in $k$ to avoid issues associated with multicollinearity (Tabachnick & Fidell, 2006). We show the results of the regression analyses in Table 4 and discuss them by measure next. We conducted the correlations and regression analyses using Wilson’s (2005) SPSS macros, which use maximum likelihood estimation to conduct weighted random-effects regression.

### AAS

As seen in Table 2, the average reliability of scores produced by the AAS ranged from .733 to .761 across subscales. For each subscale, the lower and upper bounds of the 95% confidence interval places the average reliability of scores produced by the AAS squarely in the .70 range, with the lower bounds of each of the confidence intervals staying above .70. This suggests an overall level of reliability that is considered acceptable for research purposes, but still less than optimal. The average interitem correlations between AAS items were
Table 3.—Bivariate random-effects correlations (and pairwise $k$) between reliability and sample characteristics.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subscale</th>
<th>% Age</th>
<th>% Male</th>
<th>% White</th>
<th>% Non-U.S.</th>
<th>% College</th>
<th>Current Relationship</th>
<th>Relationship Length</th>
<th>% Married</th>
<th>% Heterosexual</th>
<th>% Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAS</td>
<td>Close</td>
<td>.27*</td>
<td>.56***</td>
<td>.66</td>
<td>.42*</td>
<td>.48</td>
<td>.55**</td>
<td>.13</td>
<td>.18</td>
<td>.28*</td>
<td></td>
</tr>
<tr>
<td>Depend</td>
<td>.25</td>
<td>.32*</td>
<td>.47***</td>
<td>.37**</td>
<td>.18</td>
<td>.62*</td>
<td>.19</td>
<td>.12</td>
<td>.28*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>.01</td>
<td>.02</td>
<td>.24</td>
<td>.11</td>
<td>.02</td>
<td>-.15</td>
<td>.11</td>
<td>-.39*</td>
<td>.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAAS</td>
<td>Close</td>
<td>-.31</td>
<td>.31</td>
<td>.18</td>
<td>-.63***</td>
<td>.58**</td>
<td>.33</td>
<td>-.84**</td>
<td>.76</td>
<td>-.23</td>
<td></td>
</tr>
<tr>
<td>Depend</td>
<td>-.43*</td>
<td>-.06</td>
<td>-.33</td>
<td>-.58***</td>
<td>-.09</td>
<td>.48</td>
<td>-.87*</td>
<td>.70</td>
<td>-.61</td>
<td>-.40*</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>.08</td>
<td>.24</td>
<td>-.66**</td>
<td>-.47**</td>
<td>.10</td>
<td>-.19</td>
<td>.52</td>
<td>.20</td>
<td>.12</td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>AAQ</td>
<td>Avoidance</td>
<td>.29</td>
<td>-.16</td>
<td>.81***</td>
<td>-.13</td>
<td>-.24</td>
<td>.05</td>
<td>.30</td>
<td>.22</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Ambivalence</td>
<td>.00</td>
<td>-.32*</td>
<td>.08</td>
<td>.11</td>
<td>-.07</td>
<td>-.16</td>
<td>-.25</td>
<td>-.18</td>
<td>.29</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>ECR</td>
<td>Anxiety</td>
<td>-.10</td>
<td>-.04</td>
<td>-.06</td>
<td>-.23***</td>
<td>.16***</td>
<td>-.14</td>
<td>-.30***</td>
<td>-.48***</td>
<td>-.28*</td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td>-.22***</td>
<td>-.11*</td>
<td>-.01</td>
<td>-.31***</td>
<td>-.27***</td>
<td>-.17**</td>
<td>-.38***</td>
<td>-.41***</td>
<td>-.26**</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td>ECR–R</td>
<td>Anxiety</td>
<td>.10</td>
<td>-.11</td>
<td>-.17</td>
<td>-.01</td>
<td>-.05</td>
<td>.07</td>
<td>-.07</td>
<td>.01</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td>-.05</td>
<td>-.09</td>
<td>-.32*</td>
<td>-.17</td>
<td>.07</td>
<td>.17</td>
<td>-.15</td>
<td>.09</td>
<td>.17</td>
<td>-.09</td>
<td></td>
</tr>
</tbody>
</table>

Note. AAS = Adult Attachment Scale; RAAS = Revised Adult Attachment Scale; AAQ = Adult Attachment Questionnaire; ECR = Experiences with Close Relationships; ECR–R = Experiences with Close Relationships–Revised.

