Published in final edited form as:
J Pers Assess. 2014 ; 96(4): 417–425. doi:10.1080/00223891.2013.832261.

Adult Attachment Ratings (AAR): An Item Response Theory Analysis

Paul A. Pilkonis[1], **Yookyung Kim**[1], **Lan Yu**[1,2], and **Jennifer Q. Morse**[1]

[1]Department of Psychiatry, University of Pittsburgh School of Medicine

[2]Department of Medicine, University of Pittsburgh School of Medicine

Abstract

The Adult Attachment Ratings (AAR) include 3 scales for anxious, ambivalent attachment (excessive dependency, interpersonal ambivalence, and compulsive care-giving), 3 for avoidant attachment (rigid self-control, defensive separation, and emotional detachment), and 1 for secure attachment. The scales include items (ranging from 6–16 in their original form) scored by raters using a 3-point format (0 = *absent*, 1 = *present*, and 2 = *strongly present*) and summed to produce a total score. Item response theory (IRT) analyses were conducted with data from 414 participants recruited from psychiatric outpatient, medical, and community settings to identify the most informative items from each scale. The IRT results allowed us to shorten the scales to 5-item versions that are more precise and easier to rate because of their brevity. In general, the effective range of measurement for the scales was 0 to +2 *SD*s for each of the attachment constructs; that is, from average to high levels of attachment problems. Evidence for convergent and discriminant validity of the scales was investigated by comparing them with the Experiences of Close Relationships–Revised (ECR–R) scale and the Kobak Attachment Q-sort. The best consensus among self-reports on the ECR–R, informant ratings on the ECR–R, and expert judgments on the Q-sort and the AAR emerged for anxious, ambivalent attachment. Given the good psychometric characteristics of the scale for secure attachment, however, this measure alone might provide a simple alternative to more elaborate procedures for some measurement purposes. Conversion tables are provided for the 7 scales to facilitate transformation from raw scores to IRT-calibrated (theta) scores.

The relationships among adult attachment styles, personality, and psychopathology, especially personality disorders (PDs), have received considerable attention (Bakermans-Kranenburg & van IJzendoorn, 2009; Patrick, Hobson, Castle, Howard, & Maughan, 1994; Shorey & Snyder, 2006). Integrating results across studies has been challenging, however, because of differences in methods and measures for assessing attachment styles. The methods include both interviews (and the ratings derived from them), which grow out of a research tradition in developmental psychology, and questionnaires that rely on self-report, which have been used in research in social and personality psychology. As examples, the Adult Attachment Interview (AAI; Main & Goldwyn, 1985/1991) is the prototypic

Address correspondence to Paul A. Pilkonis, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213; pilkonispa@upmc.edu.

Editor's Note: This manuscript was accepted under the Editorship of Gregory Meyer.

interview, and the Experiences of Close Relationships–Revised scale (ECR–R; Fraley, Waller, & Brennan, 2000) is a commonly used questionnaire. Controversy continues, however, over the merits of these different methods, with the controversy being more conceptual than methodological (Davila & Cobb, 2003; Fortuna & Roisman, 2008; Riggs et al., 2007; Roisman et al., 2007; Shaver, Belsky, & Brennan, 2000). That is, the debate focuses largely on the construct validity and utility of self-reports versus expert judgments rather than on the psychometric properties of different instruments. In preference to self-report, Main and others have argued for the need to "surprise the unconscious" using an interview method that can be scored not only in terms of content (e.g., descriptions of greater vs. lesser interpersonal involvement), but also in terms of process (e.g., the fluency and coherence of the narrative elicited). In contrast, advocates of self-report point to the value for many purposes of "conscious appraisals and evaluations" about attachment (Roisman et al., 2007, p. 679).

One of the major goals in our own research on PDs has been to investigate the reciprocal relationships between interpersonal attachments and emotion regulation, especially in patients with borderline personality disorder (BPD). Our general hypothesis is that many of the interpersonal behaviors of persons with BPD can be understood as frustrated (and frustrating) bids for attachment as they cope with frequent episodes of emotion dysregulation. These attempts at coping result in self-defeating efforts to secure the usual provisions of attachment—a secure base in general and a safe haven in times of acute distress, reflected in proximity-seeking to attachment figures and separation distress when apart. To explore these ideas empirically, we have wanted a measure that could be used flexibly to organize information gathered across different clinical settings and research protocols and that was not constrained by the costs in training and implementation required, for example, by the AAI. At the same time, we were aware of the need to examine the convergent and discriminant relationships of any such measure with "mainstream" assessments used in the attachment literature.

With these aims in mind, we developed the Adult Attachment Ratings (AAR), which can be used with any social and developmental history that elicits relevant data about important attachments across the life span. The AAR includes seven scales reflecting traditional themes from the attachment literature: anxious, ambivalent attachment (with three scales for excessive dependency, interpersonal ambivalence, and compulsive caregiving), avoidant attachment (with three scales for rigid self-control, defensive separation, and emotional detachment), and secure attachment (with a single scale of the same name). The scales were developed from a large pool of 88 descriptive phrases characterizing the variants of attachment from the Bowlby–Ainsworth tradition (and more broadly, the literature on excessively dependent vs. overly autonomous personality styles, e.g., Blatt, 2008); that is, anxiety and preoccupation about attachment, avoidance and dismissive attitudes regarding attachment, and secure attachment displayed in a modulated blend of affiliation and autonomy. A panel of judges sorted the phrases on the basis of similarity, and these similarity ratings were subjected to a cluster analysis (Pilkonis, 1988), resulting in scales that varied in length from 6 to 16 features in their original form, with each feature rated using a 3-point format (*absent*, *present*, and *strongly present*). The AAR has proven

valuable in examining differences across different PDs (Meyer, Pilkonis, Proietti, Heape, & Egan, 2001). The scales have also been used more broadly to investigate the course of other disorders, both psychiatric (Strauss et al., 2006) and physical (Schmidt, Nachtigall, Wuethrich-Martone, & Strauss, 2002), and to examine related research questions; for example, the role of attachment style as a moderator of treatment outcome (Kirchmann et al., 2009).

