

## Emotional Quotient Inventory Standardization, Reliability, and Validity

### Overview

This page describes the psychometric properties of the EQ-i® 2.0 and EQ 360® 2.0, including standardization, reliability, and validity. The pilot study and standardization studies are described, including a description of the normative data collection, creation of norm groups, and demographic analyses. The reliability sections describe the internal consistency and test-retest reliability of the instruments. The validity sections describe the ability of the EQ-i 2.0 and EQ 360 2.0 to show expected associations with other psychological instruments and expected group differences, which would support the notion that the EQ-i 2.0 and EQ 360 2.0 are valid measures of emotional intelligence. The first few sections on this page are devoted to the standardization, reliability, and validity of the EQ-i 2.0, followed by sections describing these same properties for the EQ 360 2.0. To begin, a brief explanation of effect size, which is instrumental in interpreting these results, will be provided.

All tables and figures representing detailed depictions of these properties are available in [Appendix A](#) (Standardization, Reliability, and Validity).

### Effect Size

When analyzing data from an extremely large sample (such as the ones described on this page), the proper interpretation of what constitutes a significant result is important. There will be several instances throughout this page where tests of significance (e.g., *F*-tests) will be reported. As Thompson (2002) noted, significance tests do not inform as to the importance, or practical significance of the test result. Significance tests are greatly influenced by sample size; that is, the larger the sample, the more likely a test will be statistically significant (Thompson, 2002). With a normative sample size of 4,000 in the EQ-i 2.0 and 3,200 for the EQ 360 2.0, it is therefore necessary to examine the practical significance of all analyses, in addition to the statistical significance.

In order to accomplish this, estimates of effect size (e.g., Cohen's *d*) that estimate the strength of the effect are provided for analyses where appropriate. Effect sizes permit the comparison of results across studies, in which sample sizes may differ dramatically. For example, Cohen's *d* illustrates the difference between two means in terms of pooled standard deviations (i.e., a value of 1.00 means that the mean scores from the two groups differ by one pooled standard deviation). Standard criteria, which are not influenced by sample size (Cohen, 1988), are available for determining small, medium, and large effect sizes. For instance, marker values for interpreting small, medium, and large effects with Cohen's *d* are .20, .50, and .80, respectively.

Correlations are also commonly reported on this page. Although the interpretation of correlation coefficients varies depending on how you are using them, for the data reported on this page, ranges for interpreting small, medium, and large effects with the correlation coefficient (*r*) are .10, .30, and .50 (absolute values), respectively.

Partial eta-squared ( $\eta^2$ ) is used to summarize differences between multiple categorical groups or

to summarize non-linear differences between groups (e.g., age groups). This statistic is preferable to  $d$  in analyses where differences between more than two groups are examined (e.g., racial/ethnic groups), or where a non-linear effect is expected, such as the EI age trends, where scores increase up to a point and then decrease over the life span. Partial  $\eta^2$  is also used to quantify interaction effects between multiple variables (e.g., between age groups and gender). Cutoffs for evaluating partial  $\eta^2$  as small, medium, and large are .01, .06, and .14, respectively (Cohen, 1988).

## EQ-i 2.0 Pilot Study and Standardization

Standardization is an important part of test development, involving the collection of pilot and normative data. Pilot data is used to test the basic functions of an assessment, such as its reading level, response instructions, and completion time. Issues that may arise in these areas can then be addressed before normative data collection begins. Normative data establish a baseline against which all subsequent results are compared, and enable the test developer to capture the characteristics of an “average” respondent. Norms indicate the average performance on a test and the distribution of scores above and below the average (Anastasi, 1988). A large, representative normative database ensures that the reference group is inclusive with respect to demographic variables such as age, gender, education level, and race/ethnicity, increasing the audience to which the assessment is relevant. This section describes the method of data collection and the breakdown of the pilot and normative samples, including the effects of age and gender on the EQ-i 2.0 results. Data collection for the EQ-i 2.0 followed multiple stages between June 2009 and December 2010. More than 10,000 participants completed the EQ-i 2.0 over this time period. These data were collected for pilot testing, the creation of norms, and validation analyses.

### Data Collection

#### Pilot Phase

This first stage of data collection took place between June 2009 and November 2009. Participants in the EQ-i 2.0 pilot dataset ( $N = 1,346$ ; [Table A.1](#)) were 58.8% female with a mean age of 35.5 years ( $SD = 14.6$  years). The majority of the sample was White (74.9%) and the largest education level group was college/university degree or higher (45.7%). These data were collected to ensure that the basic functionality of the EQ-i 2.0 (e.g., instructions, response options, administration time) was adequate. The pilot study confirmed most aspects of the test protocol and where needed, some adjustments were made to fine tune the assessment.

#### Normative Phase

This second phase that involved the collection of data that were included in the normative sample, as well as reliability and validity data, took place between March 2010 and December 2010.

Data were gathered from all 50 U.S. states and the District of Columbia, as well as from all 10 Canadian provinces. All raters were sent an email invitation to participate in the EQ-i 2.0 data collection process. Those who agreed to participate completed the assessment online, and were compensated for their time. Various measures were undertaken to ensure all data met the highest levels of data authenticity. For example, data were screened so that any potentially illegitimate

assessments (e.g., participant responded with only a single response option for a significant number of items in a row, left too many items missing, took less than 10 minutes or more than 90 minutes to complete the assessment, etc.) were excluded from the dataset. The following section focuses on a description of the normative sample; see the [EQ-i 2.0 Reliability](#) and [EQ-i 2.0 Validity](#) sections further down this page for more information on the reliability and validity samples.

In order to create representative normative samples, specific demographic (i.e., age, gender, race/ethnicity, education level and geographic targets), guided by recent Canadian and U.S. Census information (i.e., Statistics Canada, 2006; U.S. Bureau of the Census, 2008), were utilized during the data collection procedure. In order to create the representative samples for the EQ-i 2.0, information was collected on each participant's gender, age, race/ethnicity (Asian/Pacific Islander, Black/African-American/African-Canadian, Hispanic/Latino, White, Multiracial, or Other), highest level of education attainment (high school or less, some college/university, college/university degree or higher), employment status (employed/self-employed, unemployed, retired, or other), and geographic location (state/province and country). For ease of presentation, race/ethnicity groups are referred to in this manual as follows: Black, Hispanic/Latino, White, and Other.

### **Standardization**

This section describes the process of standardization for the EQ-i 2.0, including a description of the normative sample, and the statistical analyses that were conducted in order to create normative groups and standardized scores.

### **Normative Sample**

Normative data were collected between March 2010 and April 2010. During this time period, 4,996 participants provided EQ-i 2.0 data for standardization purposes. A final sample of 4,000 participants was selected as the normative dataset. Statistical analyses showed no meaningful differences between U.S. and Canadian participants in EQ-i 2.0 scores (i.e., none of the Cohen's *d* values even reached a small effect size; [Table A.2](#)), so data from both countries were included together in a single normative sample. The EQ-i 2.0 normative sample was collected within ten age ranges (400 cases in each age range), equally proportioned by gender ([Table A.3](#)). The data provided in [Tables A.4 through A.6](#) indicate that the normative sample is very similar to the Censuses (within 3%) in terms of race/ethnicity, geographic region, and education level. Therefore, the reference group against which individual EQ-i 2.0 scores are compared is representative of the North American general population.