I2 statistics suggest that there is a meaningful amount of variance in reliabilities across studies. As such, a single population estimate does not seem reasonable.

As seen in Table 3, a number of the correlations between AAS reliabilities and sample characteristics are statistically significant. The Close subscale of the AAS produces moderate. These results appeared moderately susceptible to the effect of publication bias—the existence of 20% to 50% more studies with reliabilities of .50 would drop the average reliability below .70. The $I^2$ and $R^2$ statistics suggest that there is a meaningful amount of variance in reliabilities across studies. As such, a single population estimate does not seem reasonable.

Table 4.—Random-effects multiple regression results predicting reliability with sample characteristics.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subscale</th>
<th>Model</th>
<th>Predictor</th>
<th>$R^2$</th>
<th>$k$</th>
<th>$p$</th>
<th>Age</th>
<th>Male</th>
<th>College</th>
<th>Non-U.S.</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAS</td>
<td>Close</td>
<td>.436</td>
<td>.39</td>
<td>&lt;.001</td>
<td>-.03</td>
<td>.31*</td>
<td>-.02</td>
<td>-.22</td>
<td>.55***</td>
<td>.47***</td>
<td></td>
</tr>
<tr>
<td>Depend</td>
<td></td>
<td>.561</td>
<td>.38</td>
<td>&lt;.001</td>
<td>.26</td>
<td>-.03</td>
<td>.27*</td>
<td>.50***</td>
<td>.47***</td>
<td>.47***</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td>.174</td>
<td>.48</td>
<td>.061</td>
<td>-.10</td>
<td>.08</td>
<td>.06</td>
<td>.34*</td>
<td>.23</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>RAAS</td>
<td>Close</td>
<td>.396</td>
<td>.19</td>
<td>.005</td>
<td>-.24</td>
<td>.22</td>
<td>.43*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Depend</td>
<td></td>
<td>.267</td>
<td>.19</td>
<td>.064</td>
<td>-.59***</td>
<td>.21</td>
<td>-.33</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td>.060</td>
<td>.24</td>
<td>.641</td>
<td>.03</td>
<td>.22</td>
<td>.07</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>AAQ</td>
<td>Avoidance</td>
<td>.103</td>
<td>.47</td>
<td>.208</td>
<td>—</td>
<td>-.18</td>
<td>-.24</td>
<td>-.14</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Ambivalence</td>
<td></td>
<td>.117</td>
<td>.44</td>
<td>.115</td>
<td>—</td>
<td>-.32*</td>
<td>-.08</td>
<td>.09</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>ECR</td>
<td>Anxiety</td>
<td>.018</td>
<td>.188</td>
<td>.635</td>
<td>.15</td>
<td>-.05</td>
<td>.09</td>
<td>-.10</td>
<td>.03</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td></td>
<td>.047</td>
<td>.188</td>
<td>.107</td>
<td>.10</td>
<td>-.15</td>
<td>.21*</td>
<td>-.12</td>
<td>.06</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>ECR–R</td>
<td>Anxiety</td>
<td>.145</td>
<td>.59</td>
<td>.074</td>
<td>.37</td>
<td>-.05</td>
<td>.22</td>
<td>.09</td>
<td>-.23</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td></td>
<td>.179</td>
<td>.56</td>
<td>.043</td>
<td>.30</td>
<td>-.07</td>
<td>.33</td>
<td>-.08</td>
<td>-.34***</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

Note. AAS = Adult Attachment Scale; RAAS = Revised Adult Attachment Scale; AAQ = Adult Attachment Questionnaire; ECR = Experiences with Close Relationships; ECR–R = Experiences with Close Relationships–Revised.