Our observational, longitudinal studies of PDs now include aggregated samples of more than 400 participants, making it possible to use models from item response theory (IRT) to examine the informational value of the scales and each of their individual features. We report here on the results of these IRT analyses, with the goal of simplifying and abbreviating the seven scales (to just five items for each). Reducing the length of the scales has several advantages. It makes them easier and more efficient to use (fewer items to rate). This innovation enhances their feasibility, with the goal of promoting their adoption by other investigators. The application of IRT models ensured that the most informative items were retained (making the scales more precise); thus, the choice of items was not subjective but rather determined by the best current psychometric methods. With fewer items to rate, the expectation was also that it would be easier to establish reliability among raters using the scales.

Our aggregated sample is composed primarily of psychiatric outpatients, but it also includes community participants (with and without psychiatric diagnoses) and some medical patients. It is largely a sample of identified patients, however, and this is an important consideration for interpreting the IRT calibrations and for generalizing to new research uses and settings. That is, the calibrations should be used primarily for samples seeking mental health treatment (with the understanding that they might not be as representative in the general population). Nonetheless, the scales (now in a shorter version) provide a brief and convenient way to characterize the members of such samples. Calibration in the general population remains a future goal.

Item Response Theory

Traditional assessment tools have relied on methods derived from classical test theory (CTT), and comparing results across different measures of the same construct (e.g., attachment style) presents two major problems. First is the problem of test dependence—total scores are dependent on the particular choice of test items, regardless of whether those items accurately describe a participant's experience or provide meaningful information. The second problem is group dependence—samples that differ in ways unrelated to the construct being measured often yield different scores (Hambleton & Jones, 1993). With IRT, however, the unit of information is the individual item, not the total test score. The assumption is that respondents will have some amount of the underlying trait or ability being assessed and that their level of the trait determines the probability that they will answer an item in a specific way (Embretson & Reise, 2000). Item statistics are independent of the groups from which they are derived, an obvious benefit when attempting to compare results across studies and samples (Hambleton & Jones, 1993). These statistics reflect the ability of

each item to distinguish between higher and lower levels of the trait (discrimination) and to identify how much of the trait is required for the item to be endorsed (difficulty).

Our goal was to abbreviate each of the seven scales from the AAR by identifying the five most informative items for each scale from the 6 to 16 items used previously. A model for this work was our prior development of scales reflecting important features of PDs using items from the Inventory of Interpersonal Problems (IIP; Pilkonis, Kim, Proietti, & Barkham, 1996) and the creation of five-item versions of the IIP–PD scales (Kim & Pilkonis, 1999) based on IRT calibrations and information functions. For our five-item attachment scales, raw scores can range from 0 to 10. Because IRT (theta) scores are a nonlinear transformation of raw scores, we provide conversion tables to map raw scores onto IRT-based theta scores of "severity" for each of the seven scales.

We also examined the convergent and discriminant validity of the IRT-based AAR scales with other measures used in attachment research. Given the controversy about the relative merits of different methods for assessing attachment, we wanted one measure based on expert judgment and one based on self-report. For the former, we chose the Attachment Q-sort with summary scores for secure, preoccupied, and dismissive attachment (Kobak, 1989), and for the latter, we chose the ECR–R, which has itself been refined using IRT methods and yields two scores, one for anxious attachment and one for avoidant attachment (Fraley et al., 2000). ECR–R data were available from both participants and collateral informants who were asked to rate the attachment style of the participants. Although the constructs assessed by these different measures might not be identical, we expected convergence among the indicators for anxious attachment (the Q-sort score for preoccupied attachment, the ECR–R subscale for anxious attachment, and the AAR scales for excessive dependency, ambivalent features, and compulsive caregiving). In a similar way, we expected convergence among the indicators for avoidant attachment (the Q-sort score for dismissive attachment, the ECR–R subscale for avoidant attachment, and the AAR scales for rigid self-control, defensive separation, and emotional detachment). Finally, we expected convergence between the Q-sort score and the AAR scale for secure attachment (and negative correlations between these measures and the ECR–R subscales for both anxious and avoidant attachment). We assumed that other relationships among the measures would be smaller and provide evidence of discriminant validity. Note that a careful examination (including a meta-analysis) of the relationships between variables derived from the AAI and the Attachment Q-sort and self-reported dimensions of attachment style found such relationships to be "trivial to small," with a mean correlation of $r = .09$ (Roisman et al., 2007, p. 678). Therefore, our expectation was that the convergent relationships predicted here would be in the anticipated directions but would also be modest in magnitude.

Method

Sample

Participants ($N = 414$) were recruited from psychiatric outpatient, medical, and community settings across three different research protocols. Almost two-thirds were psychiatric patients (64%). Potential participants were excluded for psychotic disorders, organic mental disorders, mental retardation, and major medical illnesses that influence the nervous system

and might be associated with organic personality changes (e.g., Parkinson's disease, cerebrovascular disease, seizure disorders). Among the nonpsychiatric participants, we stratified on the basis of PD status using such self-report screening tools as the PD scales from the IIP (Pilkonis et al., 1996), the self-directedness subscale from the Temperament and Character Inventory (Cloninger, Przybeck, Svrakic, & Wetzel, 1994), the Iowa Personality Disorder Screen (Langbehn et al., 1999), and the McLean Screening Inventory for BPD (Zanarini et al., 2003) to sample across the spectrum of severity of personality pathology. Therefore, these participants often had current Axis I (39%) and Axis II (34%) diagnoses, making them different from general population samples with lower prevalences of disorder. (In the entire sample, 75% of participants had an Axis I diagnosis and 64% had an Axis II diagnosis.) We emphasize this point to make it clear that our IRT calibrations are based largely on samples with current psychopathology and that the version of the AAR discussed here is best applied and interpreted in this context.