### **Norming Procedures**

The first step in preparation of the norms was to determine if any trends existed in the data. For instance, large differences in scores between men and women, or across various age groups, could provide an argument for creating separate gender- or age-based norm groups. Conversely, a lack of such differences may dictate the use of a single norm group with genders and age groups combined. A series of analyses of covariance (ANCOVA; for Total EI) and multivariate analyses of covariance (MANCOVA; for the Composites and Subscales) were used to examine the relationships between gender and age with EQ-i 2.0 scores. For ease of interpretability, the ten age groups were condensed into five (18–29 years, 30–39 years, 40–49 years, 50–59 years, and 60+ years) for these analyses, with education level and race/ethnicity as covariates (in order

to control for the effects of these demographic variables). In an attempt to control for Type I errors that might occur with multiple analyses, a more conservative criterion of  $p < .01$  was used for all  $F$ -tests.

The Wilk's lambda statistic generated from these analyses ranges from 0.00 to 1.00 and conveys the proportion of variance that is not explained by the effect (in this case, the interaction between gender and age) in the multivariate analyses. These values were all close to 1.00, suggesting that only a small amount of variance could be explained by the interaction. However,  $F$ -tests revealed significant effects of gender, age, and the interaction of gender and age (see [Table A.7](#)). Given these results, the univariate effects are described in detail below.

**Focus on Effect Size.** The large sample size dictates that effect sizes should be considered more strongly than significance tests (see the previous section on Effect Size). The effect sizes are provided in [Table A.8](#). While Cohen's  $d$  values are reported to describe the size of the gender effects, Cohen's  $d$  values are not appropriate for describing age effects (where there are more than two groups). Furthermore, previous research has determined that associations between age and EI are generally non-linear, with scores increasing up to a certain age (around age 40–50) then either decreasing slightly or stabilizing (Bar-On, 1997). Therefore, it is inappropriate to examine correlations between age and EI, because Pearson's correlations are used to estimate linear trends and can therefore underestimate or completely overlook non-linear relationships. Instead, partial eta-squared (partial  $\eta^2$ ) values are reported and are used to summarize the overall effect of age on EI (technically speaking, it quantifies the proportion of variance in EI scores accounted for by the age groups).

**Gender Effects.** Results of the gender analyses showed that males and females did not differ significantly on the EQ-i 2.0 Total EI score, indicating that overall emotional intelligence as measured by the EQ-i 2.0 is the same for males and females; however, small to medium gender effects were found for some subscales (see [Table A.8](#) for effects sizes and [Table A.9](#) for descriptive statistics and significance test results). The largest difference was on Empathy, with women scoring higher than men with a moderate effect size ( $d = -0.49$ ). Smaller differences were found with women scoring higher than men on the Interpersonal Composite ( $d = -0.33$ ), Emotional Expression ( $d = -0.31$ ), and Emotional Self-Awareness ( $d = -0.22$ ). Men scored higher than women with small effect sizes on Stress Tolerance ( $d = 0.30$ ), Problem Solving ( $d = 0.26$ ), and Independence ( $d = 0.21$ ). These differences are compatible with the logic of the EQ-i 2.0 conceptual framework and show empirical precedence, such as in the original EQ-i (see Bar-On, 2004). However, it is important to note that these effects were small and represent only a few absolute standard score points.

**Age Effects.** Significant but small age effects were found for the EQ-i 2.0 (see [Table A.8](#) for effect sizes and [Table A.10](#) for descriptive statistics and significance test results). The age differences varied from scale to scale. In some instances, scale scores increased with age (i.e., Total EI, Self-Regard, Interpersonal Composite, Interpersonal Relationships, Empathy, Stress Management Composite). In other cases, scores increased until about age 40–49 years, then the scores stabilized or decreased slightly (i.e., Self-Expression Composite, Independence, Problem Solving, Flexibility, Stress Tolerance). Differences between age groups were generally only a few standard score points in magnitude. Previous research has demonstrated similar age trends (see Bar-On, 2004). Emotional Self-Awareness and Assertiveness were the only subscales that failed to show at least a small effect size.

**Gender × Age Interaction.** There were no interactions between age and gender; partial  $\eta^2$  values did not reach the minimum criterion for a small effect size ([Table A.8](#)). In fact, partial  $\eta^2$  values were .00 for all scales. In other words, any age effects were consistent within males and females, and any gender effects were consistent within age groups.

Overall, the age and gender analyses revealed significant, but small effects. Therefore, both specific “Age and Gender norms” (i.e., age and gender specific) as well as “General population norms” (i.e., neither age nor gender specific) were developed. Actual construction of the norms was conducted by a multi-step statistical process. Results revealed that skewness and kurtosis values were close to 0 (skewness values ranged from -0.93 to -0.15; kurtosis values ranged from -0.17 to 0.77), and an examination of the scale histograms did not reveal any significant departures from normality (an example histogram for the EQ-i 2.0 Total EI score is provided in [Figure A.1](#)). Therefore, artificial transformation of scores to fit normal distributions was deemed unnecessary.

In the next step, means were statistically smoothed for the Age and Gender norms. Data points that diverged significantly from a smooth curve partly reflect true differences and partly reflect sampling variability (Zachary & Gorsuch, 1985). To mitigate the effect of sampling variability, the data were smoothed using the following technique. Means and standard deviations were computed at each age group, separately for males and females, for every score. For each scale, regression analysis was used to find the best fitting curve (linear or curvilinear) across age. Linear and quadratic effects of age were the independent variables, and the mean scores at each age were the dependent variables. At each age, the predicted score mean from the regression was used in conjunction with the original (unsmoothed) mean to produce the final norms. Specifically, the final “smoothed” mean was a weighted mean of the regression generated value, and the original, unsmoothed mean (each a 50% weighting). Use of this smoothed normative value allows for irregular but real differences between age groups to have an effect, while reducing the impact of random fluctuation. The smoothed values were averaged within each of the five age groups for the computation of the standard scores. For example, the mean of the means and standard deviations for 18-year-olds, 19-year-olds, 20-year-olds, and so on up to 29-year-olds were computed for the 18-29 years group.

### **Standardization Summary**

Over 10,000 EQ-i 2.0 assessments were collected between 2009 and 2010 for the standardization of the tool. A sample of 4,000 participants was chosen as the EQ-i 2.0 normative sample. The sample was evenly distributed by gender and age, and matched to the Census based on race/ethnicity, geographic region, and highest level of educational attainment. Statistical analyses revealed small differences across gender and age; therefore, general norms as well as separate age and gender norms are available as options in the use of the EQ-i 2.0. The norming process resulted in standard scores with means of 100 and standard deviations of 15 for the Total EI score, Composite Scales, and Subscales.

### **EQ-i 2.0 Reliability**

Reliability is defined as “the consistency of scores obtained by the same person when re-examined with the same test on different occasions, or with difference sets of equivalent items,

or under other variable examining conditions” (Anastasi, 1988, p. 102). Two basic statistical methods for evaluating a test’s reliability are internal consistency and test-retest reliability analyses. Internal consistency refers to the general cohesiveness of its items, or the degree to which a particular set of items assess a single construct. Test-retest reliability refers to the stability of scores over time. Each of these two analyses was conducted for the EQ-i 2.0. From a practical perspective, internal consistency may be used to calculate the precision or “margin of error” associated with an individual’s EQ-i 2.0 score. These values are also referred to as *confidence intervals (CI)*. Reliability analyses are also used to determine which subscales are in balance or out of balance with one another within a client’s EQ-i 2.0 profile (i.e., what is a meaningful difference between an individual’s subscale scores?).