*p < .05. **p < .01. ***p < .001.
statistically significantly more reliable scores with men, White individuals, individuals currently involved in a romantic relationship, and married individuals. The Depend subscale of the AAS produces statistically significantly more reliable scores with White individuals, U.S. samples, and college students. Furthermore, the Depend subscale reliabilities have decreased over time to a statistically significant degree. The Anxiety subscale of the AAS is influenced less by sample characteristics than the other subscales, producing statistically significantly more reliable scores with gays and lesbians than with heterosexual couples.

As shown in Table 4, the age of the sample and the percentage of the sample that was male, college students, from outside the United States, and White failed to predict a statistically significant amount of the variance in AAS Anxiety scores. However, these five predictors explained a large amount of the variance in Close (44%) and Depend (56%) subscale reliabilities. Most notably, both subscales provided less reliable scores with non-White than with White populations, and the Depend subscale provided much less reliable scores when administered to individuals from countries other than the United States.

From the standpoint of reliability, the AAS appears to be a poor choice for a measure of adult attachment. The overall reliabilities, although in a range considered generally acceptable for research purposes, were the lowest of the measures examined in this meta-analysis. Furthermore, the reliability of the Close and Depend subscales appeared to be highly susceptible to sample characteristics. Specifically, the measure appears to produce the most reliable scores when administered to White, college-age, U.S. students. This represents the population on which the measure was initially developed. As such, we encourage caution when using the AAS for studying adult attachment across diverse groups and countries. Instead, we encourage researchers desiring a measure of attachment similar to the AAS to consider the RAAS. The reliability of Anxiety subscale scores appears more stable across samples.

**RAAS**

As seen in Table 2, the average reliability of scores produced by the RAAS was notably improved over the AAS, ranging from .768 to .855 across subscales. On average, the reliability of the RAAS scores was higher than that of their associated AAS scores. This suggests that the revisions undertaken by Collins (1996) were effective in improving the overall reliability of scores. The average interitem correlations between RAAS items were moderate, save for items from the Anxiety subscale, which were substantially larger. These results appeared fairly robust to the effects of publication bias, which needs 40% to 86% more studies with low reliabilities to drop the average reliabilities below .70. There was a meaningful amount of variance in reliability scores across studies.

Because the more widely used ECR came into regular use 2 years after the publication of the RAAS, there were fewer data available for the RAAS than for other measures. As such, although some of the correlations between RAAS reliabilities and sample characteristics are higher than those for AAS reliabilities, they fail to achieve statistical significance due to the decreased sample size. All of the RAAS subscales produced statistically significantly less reliable scores in non-U.S. than in U.S. samples. The Close subscale produces more reliable scores in college student populations than in non-college-student populations. The reliability of the Depend subscale scores is lower in older individuals than in younger individuals. The Anxiety subscale produces more reliable scores in non-White than in White populations. Several other correlations emerged as statistically significant, but had fewer than 10 reliability coefficients contributing valid data.

As shown in Table 4, the percentage of the sample that was male, college students, and from outside the United States did not predict a statistically significant amount of the variance in RAAS scores across subscales. The lack of statistical significance was largely due to the small number of studies contributing data, given that the $R^2$ values for both the Close and Depend subscales were large. In the context of the other predictors, the Close subscale produced more reliable results with college students, and the Depend subscale produced more reliable results with younger participants.

Collins (1996) appears to have been successful at the stated goal of improving the psychometric properties of the AAS. The average reliability of RAAS scores is higher than those of AAS scores. The reliability of the Anxiety subscale (.855) is particularly noteworthy, considering that it is only six items long. Because Cronbach’s alpha is a function of item length and average interitem correlation, this suggests that the items making up the RAAS Anxiety subscale are strongly interrelated. The Anxiety subscale might produce particularly reliable attachment anxiety scores, for those preferring a shorter measure. Although sufficient data are not available to make broader conclusions, the reliability of RAAS scores does seem consistently biased in favor of U.S. samples. As such, we encourage caution when using the RAAS in non-U.S. samples and, in particular, when making cross-national comparisons.