The mean age of the sample was 40 years old (*SD* = 11). Sixty-nine percent of the participants were female. In terms of ethnicity, 2% were Hispanic, and in terms of race, 74% were White. Among the non-White participants, the large majority was African American (23% of the total group). Forty-five percent of the participants were single, 32% were married or living with a long-term domestic partner, and 24% were separated, divorced, or widowed. Eighteen percent had a high school education or less, and 52% were employed.

All study procedures were approved by the University of Pittsburgh Institutional Review Board. All participants agreed to participate voluntarily and provided written informed consent after receiving a complete explanation of study procedures.

Measures

Adult attachment ratings—The AAR requires rater judgments and includes seven scales organized around traditional themes from the attachment literature: anxious, ambivalent attachment (excessive dependency, interpersonal ambivalence, and compulsive care-giving), avoidant attachment (rigid self-control, defensive separation, and emotional detachment), and secure attachment (with a single scale of the same name). In their original form, the scales varied in length from 6 to 16 items, with each item requiring a 3-point rating—0 (*absent*), 1 (*present*), or 2 (*strongly present*).

Experiences of close relationships-revised scale—Based on an analysis of extant self-report measures assessing adult romantic attachment, the ECR–R yields continuous scores for two dimensions of attachment (anxiety and avoidance), each of which requires 18 items (Fraley et al., 2000). Each item is rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). It was refined using models from IRT to identify items that provide the maximum amount of information about the attachment dimensions. It can also be used to classify individuals into one of four attachment categories (secure, fearful, preoccupied, or dismissive).

Attachment Q-sort—The items in the Attachment Q-sort (Kobak, 1989) are based on Main and Goldwyn's system for identifying attachment styles. Correlations between individual Q-sorts and each of three "gold-standard" Q-sorts (reflecting secure, preoccupied,

and dismissive styles) provide three attachment scores. For this study, the Q-sort was based largely on information collected with the Interpersonal Relations Assessment (Heape, Pilkonis, Lambert, & Proietti, 1989), an interview that asks about important relationships across the life span, including childhood relationships with parents, siblings, and friends and current romantic relationships and friendships.

Data Analysis

Overview—As a first step, we inspected frequency distributions of individual scale items for sparse cells. We then began our examination of dimensionality by dividing the sample randomly into two subsamples (each with 207 participants), with an eye toward repeated confirmatory factor analyses (CFAs). The development of the scales, based on cluster analysis (Pilkonis, 1988), ensured that they were relatively homogeneous groups of items, allowing us to move directly to CFA rather than exploratory factor analysis. CFAs were conducted using M*plus* 4.21 with promax rotation (Muthén & Mutheén, 2006). In the CFAs, the items were treated as categorical variables, and the robust weighted least squares (WLSMV) estimator was used. We emphasized the magnitude of factor loadings that appeared in the CFAs and the fit and information reflected in IRT models.

Our primary goal was to document sufficient unidimensionality for IRT analyses. There are trade-offs, however, between bandwidth (a set of items that has good content validity and captures a somewhat varied pool of clinical indicators) and fidelity (a set of items that is narrower because of unidimensionality constraints). We tried to strike appropriate compromises by ensuring that the items for each scale were suitable for unidimensional scaling without unduly narrowing the construct. As Reise and colleagues have argued (Reise, Moore, & Haviland, 2010), there is a difference between studying the dimensionality of a correlation matrix versus determining the degree to which scores are influenced by a single common factor. Even multidimensional data can result in scores that still reflect essentially only one common influence, and we believe that our scales strike a reasonable compromise in this regard.

Item response theory analysis—The most commonly used IRT model for polytomous items (i.e., items with three or more ordinal response categories) is the two-parameter graded response model (GRM; Samejima, 1969). The GRM has a slope parameter and $n - 1$ threshold parameters for each item, where n is the number of response categories. The slope parameter measures item discrimination: that is, how well the item differentiates between higher versus lower levels of severity (or θ in IRT terms). Useful items have large slope parameters. Threshold parameters measure item difficulty; that is, the ease versus difficulty of endorsing different response options for an item. For example, the first threshold parameter for an item tells us where along the θ scale of severity a respondent is more likely to receive a score of 1 (*present*) rather than 0 (*absent*).

Items remaining in the pool for each scale following the CFAs were calibrated with the two-parameter GRM using MULTILOG 7.03 (Thissen, Chen, & Bock, 2003). The convergence criterion for the EM cycles was set to .0001, with the number of cycles set to 100. IRT model fit was examined for each item using the IRTFIT macro program and the option for

the sum-score-based method, which uses the sum score instead of theta for computing the predicted and observed frequencies (Orlando & Thissen, 2003).

Results

Frequency Distributions of Items

The seven scales represented an initial pool of 67 items: 16 for excessive dependency, 10 for ambivalent features, 7 for compulsive caregiving, 11 for rigid self-control, 10 for defensive separation, 6 for emotional detachment, and 7 for secure attachment. In general, there was adequate spread across the three response categories for the items, but there were four items that had fewer than five observations in the highest category (*strongly present*), one each for compulsive caregiving ("In the extreme, plays the role of the 'martyr' in order to elicit thanks and appreciation from others"), rigid self-control ("Is persistent in working toward personal goals; tends to be an overachiever"), defensive separation ("Disclaims, and even ridicules, any desire for close relations with someone who might provide love and care"), and emotional detachment ("Is relatively uninfluenced by external feedback, either praise or criticism"). A limited number of observations makes it difficult to estimate IRT parameters for sparse cells. However, we chose to include all items in the initial CFAs, given the small number of items for the caregiving and detachment scales. In any case, three of the four items with sparse cells were eliminated by our various decision rules, with only the persistence item from rigid self-control surviving into the final pool.