### **Internal Consistency**

Internal consistency conveys the degree to which a set of items are associated with one another. High levels of internal consistency suggest that the set of items are measuring a single, cohesive construct. Internal consistency is typically measured using Cronbach’s alpha (Cronbach, 1951). Cronbach’s alpha ranges from 0.0 to 1.0 and is a function of (a) the interrelatedness of the items in a test or scale and (b) the length of the test (John & Benet-Martinez, 2000). Higher values reflect higher internal consistency.

Cronbach’s alpha values for the EQ-i 2.0 scales in the normative sample are presented in [Table A.11](#) (see the the previous Standardization section for a description of the normative sample). Given that Cronbach’s alpha is influenced by the number of items in a set (with more items generally leading to higher alphas), the number of items per scale is also displayed in this table. Though there is no universal criterion for a “good” alpha level, informal cutoffs for evaluating alpha are typically .90 is “excellent,” .80 is “good,” .70 is “acceptable,” and less than .70 is “unacceptable.” Most of the values found in [Table A.11](#) demonstrate excellent reliability for the EQ-i 2.0, particularly notable given the small number of items included in most subscales. Looking at the General (Total Sample) column, the alpha value of the Total EI scale was .97, values for the composite scales ranged from .88 to .93, and values were .77 or higher for all subscales. These values were similar within the various age and gender normative groups, including a Total EI alpha of at least .97 in each norm group. Furthermore, these values are generally higher than those found in the original EQ-i normative samples. For instance, the average alpha reliability value for the original EQ-i Total EI score across nine normative samples was .79 (Bar-On, 2004). The high level of internal consistency found in the EQ-i 2.0 Total EI score supports the idea that, taken together, the EQ-i 2.0 items are measuring a single cohesive construct—namely, emotional intelligence.

### **Confidence Intervals**

A practical application of alpha values is that they may be used to calculate the precision or margin of error associated with individual scores. Specifically, alpha values may be used to calculate confidence intervals for each individual score. Unlike physical attributes, such as height and blood pressure, psychological characteristics (such as EI) cannot be measured directly. Psychological assessments serve as estimates of an individual’s true score on these dimensions, and therefore some degree of uncertainty is associated with the obtained scores. Confidence intervals are a method of measuring the degree of this uncertainty. The relationship between alpha values and confidence intervals is inverse; as alpha values increase, confidence intervals decrease. In other words, as the internal consistency of an assessment increases, the degree of

uncertainty decreases. □ Confidence intervals at the 90% confidence level for all EQ-i 2.0 scores are integrated into the computerized reports as an option the user may select. For example, if a client obtains a score of 105 on the EQ-i 2.0 Total EI scale, 90% confidence intervals suggest that the margin of error is  $\pm 4$  points, with the true score ranging from a low of 101 to a high of 109. In other words, this individual's actual level of EI will fall within this interval 90% of the time. Note that the score of 105 still remains the best single point estimate of the client's Total EI.

### **Balancing Ei: Comparing Differences In Subscale Scores**

The EQ-i 2.0 report includes an optional Balancing Your EI section. This section compares scores from every subscale to three related subscales. For example, Self-Regard is compared to Self-Actualization, Problem Solving, and Reality Testing (see [Understanding the Results](#) for details on interpreting the Balancing Your EI section). Analyses similar to those used to generate confidence intervals were used to calculate the size of “gaps” between EQ-i 2.0 subscales. Results from these analyses were used to guide the critical value at which point scales were determined to be “in balance” or “out of balance” with each other. Specifically, considering the results of these analyses as well as practical functionality, a critical value of 10 points was selected for the Balancing Your EI section. This value is actually slightly smaller than those suggested by the statistical analyses, but was selected so the user can be confident that they are identifying any potentially important imbalances in EI abilities. For example, if two subscales in the Balancing Your EI section are less than 10 points apart, they will be reported as being “in balance,” whereas subscale scores that are 10 or more points apart will be described as being “out of balance.”

### **Test-Retest Reliability and Stability**

The test-retest reliability of an assessment refers to the consistency of scores over time. This type of reliability is typically calculated by examining the correlation between an individual's scores on the same assessment at two different times. This time interval must not be too long (Anastasi, 1982) to ensure that factors such as developmental changes do not overly obscure the assessment of the instrument's reliability, and must not be too short as to be contaminated by memory effects (Downie & Heath, 1970). A two- to eight-week interval between administrations is usually recommended.

When test-retest reliability is assessed at the group level, high correlations indicate that the rank-order of individuals' assessment scores have remained consistent over time. However, differences in mean scores may confound these results. For example, if each individual's score increases or decreases in a dramatic but uniform manner over time, the test-retest correlation for the overall sample will remain high. Test-retest stability analyses can be used to determine not only if the rank-order of scores remains consistent, but if the actual scores themselves remain stable over time. Test-retest stability was examined by calculating the difference between Time 1 and Time 2 standard scores for each individual in the test-retest samples.

For the EQ-i 2.0, test-retest data was available for 204 individuals who were assessed two to four weeks apart (mean interval = 18.41 days,  $SD = 3.22$  days), and for 104 individuals who were assessed approximately eight weeks apart (mean interval = 56.80 days,  $SD = 1.25$  days). Demographic characteristics of the two retest samples are displayed in [Table A.12](#). EQ-i 2.0 test-retest correlations are expected to be high for the two- to four-week interval, supporting the

reliability of the EQ-i 2.0 as a tool, because a person's EI should not change much over two to four weeks, especially in the absence of any EI-targeted intervention, as was the case in our data (see Stein & Book, 2000). However, in general, test-retest correlations also tend to decrease as the time interval between assessments increases because there is more opportunity for developmental changes or other events to occur. Therefore, the 8-week test-retest values are expected to be slightly lower than the 2- to 4-week values. Nonetheless, test-retest correlations (see [Table A.13](#)) were high for the EQ-i 2.0 Total EI score in both the 2- to 4-week ( $r = .92$ ) and 8-week samples ( $r = .81$ ). Test-retest correlations for the various Composite scales were very high, ranging from  $r = .86$  (Self-Expression Composite) to  $r = .91$  (Interpersonal Composite) in the 2- to 4-week sample, and from  $r = .76$  (Interpersonal Composite) to  $r = .83$  (Decision Making Composite) in the 8-week sample. Finally, results for the subscales were also high, ranging from  $r = .78$  (Impulse Control) to  $r = .89$  (Empathy) in the 2-4-week sample and from  $r = .70$  (Flexibility) to  $r = .84$  (Self-Regard, Happiness) in the 8-week sample. These values were generally similar to those found in the original EQ-i (Bar-On, 2004).