**AAQ**

As seen in Table 2, the average reliability of scores produced by the AAQ was .785 for both the Avoidance and Ambivalence subscales. For each subscale, the lower and upper bounds of the 95% confidence interval place the average reliability of scores produced by the AAQ in the high .70s range. This suggests an overall level of reliability that is considered acceptable for research purposes, and is generally comparable with reliabilities from the AAS and RAAS. The average interitem correlations between AAQ items were moderate, although the lowest of the measures examined here. The reliability of AAQ scores appeared moderately susceptible to the effect of publication bias—the existence of 20% to 50% more studies with reliabilities of .50 would drop the average reliabilities below .70. The $Q$ and $I^2$ statistics suggest that there is a meaningful amount of variance in reliabilities across those studied. As such, a simple population estimate does not seem reasonable.

As seen in Table 3, very few of the correlations between AAQ reliabilities and sample characteristics were statistically significant. The Avoidance subscale produced statistically significantly more reliable scores with White than with non-White individuals. The Ambivalence subscale reliabilities produced statistically significantly more reliable scores with women than with men. None of the other correlations emerged as statistically significant, suggesting that the reliability of
AAQ scores is relatively stable across samples. As shown in Table 4, the percentage of the sample that was male, college students, and from outside the United States failed to predict a statistically significant amount of the variance in AAQ Avoidance or Ambivalence scores.

The average reliability of AAQ scores appear on par with the AAS and RAAS. Furthermore, the reliability of the Avoidance and Ambivalence subscales appeared to be relatively robust to the effect of sample characteristics. With the exception of the Avoidance subscale producing more reliable scores in White than in non-White participants, the AAQ appears to be a good potential choice for studying diverse samples and making cross-group comparisons.

**ECR**

As seen in Table 2, the average reliability of scores produced by the ECR was .892 and .898 for the Anxiety and Avoidance scores, respectively, suggesting strong overall reliability. For each subscale, the lower and upper bounds of the 95% confidence interval place the average reliability of scores produced by the ECR squarely at about .89. The average inter-item correlations between ECR items were moderate. The reliability of ECR scores was quite robust to the possible effect of publication bias—more than double the number of existing studies with reliability of .50 would be needed to drop the overall reliability below .70. Although the overall high reliability of ECR scores did not appear to be the result of publication bias, there was a meaningful amount of variance in reliability scores across studies.

As seen in Table 3, a large number of the correlations between ECR reliabilities and sample characteristics were statistically significant. In part, this is a function of the wide use of the ECR, and the number of reliability coefficients reported. Because of the large effective sample size, correlation coefficients (even if relatively modest in size) are more likely to be statistically significant for the ECR in comparison to the other measures examined here. Large sample size notwithstanding, both subscales of the ECR tend to produce more reliable scores with nonmarried versus married individuals, with early-stage versus well-established relationships, with gays and lesbians versus heterosexual individuals, with individuals from inside the United States versus outside the United States, and with college student samples versus non-college-student samples. Furthermore, the ECR Avoidance subscale appears to produce more reliable scores in younger, versus older, samples.

As shown in Table 4, the age of the sample and the percentage of the sample that was male, college students, from outside the United States, and White failed to predict a statistically significant amount of the variance in ECR scores, although the percentage of the sample that were college students was statistically significantly related to the Avoidance subscale reliability. This contrasts with the information obtained from the bivariate correlations, suggesting that the reliability of ECR scores might be somewhat robust to the effect of sample characteristics. In part, the contrast between these results and those obtained from the bivariate correlations can be explained by the reduction in sample size necessitated by the listwise-deletion of cases in the regression analysis. Additionally, those variables included in the multiple regression analysis represented some of the lowest bivariate correlations.