Confirmatory Factor Analyses

We conducted separate, single-factor confirmatory factor analyses (Step 1 CFAs) with each of the seven scales using data from the first half of the sample. Among the initial 67 items from the seven scales, 9 items were removed because of factor loadings less than .40: 3 (of 16) items from excessive dependency, 2 (of 7) items from compulsive caregiving, 1 (of 11) items from rigid self-control, 2 (of 10) items from defensive separation, and 1 (of 6) items from emotional detachment. All of the candidate items from ambivalent features ($n = 10$) and secure attachment ($n = 7$) had factor loadings larger than .40 and survived into the next round of CFAs.

To refine further the Step 1 results, we conducted another round of single-factor CFAs (Step 2 CFAs) with each of the seven scales using data from the second half of the sample. Among the 58 items that survived from the Step 1 CFAs, 7 items were removed at this stage because of factor loadings less than .40: 3 (of 13) items from excessive dependency, 2 (of 10) items from rigid self-control, and 2 (of 8) items from defensive separation.

Item Response Theory Calibrations

The resulting 51-item versions of the scales were calibrated separately using the two-parameter GRM and all data from the full sample of 414. Based on item information functions (IIFs), 7 items contributed little information (i.e., the peak of the IIF was less than 0.5): 3 (of 10) items from excessive dependency, 2 (of 10) items from ambivalent features, 1 (of 8) items from rigid self-control, and 1 (of 6) items from defensive separation. No items

were identified as misfitting ($p < .01$), and no items displayed significant residual correlations (i.e., local dependency) on any of the scales.

We selected the five best performing items for each scale based on psychometric criteria (emphasizing item information functions and discrimination parameters) and a review of their content and clinical importance. Table 1 displays the final items for each scale rank-ordered according to their a (discrimination) parameters. Test information curves for all seven scales were relatively similar, with their most effective range of measurement spanning the interval from 0 (the mean value) to +2 *SD*s for each of the attachment constructs. Thus, the scales are especially informative in characterizing respondents with average to high levels of these attachment problems; that is, from the 50th to 98th percentiles of the distribution characteristic of the calibration sample. This outcome is consistent with the goal of identifying persons at risk because of these individual differences in adult attachment style. Figure 1 displays the results for the scale for secure attachment, with the upper line providing the test information function and the lower lines providing the information functions for the five individual items. (Plots of test information functions for all the scales are available on request.)

Conversion Tables: Raw Scores to IRT Theta Scores

Table 2 is a compilation of the seven conversion tables for transforming raw scores to IRT theta scores for the short forms of the scales. The tables can be used only when all items have been rated. Because the IRT (theta) scores are a nonlinear transformation of the raw scores, the same raw scores will yield different IRT scores with different scales (see Table 2). The theta scores associated with each summed score were computed as the expected value (mean) of the sum of the likelihoods of all combinations of responses producing the given summed score. The scoring algorithm described by Thissen, Flora, Reeve, and Vevea (2000) was used to obtain the likelihoods. The posterior distributions were then computed based on the combination of the likelihoods and the prior distributions. The theta scores and associated standard errors were computed as the expected values and standard deviations of the posterior distributions.

The correlations between the raw scores from the original scales and the raw scores from the five-item versions ranged from .89 to .98; the correlations between the raw scores from the original scales and the IRT theta scores for the five-item versions, .88 to .96; and the correlations between the raw scores and the IRT theta scores for the five-item versions, .92 to .98. These results demonstrate strong convergence among the alternative forms of the scales.

Convergent and Discriminant Relationships With Other Attachment Measures

ECR–R self-report data were available from a subsample of 256 participants, ECR–R data from informants were available for 209, and the Attachment Q-sort was completed for 262. These data were collected in the latter two of the three protocols from which the entire data set was aggregated, representing an addition to the assessment battery. Participants in the latter two protocols did not differ from participants in the first protocol in terms of age, gender distribution, ethnicity, and marital status, but they did differ in terms of race,

education, and employment status. Participants in the latter two protocols displayed greater racial diversity (33% vs. 13% non-White) and were more disadvantaged socially (24% vs. 8% with a high school education or less, and 61% vs. 26% currently unemployed). It is plausible that the greater racial and social diversity of participants with ECR and Q-sort data make the results from these data more generalizable to many clinical samples.

The correlations between participant and informant reports on the ECR–R were .50 for anxious attachment and .46 for avoidant attachment, reflecting moderate-to-good agreement between these two perspectives (see Table 3). Among participants, anxious and avoidant attachment were themselves correlated .44, whereas among informants, these two forms of insecure attachment correlated .23, suggesting that informants made a greater distinction between these themes.

With ECR–R ratings, there were clear (and expected) relationships between self-reported anxious attachment and the IRT theta scores from the five-item scales for excessive dependency (.50) and interpersonal ambivalence (.48, see Table 3). There was no relationship, however, with compulsive caregiving. In addition, there were smaller and often nonsignificant relationships between self-reported avoidant attachment and the IRT scores for our avoidant scales. Note, however, that the scale for secure attachment performed better in detecting the “insecure” signal in both anxious (−.46) and avoidant (−.32) directions. A similar pattern emerged with informant data on the ECR–R. Thus, the best consensus among self-reports on the ECR–R, informant reports on the ECR–R, and interview-based judgments on the AAR emerged for anxious, preoccupied attachment, especially the themes of excessive dependency and interpersonal ambivalence.