The stability of the EQ-i 2.0 scores was examined by calculating the difference between Time 1 and Time 2 standard scores for each individual in the test-retest samples. [Tables A.14](#) (2- to 4-weeks) and [A.15](#) (8 weeks) display the frequencies of these differences, as well as the mean differences (i.e., the difference between Time 1 and Time 2 ratings for each individual averaged across the samples) and the 95% confidence interval surrounding the mean difference. Positive mean differences indicate that scores increased over time, whereas negative mean differences indicate that scores decreased over time. The results suggest scores remained highly stable over time: for almost all scales, roughly 90% or more of the individuals' scores did not change by more than one normative standard deviation (i.e., 15 standard score points) over time in both the 2- to 4-week and 8-week samples. Confidence intervals around the mean differences were also consistently small, and instances where this interval encapsulates zero suggest that the difference is not statistically significant ( $p < .05$ ). These results provide support that the EQ-i 2.0 captures the temporal stability of emotional intelligence.

### **Reliability Summary**

Overall, the EQ-i 2.0 demonstrates sound reliability. Internal consistency (alpha) values were generally high for the overall normative groups and within specific age and gender subgroups, suggesting that the items cohesively measure Total EI, as well as the constructs represented by the composite scales and subscales. Test-retest reliability and stability values were also high at both 2- to 4-week and 8-week intervals, reflecting a level of temporal stability that would be expected for emotional intelligence. Users of the EQ-i 2.0 can be confident that the scores generated by this assessment will be consistent and reliable.

### **EQ-i 2.0 Validity**

Reliability is necessary for, but does not ensure, validity. The validity of a test refers to whether the test measures what it claims to be measuring; in this case, does the EQ-i 2.0 measure emotional intelligence? The quality of inferences that can be made by the test's scores, and the validity of an instrument like the EQ-i 2.0, rests upon the weight of accumulated evidence from a number of validity studies using various methodologies (Campbell & Fiske, 1959). Various types of validity were examined for the EQ-i 2.0. Specifically, how well does the EQ-i 2.0 measure the construct(s) it was designed to measure, how well are the claims regarding its use and

applications supported by empirical evidence, and is the EQ-i 2.0 free of test-bias?

Evidence that the EQ-i 2.0 measures the constructs it was designed to measure include

- a description of the content validity of the assessment;
- the appropriateness of the scale structure (including the Positive and Negative Impression scales and Inconsistency Index); and
- an exploration of the relationship of the EQ-i 2.0 scores to those from other instruments.

In terms of the use and applications of the EQ-i 2.0, evidence is provided that the EQ-i 2.0 scores are related to external criteria, including expected differences between the following groups of individuals:

- Leaders and non-leaders
- Individuals with higher, compared to lower, levels of education
- Control group to clinical groups (i.e., individuals diagnosed with clinical depression or other psychological conditions)

Following these analyses, results from an examination of potential bias across racial/ethnic groups will be presented. As a general psychological characteristic, EI is expected to be similar across racial/ethnic groups; group differences would indicate that the EQ-i 2.0 may be biased towards certain racial/ethnic groups.

Finally, the Validity scales were validated:

- The validity of the Positive Impression and Negative Impression scales was examined by comparing scale scores between individuals instructed to present either overly positive or negative impressions to individuals who completed the scales under standard instructions.
- The validity of the Inconsistency Index was examined by comparing scores between the EQ-i 2.0 normative sample and a dataset of randomly generated EQ-i 2.0 item responses.

### **Content Validity**

Content validity is achieved when an assessment shows adequate coverage of the content it is proposed to measure, based on the conceptual framework of the construct. Support for this type of validity is often provided through non-statistical methods (Jackson, 1971). For the EQ-i 2.0, content validity of the items was analyzed by mapping their relevance to the EI construct by content experts. The conceptual framework of the EQ-i 2.0 is highly similar to that of its predecessor, the EQ-i (see Bar-On, 2004). Content validity of the original EQ-i was established through the systematic method of item generation (see *EQ-i 2.0 Stages of Development*). Specifically, the essence of each of the factors relevant to EI was articulated through detailed definitions. Items were then developed to encompass these definitions. Content experts scrutinized these items for their relevance to EI and the factors with which they were associated. Any items deemed irrelevant to a particular factor were moved to a more relevant factor, or discarded if their relevance could not be established. Based on these procedures, the final form of the EQ-i 2.0 adequately satisfied the requirements of content and face validity (Anastasi, 1988).

### **Factor Structure**

The conceptual framework of the EQ-i 2.0 can be considered hierarchical. As displayed in [Figure 3.4](#), several correlated factors comprise EI. The 15 subscales are categorized into the five composite scales, which combine to form the overall EI factor (i.e., Total EI). Evidence for the existence and appropriateness of the proposed EQ-i 2.0 factor structure was examined in several ways:

- Exploratory factor analyses (EFA) were used to determine whether the theoretically-based subscales empirically emerge from the normative data set.
- Confirmatory factor analyses (CFA) were used to determine whether the factor structure identified through theory and EFA results may be replicated in an independent data set.
- Correlations among composite scales and subscales were used to establish the degree of multidimensionality in the EQ-i 2.0. These correlations should be moderate in size; they should be high enough to indicate that the scales are all assessing a common underlying trait—emotional intelligence—yet they should not be so high as to indicate redundancy in the scales.

For the EFA and CFA analyses, the normative sample was split equally into two demographically-matched subsamples ([Table A.16](#)) to provide independent replication of the factor structure. Correlations among the scales were computed on the entire normative sample.

### Exploratory Factor Analyses

The factor structure of the EQ-i 2.0 items was determined through a series of exploratory factor analyses (EFAs). This analysis is exploratory, as the EQ-i 2.0 contains many new or revised items from the original EQ-i. Five EFAs were conducted on the exploratory subsample of the normative sample, analyzing the items within each composite scale separately. In each EFA, a three-factor solution was determined to be the most appropriate based on statistical (eigenvalues/scree plot) and non-statistical (interpretability) criteria. Principal axis factoring extraction was used because the goal of the analysis was to identify the underlying constructs expected to produce the EQ-i 2.0 scores. Direct oblimin (i.e., oblique) rotation was used because the factors were expected to correlate with each other, given that they all share a common underlying construct (i.e., the composite scale factor). Reverse-scoring was applied to relevant items prior to the analysis. Factor loadings were considered significant if they reached at least  $\pm .300$ , and an item was defined as cross-loading if it was significant on more than one factor and had loadings within .100 of each other on these factors.

For the **Self-Perception Composite** EFA, the first factor contained eight items covering areas such as self-confidence, self-respect, and a generally positive self-image, matching the definition of the Self-Regard subscale. The second factor contained seven items covering awareness and understanding of one's own emotions, matching the definition of the Emotional Self-Awareness subscale. The third factor contained nine items covering personal striving, ambition, and achievement, and matched the definition of the Self-Actualization subscale. Each item loaded significantly onto one factor and there were no cross-loadings.

For the **Self-Expression Composite** EFA, the first factor contained eight items covering areas such as autonomy and self-sufficiency, corresponding with the definition of the Independence subscale. The second factor contained eight items relating to one's ability to describe, express, and share their emotions, matching the definition of the Emotional Expression subscale. The third subscale contained seven items referring to one's tendencies towards being direct and "speaking one's mind," matching the definition of the Assertiveness subscale. Again, each item loaded significantly onto one factor and there were no cross-loadings.