Taken together, the results of the ECR analyses provide mixed results. When considering the average score reliabilities, the ECR appears to be a strong choice for a measure of adult attachment. The average reliabilities were quite high, and unlikely to be the result of publication bias. The relation between the reliability of the ECR subscales and sample characteristics was mixed. The reliability of ECR scores was relatively unaffected by many of the more common sample characteristics such as gender and percentage White. Of all of the sample characteristics, ECR scores appear most likely to be influenced by relationship length and marital status. The ECR produces more reliable scores in early-stage, unmarried relationships. Generally, the ECR seems to produce more reliable scores in the types of samples most commonly used in social psychology research conducted in the United States: unmarried American college students in early-stage relationships. Although it is often the case that the psychometric characteristics of scores produced by a measure are maximized in the population that was used to develop the measure, this procedure could create problems when making cross-group comparisons or applying the measure to more diverse populations. Examples of this issue with the ECR can be seen in the available literature. For example, Conradi and colleagues (2006) found that the reliability of ECR scores and the underlying factor structure were stronger in a Dutch college student sample than in a Dutch general population sample. Given the overall high level of reliability evidenced by ECR scores, the ECR is unlikely to produce scores of unacceptably low reliability in any of these groups. Rather, caution should be exhibited when using ECR scores to make cross-group comparisons.

**ECR–R**

As seen in Table 2, the average reliability of scores produced by the ECR–R was .897 and .908 for the Anxiety and Avoidance scores, respectively. For each subscale, the lower and upper bounds of the 95% confidence interval place the average reliability of scores produced by the ECR–R squarely at about .90. These reliabilities were the highest of all measures considered here, and within the range of what might be considered acceptable for assessing individuals, albeit not in high-stakes situations. The average interitem correlations between ECR–R items were moderate. The reliability of ECR–R scores was robust to the possible effect of publication bias—the number of studies with reliability of .50 would have to more than double the number of existing studies to drop the overall reliability below .70. The $Q$ and $I^2$ statistics suggest that there is a meaningful amount of variance in reliabilities across studies. As such, a single population estimate does not seem reasonable.

As seen in Table 3, only a single correlation between ECR–R reliabilities and sample characteristics emerged as statistically significant. The Avoidance subscale produced statistically significantly more reliable scores with non-White than with White participants. Otherwise, none of the other correlations between ECR–R reliabilities and sample characteristics were statistically significant. This relative lack of statistically significant findings is particularly noteworthy when you
consider the fact that ECR–R had the second largest sample size of all of the measures considered.

As shown in Table 4, the age of the sample and the percentage of the sample that was male, college students, from outside the United States, and White failed to predict a statistically significant amount of the variance in ECR–R Anxiety score reliabilities. However, these five predictors explained a statistically significant amount of the variance in ECR–R Avoidance score reliabilities. An examination of the beta weights suggests that the percentage of the sample that was White was the only predictor to emerge as statistically significant in the context of the other predictors.

Overall, the ECR–R appears to produce highly reliable scores across studies. Furthermore, the reliability of ECR–R scores was relatively unaffected by a wide variety of sample characteristics, the relation between the percentage of the sample that is White and ECR–R Avoidance reliabilities notwithstanding. This lack of statistically significant findings is particularly noteworthy given the large number of reliability coefficients contributing to these statistics and given the amount of power available to detect effects. All things considered, the ECR–R appears to produce highly reliable scores for a wide variety of samples. As such, it seems particularly well suited for cross-group comparisons and for applying to samples with atypical compositions.

### General Discussion

This meta-analysis considered a variety of commonly used measures of adult attachment security. Although it is erroneous to suggest that a measure itself is reliable, the aggregation of studies in these results best approximates the reliability of scores that might be produced by the measures and represent the best available population estimates of these reliability scores. Across the five measures, by all criteria considered, the ECR–R produced scores with the highest average reliability of all of the measures considered here. Furthermore, the reliability of the scores was relatively unaffected by characteristics of the sample and setting.

This suggests that the ECR–R is particularly well suited, not only as a general measure of adult attachment security, but for making cross-group comparisons between participants of a variety of backgrounds.