The associations between the AAR IRT scores and the scores from the Attachment Q-sort were stronger in general than those with the ECR–R, but these results should be interpreted with some caution because of shared-method variance—that is, the judges who rated the AAR scales during a consensus diagnostic conference also completed the Q-sort as a consensus team. On the other hand, the two tasks are considerably different and the constructs being assessed are not identical. Therefore, their convergence is reassuring, documenting that the constructs derived from the two approaches are associated in predictable ways. The strongest relationships appeared between (a) the Q-sort summary score for preoccupation and the AAR IRT scores for ambivalence (.63) and excessive dependency (.51), and (b) the Q-sort summary score for dismissiveness and the AAR IRT score for emotional detachment (.43). Again, it is worth noting that the single, five-item scale for secure attachment was strongly associated with all outcomes available from the Q-sort, which requires a much lengthier and more effortful process (r = .68 with secure attachment on the Q-sort, −.54 with preoccupied attachment, and −.55 with dismissive attachment).

Discussion

The results reported here document our success in fulfilling our initial goals. We refined our scales to make them sufficiently unidimensional for IRT analyses, and on the basis of these analyses, we were able to shorten the scales to five items. These versions of the scales are

brief yet precise, making them efficient to use in any clinical research or assessment setting where constructs relevant to adult attachment are important to measure. The scales require ratings by judges who have collected appropriate social and developmental histories, but they are not constrained by implementation with a single methodology, for example, the AAI. In addition, the scales have been calibrated using a sample composed largely of participants with a current (Axis I or Axis II) diagnosis, most of whom were receiving psychiatric care. Their most effective range of measurement is from 0 (the mean of the calibration sample) to about 2 *SD*s above the mean (see Figure 1 for an example). It is important to recognize that their validity is greatest in this context, in samples in which psychopathology and attachment problems are likely to be present. In such samples, the scales provide a brief and efficient way to assess whether a "signal" is present for attachment problems and to alert the investigator (or clinician) to the need to pursue these issues further. A goal for future work is to extend the reach of the scales; that is, to calibrate them in general population and nonclinical samples.

The scale for secure attachment provides the most information (and least measurement error) of the entire set, followed by the scales for excessive dependency, emotional detachment, and interpersonal ambivalence. The shapes of the test information curves are much the same for all the scales, but the elevations (reflecting the total information contained in the scales) vary. Secure attachment, excessive dependency, emotional detachment, and interpersonal ambivalence have peaks in a range from 12 to 16, whereas compulsive caregiving, rigid self-control, and defensive separation have peaks that are lower (range = 8–10). These results help to explain the findings for convergent and discriminant validity (see Table 3) where the former scales tend to have the strongest relationships with other measures of attachment—they provide the strongest informational signal and generate correlations less likely to be attenuated by measurement error. Given the good overall psychometric characteristics of the scale for secure attachment, this measure alone might provide a simple alternative to more elaborate procedures for some measurement purposes.

It is also worth noting that the magnitude of the convergent relationships across the measures investigated here should be interpreted in the context of even more modest associations characteristic of the literature on the measurement of attachment. The absolute values of the largest correlations in Table 3 fall in the good-to-moderate range, but previous investigations (including a meta-analysis) have documented that relationships between variables derived from the AAI and the Attachment Q-sort and self-reported dimensions of attachment style are often smaller, usually less than $r = .20$ (Roisman et al., 2007). In this context, the AAR represents a reasonable compromise between ratings derived from interview methods (e.g., the Attachment Q-sort), focused on the process and coherence of attachment narratives, and self-reports about attachment relationships, which reflect overt appraisals of those relationships.

A future goal is to promote more work on the reliability of the new scales. In much of our own work, we have relied on a consensus methodology (rating the items and scales at a diagnostic conference where a group consensus is recorded), and we have not done systematic investigation of reliability across research groups. Simplifying the scales in the ways described here makes it more feasible to investigate interrater reliability in

conventional ways (both within a single research group comparing individual raters and across research groups), although we recognize that comparing consensus groups might never be as practical. In this vein, we also recognize (a) that some of our items have multiple components, making them more challenging to rate reliably if different judges focus on different components, and (b) that regardless of the number of items, some items require rather high levels of inference, another potential impediment to reliability. At a minimum, judges should be familiar with the relevant constructs from attachment theory, and systematic examination of interrater reliability should be undertaken to ensure that judges are applying this knowledge similarly.

Another future goal is to examine the predictive validity of the IRT-calibrated scales. We have used the original versions to document the importance and specificity of problems in attachment for patients with BPD (Meyer et al., 2001; Stepp, Hallquist, Morse, & Pilkonis, 2011), and we continue to investigate the reciprocal relationships between interpersonal attachments and emotion regulation in BPD (Morse et al., 2009). The Linehan (1993) model of BPD puts affective instability and lack of emotion regulation at the core of the disorder and hypothesizes that improvement in emotion regulation mediates other changes. From an interpersonal perspective, however, it is plausible that the causal sequence is the opposite (at least for some patients)—that is, changes in attachment create a more hospitable interpersonal environment that promotes emotion regulation. We propose to examine both possibilities with longitudinal data, including those collected with the AAR. Our specific hypothesis is that there will be a subgroup (latent class) of BPD patients who display changes in interpersonal functioning that precede changes in emotional reactivity and regulation and that a critical variable associated with membership in this group will be the presence of at least one “true” attachment figure.

Acknowledgments

This research was supported in part by grants MH044672 and MH056888 from the National Institute of Mental Health. We acknowledge the help of Rosary Giang, Kimberly Nolf, and Nathan Dodds in the preparation of the article.