In the **Interpersonal Composite** EFA, the first factor contained eight items covering areas such as sociability and friendliness, corresponding to the definition of the Interpersonal Relationships subscale. The second factor contained nine items referring to one's awareness, receptiveness, and

respectfulness towards the emotions of others, corresponding with the definition of the Empathy subscale. The third factor covered consciousness of social/global issues and one's contributions towards addressing these issues, matching the definition of the Social Responsibility subscale. Each item loaded onto one factor with no cross-loadings.

The first factor emerging from the **Decision Making Composite** EFA included eight items referring to one's emotional process when faced with problems, matching the definition of the Problem Solving subscale. The second factor contained eight items describing one's general awareness and tendency to be objective and impartial, corresponding to the definition of the Reality Testing subscale. The third factor contained eight items covering one's ability to combat impulses and temptations, matching the definition of the Impulse Control subscale. Each item loaded significantly onto one factor except for one item (I interrupt when others are speaking). This item was retained on the Impulse Control factor due to its theoretical relevance and the fact that it loaded more highly on the Impulse Control factor (.248) than on the other two factors (.034 and .004). No items cross-loaded across multiple factors.

Finally, the first factor generated from the **Stress Management Composite** EFA included eight items describing one's positive outlook towards other people and the future in general, matching the definition of the Optimism subscale. The second factor contained eight items describing one's ability to manage change and unpredictability, corresponding to the definition of the Flexibility subscale. The third factor contained eight items referring to one's ability to endure and cope with high-pressure situations and matched the definition of the Stress Tolerance subscale. Each item loaded onto a single factor with no cross-loadings.

To summarize, the EFAs generated an easily interpretable set of fifteen factors from the EQ-i 2.0 items. In addition, the items empirically grouped into the factors outlined by the theoretical framework of the instrument.

### **Confirmatory Factor Analysis**

Confirmatory factor analyses (CFAs) were conducted on the confirmatory subsample of the EQ-i 2.0 normative data. Six models were tested. The first, called the Overall Model, consisted of the five composite scales loading onto Total EI. The other five CFAs were conducted at the composite scale level, each with the three relevant subscales loading onto their respective composite scale. Results from these analyses provide further support for the theoretical factor structure of the EQ-i 2.0, as well as the empirical results generated by the EFAs. Goodness of fit indices are displayed in [Table A.17](#). Specifically, the Goodness of Fit Index (GFI; Jöreskog & Sörbom, 1986), Adjusted Goodness of Fit Index (AGFI; Jöreskog & Sörbom, 1986), Normed Fit Index (NFI; Bentler & Bonett, 1980), Non-Normed Fit Index (NNFI; Bentler & Bonett, 1980), Comparative Fit Index (CFI; Bentler, 1990), and Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980) were examined to evaluate the fit of the models. General guidelines for adequate model fit are values below .10 for the RMSEA and above .90 for the remaining fit indices. Values suggested adequate fit for the models, providing further support for the factor structure of the EQ-i 2.0 as outlined by theory and EFA results.

### **Correlations Among Eq-I 2.0 Composite Scales And Subscales**

After establishing the existence of the proposed subscales through EFA and obtaining further verification through CFA, correlations among the EQ-i 2.0 composite scales and subscales were

examined to determine the degree of cohesiveness among them. It is expected that these correlations will be generally high, given that they are all measuring the same underlying construct—emotional intelligence—but they should not be so high as to indicate redundancy between the subscales. [Tables A.18](#) (Composite Scales) and [A.19](#) (Subscales) display these correlations observed in the EQ-i 2.0 normative sample. These correlations matched closely to hypotheses. Each composite scale correlation reached at least a large effect size, ranging from  $r = .50$  (Interpersonal/Decision Making) to  $r = .78$  (Self-Perception/Stress Management). Subscale correlations were also of the expected magnitude. As highlighted in [Table A.19](#), virtually all subscale correlations within a composite reached at least a medium effect size and over half reached at least a large effect size, ranging from  $r = .27$  (Reality Testing/Impulse Control) to  $r = .70$  (Self-Regard/Self-Actualization). These results support the notion that a single, underlying dimension is being represented in the EQ-i 2.0, yet there is clear evidence for the multidimensional nature of the assessment.

### **Relationship of the EQ-i 2.0 to Other Measures**

The validity of the EQ-i 2.0 was further evaluated by examining its overlap with other psychological measures. These analyses inform whether the EQ-i 2.0 assesses the construct it is intended to assess—namely, emotional intelligence. Specifically, correlations between the EQ-i 2.0 and these other measures are examined. The expected pattern of correlations (magnitude, direction) depends on the relevance and degree of overlap among the psychological constructs these measures are proposed to assess. Validity is supported by the extent to which the actual correlations correspond with these theoretical associations. For example, is the EQ-i 2.0 related to other measures of emotional intelligence but unrelated to measures of different content, like critical thinking? For the EQ-i 2.0, these external psychological measures included

- the original version of the EQ-i (Bar-On, 2004);
- the Social Skills Inventory (SSI; Riggio & Carney, 2003), a measure of emotional and social communication skills;
- the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), a measure of fundamental personality traits;
- the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002), an ability-based measure of EI, and
- the Watson-Glaser II Critical Thinking Appraisal (Watson & Glaser, 2009), a measure of critical thinking.

Demographic characteristics of the samples used in these analyses are displayed in [Table A.20](#).

### **Relationship Between □Eq-I 2.0 And The Original Eq-I**

The original EQ-i (Bar-On, 2004) is a 133-item self-report measure designed to assess emotional intelligence (EI). Bar-On defines EI as “an array of non-cognitive capabilities, competencies, and skills that influence one’s ability to succeed in coping with environmental demands and pressures” (p. 14). Other key features of the EQ-i’s conceptual framework are that it is multifactorial and relates to potential for performance rather than performance itself (i.e., the potential to succeed rather than success itself). It is process-oriented rather than outcome-oriented, unlike ability-based conceptualizations of EI such as that measured by the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2002). The 15 EI constructs assessed by the EQ-i are measured by the EQ-i subscales, organized as outlined in [Figure 3.1](#). This figure also illustrates the slight changes made to the organization of the subscales in the EQ-

i 2.0 revision. Because the models are very similar, it is expected that correlations between the EQ-i and EQ-i 2.0 will be large.

Correlations between the EQ-i 2.0 and the original EQ-i are displayed in [Table A.21](#). Correlations between overlapping subscales are presented in this table (i.e., the correlation between the two Interpersonal Relationships subscales, the correlation between the two Flexibility subscales, and so on). Despite the updates made to the EQ-i 2.0 from the original EQ-i, correlations between the subscales on the two measures were high. Correlations between the Total EI score of each measure was  $r = .90$ , suggesting a high degree of overlap between the two versions at the overall EI level. The majority of the subscale correlations between the EQ-i and EQ-i 2.0 were high. This trend was particularly evident for subscales that underwent very minor changes between the two versions of the scale, with correlations ranging from  $r = .65$  to  $r = .88$  (see shaded cells in [Table A.21](#)). Conversely, for subscales that underwent more dramatic changes between versions (see *The EQ-i 2.0 Framework*; unshaded cells in [Table A.21](#)), correlations were still high but lower, as expected, than those found for the unchanged subscales. These correlations ranged from  $r = .49$  to  $r = .57$ . One exception was the correlation between Emotional Expression and the original EQ-i Emotional Self-Awareness subscale ( $r = .84$ ). Many of the Emotional Self-Awareness items from the original EQ-i were incorporated into the new Emotional Expression subscale, which explains this high correlation. Overall, correlations between the EQ-i 2.0 and the original EQ-i reflect not only the stability of the construct measured by the two assessments, but also the changes in item content made in the recent update to the EQ-i 2.0.