In comparison, although the ECR produced scores with strong overall reliability, the reliability of these scores appeared to be highly suspect to sample characteristics. Given that much of the recent knowledge base on adult attachment security was obtained from studies using the ECR, this could present several issues. Because reliability attenuates any correlations obtained using those scores, it could be that some statistically nonsignificant results might have been statistically significant if more reliable measures were chosen. However, the potential for this issue to have adversely impacted the existing literature is largely mitigated by the high average reliability of ECR scores. The potential issues with group comparisons are more of a concern. The fact that reliability for a set of scores is lower in one group than in another might make it appear as if group membership moderates the relation between attachment and other variables. In light of this, we strongly encourage cross-cultural researchers and others making group comparisons across diverse samples to carefully consider the degree to which reliability attenuation might be affecting their results. Because of the wide use of the ECR, and the potential for inappropriate comparisons, we encourage researchers to carefully examine their past research for this possibility.

The remaining measures of adult attachment in this study, the AAS, RAAS, and AAQ, produce scores consistent with modest reliability and are likely of value in some situations. Researchers might need to use older measures to obtain data comparable to that obtained from other studies. Given the issues we have described, considering the reliability of the data in hand when making substantive interpretations is even more important when using such measures.

Of all of the measures considered here, the ECR–R was the only one of the measures to make use of IRT in its development. Rather than assuming that all items measure the same underlying trait at the same level of difficulty, IRT allows the discriminatory ability of items to differ across different levels of the underlying trait of interest. Even if IRT is not used, factor analysis can be used to obtain similar parameter estimates (Kamata & Bauer, 2008). Although it is necessary to use IRT scoring techniques to take full advantage of this methodology, the psychometric benefits of using such methods for item selection, even when considering classical psychometric markers such as reliability, can be substantial.

The fact that a larger proportion of authors reported the reliability of their data in hand than is typical for other research areas is laudable. It is of the utmost importance that researchers calculate and consider the reliability of their own data, rather than using reliability induction to assume that their data will possess sufficient reliability. In addition to always reporting the reliability of data in hand, we encourage researchers to carefully consider the potential effect of differential reliability on cross-group comparisons. Otherwise, reliability attenuation can make it appear that sample characteristics are acting to moderate the relation between adult attachment and other variables of interest. A variety of methods for correcting effect sizes for unreliability exist (e.g., Fan, 2003; Zimmerman, 2007), and we encourage attachment researchers to consider the reliability of their own data to determine whether reliability attenuation might be adversely impacting their results. Of course, reliability attenuation is only one of the many factors influencing effects, and other issues such as power, sample representativeness, and restriction in range still warrant consideration.

The data reported here do not represent a full test of the reliabilities of attachment scores across diverse populations. Because many researchers use samples of convenience (which are often White college students) for participants, the demographic variables we examined are restricted in range. Combined with the inconsistent reporting of participant demographic information across studies, this means that many of the demographic variables end up being highly correlated with one another. In cases where few studies are available, outliers with regard to demographics can have a large effect on the correlations between reliability coefficients and demographic characteristics. Likewise, problems with multicollinearity between demographic variables can result in overly large standardized weights in a multiple regression analysis. As such, the results of this meta-analysis should be considered a detailed description of the reliability evidence currently available in the literature. Purpose-built research designed to
explicitly test the psychometric properties of these instruments across diverse populations continues to be needed.

It is important to note that although reliability is of central importance to substantive research, it is not the only psychometric characteristic that can affect results. Issues pertaining to validity and measurement invariance are likewise important and should be considered when choosing a measure. One measure might be more theoretically relevant than the others, better describe change over time, or be more applicable to the population of interest. Reliability serves as a prerequisite for validity, but all of the measures we examined here had the ability to produce scores of sufficient reliability for research purposes. As such, although more reliable measures maximize the ability to detect meaningful effects, the choice of which measure to select is not unequivocal. It can be useful to use a variety of measures, as doing so can help to ensure that research results reflect the true latent traits of interest rather than idiosyncrasies that are specific to a particular measure. The ECR–R had the highest average reliabilities of the self-report measures of adult attachment security examined here, but we encourage researchers to adopt a variety of methods for measuring attachment, including interviews and observational methods.

The fact that adult attachment has blossomed into such a well-regarded field of research demonstrates the degree to which our intimate relationships are seen as central to what it means to be human. As the field continues to develop, we suspect that further connections between adult attachment and other aspects of human experience will emerge. As such, the continued development of measures that produce psychometrically sound scores will remain essential.

REFERENCES