References

Bakermans-Kranenburg MJ, van IJzendoorn MH. The first 10,000 adult attachment interviews: Distributions of adult attachment representations in clinical and non-clinical groups. Attachment & Human Development. 2009; 11:223–263. [PubMed: 19455453]

Blatt, S. Polarities of experience: Relatedness and self-definition in personality development, psychopathology, and the therapeutic process. American Psychological Association; Washington, DC: 2008.

Cloninger, CR.; Przybeck, T.; Svrakic, DM.; Wetzel, R. The Temperament and Character Inventory (TCI): A guide to its development and use. Center for Psychobiology of Personality; St. Louis, MO: 1994.

Davila J, Cobb RJ. Predicting change in self-reported and interviewer-assessed adult attachment: Tests of the individual difference and life stress models of attachment change. Personality and Social Psychology Bulletin. 2003; 29:859–870. [PubMed: 15018674]

Embretson, SE.; Reise, SP. Item response theory for psychologists. Erlbaum; Mahwah, NJ: 2000.

Fortuna K, Roisman GI. Insecurity, stress, and symptoms of psychopathology: Contrasting results from self-reports versus interviews of adult attachment. Attachment & Human Development. 2008; 10(1):11–28. [PubMed: 18351491]

Fraley RC, Waller NG, Brennan KA. An item response theory analysis of self-report measures of adult attachment. Journal of Personality & Social Psychology. 2000; 78:350–365. [PubMed: 10707340]

Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. Educational Measurement: Issues and Practice. 1993; 12(3):38–47.

Heape, CL.; Pilkonis, PA.; Lambert, J.; Proietti, JM. Interpersonal relations assessment. Department of Psychiatry, University of Pittsburgh; Pittsburgh, PA: 1989.

Kim Y, Pilkonis PA. Selecting the most informative items in the IIP scales for personality disorders: An application of item response theory. Journal of Personality Disorders. 1999; 13:157–174. [PubMed: 10372349]

Kirchmann H, Mestel R, Schreiber-Willnow K, Mattke D, Seidler K-P, Daudert E, Strauss B. Associations among attachment characteristics, patients' assessment of therapeutic factors, and treatment outcome following inpatient psychodynamic group psychotherapy. Psychotherapy Research. 2009; 19:234–248. [PubMed: 19396654]

Kobak, RR. The Attachment Interview Q-sort. University of Delaware; Newark, DE: 1989.

Langbehn DR, Pfohl BM, Reynolds S, Clark LA, Battaglia M, Bellodi L, Links P. The Iowa Personality Disorder Screen: Development and preliminary validation of a brief screening interview. Journal of Personality Disorders. 1999; 13:75–89. [PubMed: 10228929]

Linehan, MM. Cognitive-behavioral treatment of borderline personality disorder. Guilford; New York, NY: 1993.

Main, M.; Goldwyn, R. Adult attachment scoring and classification system. University of California, Berkeley; Berkeley, CA: 1991. Original work published 1985

Meyer B, Pilkonis PA, Proietti JM, Heape CL, Egan M. Attachment styles and personality disorders as predictors of symptom course. Journal of Personality Disorders. 2001; 15:371–389. [PubMed: 11723873]

Morse JQ, Hill J, Pilkonis PA, Yaggi KE, Broyden N, Stepp SD, Feske U. Anger, preoccupied attachment, and domain disorganization in borderline personality disorder. Journal of Personality Disorders. 2009; 23:240–257. [PubMed: 19538080]

Mutheén, LK.; Muthén, B. Mplus user's guide. 4th. Mutheén & Mutheén; Los Angeles, CA: 2006.

Orlando M, Thissen D. Further investigation of the performance of the S-X2: An item fit index for use with dichotomous item response theory models. Applied Psychological Measurement. 2003; 27:289–298.

Patrick M, Hobson RP, Castle D, Howard R, Maughan B. Personality disorder and the mental representation of early social experience. Development and Psychopathology. 1994; 6:375–388.

Pilkonis PA. Personality prototypes among depressives: Themes of dependency and autonomy. Journal of Personality Disorders. 1988; 2:144–152.

Pilkonis PA, Kim Y, Proietti JM, Barkham M. Scales for personality disorders developed from the Inventory of Interpersonal Problems. Journal of Personality Disorders. 1996; 10:355–369.

Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment. 2010; 92:544–559. [PubMed: 20954056]

Riggs SA, Paulson A, Tunnell E, Sahl G, Atkison H, Ross CA. Attachment, personality, and psychopathology among adult inpatients: Self-reported romantic attachment style versus Adult Attachment Interview states of mind. Development and Psychopathology. 2007; 19:263–291. [PubMed: 17241494]

Roisman GI, Holland A, Fortuna K, Fraley RC, Clausell E, Clarke A. The Adult Attachment Interview and self-reports of attachment style: An empirical rapprochement. Journal of Personality and Social Psychology. 2007; 92:678–697. [PubMed: 17469952]

Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika:Monograph Supplement 17. 1969

Schmidt S, Nachtigall C, Wuethrich-Martone O, Strauss B. Attachment and coping with chronic disease. Journal of Psychosomatic Research. 2002; 53:763–773. [PubMed: 12217450]

Shaver PR, Belsky J, Brennan KA. The adult attachment interview and self-reports of romantic attachment: Associations across domains and methods. Personal Relationships. 2000; 7(1):25–43.

Shorey HS, Snyder CR. The role of adult attachment styles in psychopathology and psychotherapy outcomes. Review of General Psychology. 2006; 10(1):1–20.

Stepp SD, Hallquist MN, Morse JQ, Pilkonis PA. Multimethod investigation of interpersonal functioning in borderline personality disorder. Personality Disorders: Theory, Research, and Treatment. 2011; 2:175–192.

Strauss B, Kirchmann H, Eckert J, Lobo-Drost A, Marquet A, Papenhausen R, Hoger D. Attachment characteristics and treatment outcome following inpatient psychotherapy: Results of a multisite study. Psychotherapy Research. 2006; 16:579–594.