### **Relationship Between EQ-I 2.0 And Ssi**

The Social Skills Inventory (SSI; Riggio & Carney, 2003) is a 90-item self-report measure designed to assess “basic social communication skills” (Riggio & Carney, p. 5). The scale captures the expression, sensitivity, and control (i.e., regulation) aspects of communication in two domains: emotional (nonverbal) and social (verbal). This conceptualization results in six subscales: Emotional Expression, Emotional Sensitivity, Emotional Control, Social Expression, Social Sensitivity, and Social Control. Along with a Total SSI Score, these subscales are collapsed into Total Emotional and Social Scales as well as Total Expression, Control, and Sensitivity Scales. It is expected that the EQ-i 2.0 will correlate more strongly with the Emotional Scales than the Social Scales. For instance, although the SSI authors admit that the tool does not fully capture all aspects of EI, they specifically state that the SSI Emotional subscales “can be used as indicators of emotional intelligence” and “could be used as an alternative to existing self-report measures of emotional intelligence” (Riggio & Carney, p. 6). These statements summarize the relevance of the SSI to the EQ-i 2.0.

Emotional intelligence is proposed to be relevant to social skills as measured by the SSI, especially the SSI Emotional Subscales. Therefore, most correlations between the EQ-i 2.0 and the SSI should be strong and positive. As illustrated in [Table A.22](#), the EQ-i 2.0 Total EI score correlated positively with the SSI Total Score ( $r = .54$ ;  $p < .01$ ). With the exception of Impulse Control ( $r = -.13$ ;  $p = .19$ ), each of the EQ-i 2.0 composite scales and subscales correlated significantly with the SSI Total Score. The EQ-i 2.0 Total EI score also showed significant positive correlations with most of the SSI Subscales. Exceptions were a non-significant correlation with the Total Sensitivity Scale ( $r = .08$ ;  $p = .43$ ) and a significant negative correlation with the Social Sensitivity Scale ( $r = -.35$ ;  $p < .01$ ). Riggio and Carney describe the

Social Sensitivity Scale as measuring “an individual’s sensitivity to and understanding of the norms governing appropriate social behavior” (p. 5), and also suggest that extremely high scores may indicate self-consciousness and general insecurity, which could explain the negative correlation with the EQ-i 2.0. The nonsignificant correlation between the EQ-i 2.0 and the Total Sensitivity Scale is likely due to the former’s positive correlation with the Emotional Sensitivity Scale and negative correlation with the Social Sensitivity Scale cancelling each other out. These results provide support for the idea that higher EI is related to stronger social skills.

### **Relationship Between □Eq-I 2.0 And Neo-Ffi**

The NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) is a shortened, 60-item version of the NEO Personality Inventory-Revised (NEO-PI-R; Costa & McCrae, 1992). This scale measures what are considered to be the five fundamental personality traits according to the Five-Factor Model of personality: Neuroticism, Conscientiousness, Openness to Experience, Agreeableness, and Extraversion. Conceptually, the Big Five and emotional intelligence share certain features, such as positive correlations with occupational performance (e.g., Mount & Barrick, 1998). In a recent meta-analysis, Van Rooy and Viswesvaran (2004) found significant positive correlations between EI and each of the Big Five factors, ranging from  $r = .23$  (Agreeableness, Openness to Experience) to  $r = .34$  (Extraversion). Therefore, it is expected that the EQ-i 2.0 will correlate positively with the NEO-FFI subscales (except for Neuroticism, where negative correlations are expected).

The EQ-i 2.0 Total EI score correlated significantly with the NEO-FFI Neuroticism (note that the negative correlations are in the expected direction), Extraversion, Agreeableness, and Conscientiousness subscales, but not with Openness to Experience ([Table A.23](#)). The pattern of correlations suggests that EI is distinct from personality. The correlations also support the hypotheses that high levels of Neuroticism may inhibit EI development, whereas high levels of Extraversion and Conscientiousness may help facilitate EI skills.

### **Relationship Between □Eq-I 2.0 And Msceit**

The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2002) is a 141-item ability-based measure of EI. The MSCEIT is ability-based in that it considers EI as a skill and measures it through items that require the respondent to demonstrate their level of EI by performing various relevant tasks and solving emotional problems. The scale is a “test” in the true sense of the word in that items are considered to have correct and incorrect responses, based on either general consensus or expert consensus. This feature defines the MSCEIT as outcome-oriented as opposed to process-oriented as in the EQ-i 2.0 (see Bar-On, 2004). The distinction between ability-based measures like the MSCEIT and trait-based measures like the EQ-i 2.0 has long been established by researchers (Austin, 2010; Brackett & Mayer, 2003; Mayer et al., 2002; O’Boyle et al., 2011; Van Rooy & Viswesvaran, 2004). In line with this research it is expected that the EQ-i 2.0 and the MSCEIT would not be strongly correlated.

In the MSCEIT, the Total Emotional Intelligence Quotient (EIQ) score comprises eight subscales called Tasks: Faces, Pictures, Sensations, Facilitation, Changes, Blends, Emotion Management, and Emotional Relations. These Tasks are categorized into four Branch scores: Perceiving Emotions (Faces, Pictures), Facilitating Thought (Sensation, Facilitation), Understanding Emotions (Changes, Blends), and Managing Emotions (Emotion Management, Emotional Relations). The Branch scores are further categorized into two Area scores: Experiential

(Perceiving Emotions, Facilitating Thought) and Strategic (Understanding Emotions, Managing Emotions). Task, Branch, and Area scores provide different levels of scope of the individual's EI abilities. Further description of these scales is provided by Mayer et al. (2002).

The EQ-i 2.0 conceptualizes EI as a trait-based measure, whereas the MSCEIT assesses EI as an ability-based measure. For these reasons, it is expected that the relationship between MSCEIT and EQ-i 2.0 scores will be moderate, at best. In our sample, these correlations are displayed in [Tables A.24 and A.25](#). Indeed, the correlation between the EQ-i 2.0 Total EI score and the MSCEIT Total EI Score was  $r = .12$  ( $p = .22$ ). The vast majority of MSCEIT Task Scores, Branch Scores, and Area Scores were not significantly correlated with EQ-i 2.0 composite or subscale scores. This pattern of results demonstrates that the EQ-i 2.0 measures trait-based EI that does not overlap with EI as measured by the MSCEIT. On a larger, conceptual level, these results support the idea that trait-based EI and ability-based EI are independent constructs.

### **Relationship Between EQ-i 2.0 And Watson-Glaser II**

The Watson-Glaser II Critical Thinking Appraisal (Watson & Glaser, 2009) is “(d)esigned to measure important abilities and skills involved in critical thinking” (p. 1). Along with a Total Score, the three subscales of the Watson-Glaser II are Recognize Assumptions, Evaluate Arguments, and Draw Conclusions. Individuals must evaluate a series of exercises that cover these areas, such as rating the degree of truth or falsity of various inferences. Validity of the scale is demonstrated through correlations with similar ability measures such as the Wechsler Adult Intelligence Scales-IV (WAIS-IV; Wechsler, 2008) and occupational and academic success. However, emotional intelligence is considered to be independent of more traditional cognitive abilities such as critical thinking. Therefore, it is presumed that the EQ-i 2.0 will be largely uncorrelated with the Watson-Glaser II.

The correlations between the Watson-Glaser II and the EQ-i 2.0 are displayed in [Table A.26](#). The correlation between the Total Scores of the EQ-i 2.0 and the Watson-Glaser II was not statistically significant ( $r = -.05$ ;  $p = .62$ ). Regarding the subscales of the Watson-Glaser II, the EQ-i 2.0 Total EI score was also uncorrelated with the Recognize Assumptions ( $r = .03$ ;  $p = .76$ ) and Draw Conclusions ( $r = .02$ ,  $p = .84$ ) subscales, but was significantly negatively correlated with the Evaluate Arguments subscale ( $r = -.25$ ,  $p < .01$ ). Watson and Glaser (2009) state that lower scores on the Evaluate Arguments subscale may be found in individuals who allow high levels of emotion to “cloud objectivity and the ability to accurately evaluate arguments” (p. 3). This trend was also found for the EQ-i 2.0 composite scales and subscales. That is, the majority of EQ-i 2.0 composite scales and subscales were uncorrelated with the Watson-Glaser II Total Score and the Recognize Assumptions and Draw Conclusions subscales, but were negatively correlated with the Evaluate Arguments subscale. These results provide support for the independence of EI and cognitive intelligence; however, they also demonstrate the impact of emotional skills on the ability to effectively evaluate arguments. In summary, strong evidence has been provided that the EQ-i 2.0 measures the constructs it was designed to measure. It shows strong correlations with measures of similar constructs, and little or no correlation with measures of divergent constructs.

### **Group Differences in EQ-i 2.0 Scores**

The validity of the EQ-i 2.0 was further evaluated by examining scores among groups that are expected to show differences in EI. Specifically, validity was assessed by examining (a)

corporate job success: corporate leaders vs. the general population; (b) academic achievement: individuals with higher (i.e., post-graduate) compared to lower (high school or less) levels of education; and (c) clinical group differences: individuals with a diagnosed psychological illness vs. a demographically matched control group.

### **Relationship Between □Eq-I 2.0 And Corporate Job Success**

Occupational success is one highly relevant, consistent, and important outcome of high emotional intelligence. Therefore, the EQ-i 2.0 would be validated by showing higher scores among individuals who have excelled in their profession. To test this hypothesis, EQ-i 2.0 scores were compared between 221 corporate leaders (i.e., CEOs and other C-level leaders, senior executives, directors, and managers; see [Table A.27](#) for demographics) and the normative sample. Results are displayed in [Table A.28](#). Relative to the normative mean score of 100, leaders scored consistently higher on the EQ-i 2.0 Total EI score and all composite scales and subscales. Leaders produced a mean score of 112.2 ( $SD = 11.7$ ) on the Total EI score, which represents a large difference when compared to the normative average ( $d = 0.82$ ). Mean scores on the composite scales and subscales ranged from 104.2 ( $SD = 14.0$ ; Impulse Control) to 113.1 ( $SD = 10.4$ ; Self-Actualization), with most differences representing medium or large effects. These results indicate that occupational success, measured by one's advancement into a senior-level corporate position, is related to greater emotional intelligence.

### **Relationship Between □Eq-I 2.0 Scores And Academic Achievement**

Academic achievement is another key outcome related to emotional intelligence. Therefore, EQ-i 2.0 scores are expected to be higher among individuals who have achieved higher levels of accomplishment in educational pursuits. EQ-i 2.0 scores were compared between individuals in the normative sample who achieved a post-graduate degree (e.g., M.A., Ph.D., MBA;  $N = 402$ ) and those who progressed no farther than a high school degree ( $N = 1,451$ ). Comparisons were conducted using analysis of covariance (ANCOVA) for the EQ-i 2.0 Total EI score and two multivariate analyses of covariance (MANCOVA) for the subscales and composite scales, with age group, gender, and race/ethnicity (White vs. non-White) included as covariates. As illustrated in [Table A.29](#), higher Total EI scores were found for post-secondary graduates ( $M = 103.2$ ,  $SD = 14.8$ ) relative to high school graduates ( $M = 98.1$ ,  $SD = 15.5$ ), showing a small-to-moderate effect size ( $d = 0.33$ ). Post-graduates also scored higher on most of the composite scales and subscales. Scores that showed at least a small difference ranged from Problem Solving ( $d = 0.22$ ) to Self-Actualization ( $d = 0.54$ ), with most differences being found in the Decision Making and Stress Management areas. Overall, the results demonstrate that greater academic achievement tends to be associated with higher EI.

### **Clinical Group Differences In Eq-I 2.0 Scores**

Because emotional intelligence is associated with daily functioning, it is presumed to be lower in individuals with various psychiatric or psychological conditions. Based on this assumption, it follows that differences in EQ-i 2.0 scores should be found between clinical and non-clinical (i.e., general population) individuals. Analysis of covariance (ANCOVA) was used to compare mean EQ-i 2.0 Total EI scores across three groups: general population (or the control group taken from the normative sample), individuals diagnosed as depressed or dysthymic, and individuals with another clinical diagnosis (see [Table A.30](#) for demographic characteristics of the samples). Age, gender, and race/ethnicity were included in the analysis as covariates. This

procedure was repeated using two separate multivariate analyses of covariance (MANCOVAs) to examine the EQ-i 2.0 composite scales and subscales. Results ([Table A.31](#)) demonstrated a significant effect of clinical status for the EQ-i 2.0 Total EI score ( $F [2, 221] = 7.89, p < .01$ ). Specifically, the mean score for the general population group was higher than for each of the depressed/dysthymic and other clinical groups, and each difference approached or exceeded a medium effect size (Cohen's  $d = 0.57$  and  $0.45$ , respectively). This trend was replicated for all of the composite scales except Interpersonal. The Interpersonal composite was not significant because there were no differences for Empathy or Social Responsibility. The Interpersonal Relationships subscale, however, did show significant differences between groups. At the subscale level, the effect of the general population group scoring higher than the clinical group was found for more than half of the subscales. The subscales that showed the largest differences were those that would be expected on a conceptual level. For example, the largest differences between the general population and depressed/dysthymic groups were found for the Self-Regard and Happiness subscales. These results provide further evidence for the validity of the EQ-i 2.0.

### **Comparisons among Racial/Ethnic Groups**

The examination of potential racial or ethnic bias is always of critical importance in the development of an assessment. Specifically, it is vital to ensure that assessment scores do not show large differences among racial/ethnic groups when they are not expected to. For the EQ-i 2.0, test bias was examined by comparing mean scores across various racial/ethnic groups (White, Black, Hispanic/Latino) in the normative sample. Analysis of covariance (ANCOVA) was used to compare these three groups on the EQ-i 2.0 Total EI score, using gender, age group, and education level as covariates (in order to control for the effects of these demographic variables). Two separate multivariate analyses of covariance (MANCOVAs) were used to examine the composite scales and subscales. Results demonstrated that the effect of race/ethnicity on EQ-i 2.0 scores was statistically significant; however, the effect sizes were in the small-to-medium range. In scales that did show differences, Black and Hispanic/Latino respondents generally showed slightly higher scores than White respondents, though these differences were typically only a few standard score points in magnitude (see [Table A.32](#)). These results demonstrate that the EQ-i 2.0 does not show strong differences among racial/ethnic groups and there was no evidence of test bias toward minority groups.