Thissen, D.; Chen, WH.; Bock, RD. Multilog (version 7. Scientific Software International; Lincolnwood, IL: 2003.

Thissen, D.; Flora, D.; Reeve, B.; Vevea, JL. Linear approximations for item response theory response pattern scores; Paper presented at the annual meeting of the Psychometric Society; Vancouver, BC, Canada. Jul. 2000

Zanarini MC, Vujanovic AA, Parachini EA, Boulanger JL, Frankenburg FR, Hennen J. A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). Journal of Personality Disorders. 2003; 17:568–573. [PubMed: 14744082]

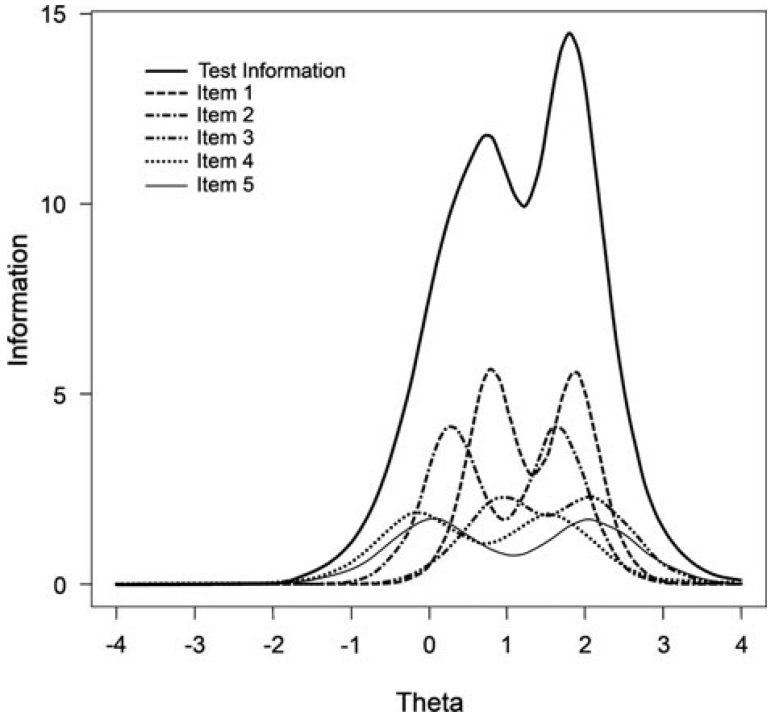

Figure 1.

Test information and item information curves for the Secure Attachment scale. Numbers for the individual items correspond to those in Table 1.

Table 1

Five-item versions of the Adult Attachment Ratings.

		Location Thresholds	
Item Stem	**Slope**	**0 vs. 1**	**1 vs. 2**
Excessive dependency			
1. Tends to depend too much on other people; becomes “clingy” in relationships	5.35	0.45	1.27
2. Has a great fear of rejection and “abandonment”; is sensitive to real (and perceived) deprivation of care, affection, and love and to disruptions in interpersonal relationships	3.03	−0.05	0.94
3. Has many passive-receptive wishes: wants to be loved, supported, understood, and guided by others	2.80	−0.33	0.94
4. Tends to be anxious and insecure because of the fear that he or she may lose an important relationship or person	1.94	0.24	1.58
5. Tends to give up control to others; underestimates his/her own abilities and resources for coping	1.58	0.31	1.64
Ambivalent features			
1. Feelings tend to be intense and can alternate rapidly between positive and negative emotions	3.55	0.17	1.06
2. Experiences anger (and even rage) over real (and perceived) deprivation	3.10	0.08	1.06
3. Tends to be manipulative in relationships	2.69	0.76	1.58
4. Has little capacity for delay of gratification and control of affect; experiences strong desires to be comforted and soothed in a direct and immediate way	2.49	−0.05	0.82
5. Interpersonal relationships are ambivalent, with a relative inability to resolve and integrate contradictory feelings (i.e., warm, loving feelings and hostile, angry feelings) toward the same person	2.27	−0.32	0.67
Compulsive caregiving			
1. Has close relationships, but always takes the role of giving care and not that of receiving it	5.08	0.94	1.79
2. Feels that it is easier to give help than to receive it	2.05	0.58	1.76
3. Insists on providing help to others, even when they do not need or want it; is confused or hurt when others spurn his/her help	1.38	1.89	3.49
4. Tends to pick partners or friends who are “lame ducks” (e.g., befriends people who are needy and vulnerable, either physically or emotionally)	1.38	1.50	2.71
5. Occasionally becomes resentful about how much he or she is doing for others and how little he or she is receiving in return	1.16	1.07	3.07
Rigid self-control			
1. Thinking is analytical, critical, and precise; attention is focused on details, differences, and contradictions	3.40	0.51	1.44
2. Intellectual processes (as an exaggerated means of control) are overvalued	3.14	1.02	1.67
3. Is very concerned with work and productivity; often assumes that others will make allowances for his/her work in preference to other activities	1.98	1.24	2.11
4. Tends to be a perfectionist	1.95	0.72	2.12
5. Is persistent in working toward personal goals; tends to be an overachiever	1.86	1.77	3.68
Defensive separation			
1. Sense of well-being is dependent on preserving autonomy, directing his/her own activities, and being free from interference of others; values the achievement of a sense of separation, definition, and independence	3.46	0.06	1.11
2. Exaggerated struggles to establish self-control and separateness interfere with the establishment of satisfying interpersonal relationships	2.41	1.20	2.19
3. Maintains strong personal boundaries; great stress is placed on defining him/herself as separate and different from others	2.15	0.63	1.73
4. Prefers to keep his/her options open rather than making permanent commitments; works to preserve and increase independence, personal choice, and mobility	1.53	0.64	1.95

		Location Thresholds	
Item Stem	**Slope**	**0 vs. 1**	**1 vs. 2**
5. Insists on being self-sufficient, regardless of circumstances; tends to depend too little on other people and dislikes asking for help	1.22	0.77	2.57
Emotional detachment			
1. Tends to be antisocial and delinquent; tends not to display remorse or guilt when such reactions would be appropriate	4.73	0.73	1.61
2. Focuses on getting positive results and places relatively little weight on the possible negative consequences of his/her actions	3.83	0.60	1.67
3. Is somewhat oblivious to the effects of his/her actions on other people; is rather insensitive to other people's needs and wishes	2.25	−0.13	1.04
4. Is less concerned than most people about physical illness and death	2.03	1.91	2.88
5. Dislikes being held back, blocked, or deterred from doing what s/he wants to do; dislikes externally imposed directives, deadlines, demands, or pressures	1.80	0.07	1.17
Secure attachment			
1. Is relatively unconflicted about relationships; does not experience excessive dysphoria (e.g., guilt, ambivalence, separation anxiety) in this part of life	4.73	0.80	1.87
2. Has relationships that are meaningful and satisfying with no perceived risk to the autonomy or integrity of the participants; such interactions are felt to be mutually enhancing rather than draining or depleting	4.07	0.28	1.65
3. Has a good sense of his/her own identity, but also appreciates the personalities of others and finds pleasure in relating to them	2.97	0.92	2.12
4. Tends to portray relationships in optimistic terms and associates them with genuine gratification	2.69	−0.12	1.57
5. Is comfortable both being with people and being apart from people	2.57	0.03	2.09

Table 2

Conversion tables: Raw scores to item response theory theta scores.

	Excessive Dependency		Ambivalent Features		Compulsive Caregiving		Rigid Self-Control		Defensive Separation		Emotional Detachment		Secure Attachment	
Raw Score	**Theta**	***SE***	**Theta**	***SE***	**Theta**	***SE***	**Theta**	***SE***	**Theta**	***SE***	**Theta**	***SE***	**Theta**	***SE***
0	−1.15	0.61	−1.13	0.61	−0.67	0.74	−0.68	0.71	−0.92	0.67	−0.95	0.67	−0.99	0.63
1	−0.52	0.45	−0.50	0.43	0.03	0.61	0.14	0.50	−0.23	0.54	−0.27	0.50	−0.27	0.42
2	−0.19	0.40	−0.18	0.38	0.44	0.58	0.53	0.44	0.16	0.50	0.10	0.46	0.15	0.35
3	0.10	0.36	0.09	0.34	0.85	0.50	0.84	0.38	0.49	0.44	0.44	0.40	0.48	0.31
4	0.34	0.34	0.32	0.32	1.16	0.46	1.09	0.36	0.77	0.42	0.75	0.35	0.80	0.30
5	0.58	0.33	0.53	0.31	1.44	0.44	1.33	0.36	1.04	0.42	1.01	0.33	1.08	0.30
6	0.80	0.32	0.74	0.31	1.72	0.45	1.56	0.36	1.30	0.42	1.24	0.32	1.33	0.29
7	1.04	0.34	0.96	0.32	1.99	0.46	1.81	0.38	1.56	0.42	1.52	0.34	1.60	0.29
8	1.31	0.38	1.20	0.34	2.25	0.47	2.08	0.40	1.83	0.44	1.84	0.38	1.87	0.29
9	1.63	0.42	1.48	0.37	2.51	0.50	2.41	0.46	2.13	0.48	2.13	0.41	2.15	0.32
10	2.03	0.51	1.95	0.50	2.82	0.55	2.80	0.54	2.51	0.53	2.47	0.50	2.55	0.42

Table 3

Correlations among all attachment variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Attachment adult ratings														
1. Excessive dependency														
2. Ambivalent features	**0.48**													
3. Compulsive caregiving	−0.04	−0.20												
4. Rigid self-control	−0.18	− **0.32**	0.12											
5. Defensive separation	− **0.43**	−0.04	−0.10	0.21										
6. Emotional detachment	0.02	**0.57**	− **0.31**	−0.23	**0.32**									
7. Secure attachment	− **0.36**	− **0.56**	0.22	0.29	−0.13	− **0.44**								
ECR–R (self-report)														
8. Anxious attachment	**0.50**	**0.48**	−0.01	−0.17	−0.11	0.20	− **0.46**							
9. Avoidant attachment	0.14	0.23	0.10	−0.09	0.12	0.08	− **0.32**	**0.44**						
ECR–R (informant)														
10. Anxious attachment	**0.45**	**0.43**	−0.08	−0.18	−0.10	0.17	− **0.52**	**0.50**	0.21					
11. Avoidant attachment	−0.07	−0.01	0.09	0.10	0.21	−0.04	− **0.33**	0.18	**0.46**	0.23				
Attachment Q-sort														
12. Preoccupied	**0.51**	**0.63**	−0.04	−0.10	−0.11	0.24	− **0.54**	**0.42**	0.29	**0.48**	0.14			
13. Dismissive	−0.05	0.22	−0.11	−0.14	0.25	**0.43**	− **0.55**	0.16	0.25	0.22	0.20	0.29		
14. Secure	−0.21	− **0.47**	0.13	0.15	−0.12	**−0.45**	**0.68**	**−0.33**	**−0.32**	**−0.38**	−0.21	**−0.68**	**−0.89**	

Note. ECR–R = Experiences of Close Relationships–Revised scale. *N* = 414 for the attachment adult ratings, 256 for the ECR–R self-reports, 209 for the ECR–R informant reports, and 262 for the Attachment Q-sort. Correlations with absolute values greater than .30 are shown in bold.