### **Validity Scale Validation**

Dishonest or exaggerated responses are always a concern with self-report instruments. Insincere responses undermine the veracity of an individual's scores on a self-report assessment, which can have significant consequences. The original EQ-i included three scales—Positive Impression, Negative Impression, and Inconsistency Index—to detect illegitimate response styles. These scales were also developed for the EQ-i 2.0. Validity studies were conducted in order to determine if the Validity scales do, in fact, capture positive, negative, and/or inconsistent response styles.

### **Positive Impression & Negative Impression Scales**

Positive and negative impression styles might be used intentionally or unintentionally when responding to a self-report questionnaire. Positive impression occurs when an individual responds to questions in such a way as to make themselves appear in an unrealistically positive light. The reasons behind positive impression include self-deception, lack of insight, an unwillingness to face one's limitations, or various needs such as social conformity, approval,

self-protection, or avoidance of criticism (Crowne & Marlowe, 1964; Edwards, 1966; Frederiksen, 1965; Jackson, 1974). An attempt to make a positive impression is more apt to occur when, for example, one is applying for a job, seeking admission to an educational institution, or simply trying to impress someone. Conversely, a negative impression style consists of making oneself appear in an unrealistically negative light. Elevated negative impression scores can be caused by low self-esteem, or various needs such as attention, sympathy, or help in resolving personal problems (Crowne & Marlowe, 1964; Frederiksen, 1965; Jackson, 1974). To detect these response styles in the EQ-i 2.0, Positive Impression (PI) and Negative Impression (NI) scales were developed (see *EQ-i 2.0 Stages of Development*). PI and NI scales are traditionally validated by examining the scores of individuals who are motivated to present themselves favorably or unfavorably, respectively, to individuals who respond to the assessment under standard instructions without such motivation. The PI and NI scales were validated using a standard between-subjects simulation study conducted during the norming phase of development. Participants were given instructions designed to elicit either a positive or negative response style while completing the EQ-i 2.0. Instructions designed to elicit a positive response style asked the respondent to imagine they are completing the EQ-i 2.0 as part of an application for a highly desirable job, and must therefore try to give themselves the highest scores possible. Instructions for the negative response style condition asked respondents to imagine they are completing the EQ-i 2.0 as part of a mandatory application for a mentoring program that he or she does not want to participate in, and must therefore try to give themselves the lowest scores possible in order to be selected out of the program. Two demographically-matched control groups who completed the EQ-i 2.0 under standard instructions were selected for comparison with the two simulation groups. Presumably, PI and NI scores would be higher in individuals who were instructed to simulate positive or negative response styles, respectively, than those who responded under standard conditions.

Results from the simulation studies are displayed in [Table A.33](#). As expected, PI scores from the positive response style group were significantly higher than those in the control group. The difference between the two groups is quantified as a medium-to-large effect size. Similarly, NI scores from the negative response style group were significantly higher than those in the control group. This difference exceeded the standard guideline for a large effect size. These results provide support for the validity of the PI and NI scales.

### **Inconsistency Index**

Inconsistent responding occurs when a respondent rates similar items in dissimilar or opposite ways. For example, a respondent who endorses (i.e., responds “Always/Almost Always”) both of the items “I like parties” and “I don’t like parties” would be responding inconsistently. Like positive impression and negative impression styles, inconsistent responding might occur intentionally or unintentionally. Various reasons for inconsistent responding include deliberate sabotage or noncompliance, fatigue, incomprehension of the items or instructions, inattention, disinterest, and a lack of motivation.

To detect inconsistent responding in the EQ-i 2.0, an Inconsistency Index (IncX) was developed. This scale is comprised of 10 pairs of highly related items, which should elicit similar responses within each pair of items. If the respondent provides very different ratings to several pairs of items that should be rated similarly, then inconsistent responding may be suspected (see *The EQ-i 2.0 Framework*). Traditionally, inconsistency scales are validated by comparing scores

generated from individuals who respond to assessment items randomly, to individuals who respond under standard conditions. These random protocols can be generated by human respondents or computer programs. A computer program (IBM SPSS Statistics 19.0.0, 2010) was used to generate a data set of 4,000 random EQ-i 2.0 response sets to compare to the normative data. Evidence of the validity of the IncX would be demonstrated if the cutoff identified a large proportion of the random response sets, and if IncX scores were higher, on average, than those in a control sample. Furthermore, these results would provide independent validation of the choice of cutoff to be used to identify scores as potentially invalid that was developed from the normative sample. [Table A.34](#) illustrates the proportion of response sets at each IncX raw score. Results demonstrated that a score of 3, which identified only 3.5% of the normative sample as potentially inconsistent, identified 93.3% of the random response sets as potentially inconsistent. Furthermore, mean IncX scores were dramatically higher in the random sample than in the normative sample ( $d = 3.36$ ; [Table A.34](#)), a difference that easily exceeded the criteria for a large effect size. These results demonstrate a high degree of predictive validity for the EQ-i 2.0 IncX.

### **Validity Summary**

Several analyses were conducted to examine the validity of the EQ-i 2.0. Content validity analyses suggest that all relevant facets of the Bar-On conceptualization of EI are being captured by the EQ-i 2.0. Exploratory factor analyses suggested that this overarching single factor (EI) may be represented by 15 correlated subscales, which in turn may be combined into five correlated composite scales (i.e., a 1-5-15 Factor Model of Emotional Intelligence). This factor structure was corroborated through confirmatory factor analyses. Correlations among the composite scales and subscales provide support for the unidimensionality of the EQ-i 2.0. Validity was supported by expected correlations with the original EQ-i and measures of social skills and general personality, as well as a lack of correlation with measures of ability-based EI and cognitive intelligence. Further validity evidence was provided by expected group differences with regard to occupational success, academic achievement, and psychological adjustment. Comparisons among racial/ethnic groups in the normative sample provided no evidence for racial/ethnic bias against minority groups in the EQ-i 2.0. The validity scales (Positive Impression, Negative Impression, and Inconsistency Index) were validated through expected differences in scores between known invalid responses and those of control groups. Overall, the analyses suggest that the EQ-i 2.0 is a valid measure of EI.

## **EQ 360 2.0 Pilot Study and Standardization**

This section describes the EQ 360 2.0 standardization procedure, including the method of data collection, the properties of the normative sample, and the effects of age and gender on the results.

### **Data Collection**

Data collection for the EQ 360 2.0 followed multiple stages between July 2009 and August 2010. More than 4,000 participants completed the EQ 360 2.0 over this time period.

### **Pilot Phase**

The first stage of data collection, the collection of pilot data, took place between July 2009 and

November 2009. Raters were required to provide demographic information of the individuals they rated (i.e., “ratees”) along with EQ 360 2.0 ratings. The ratees ( $N = 759$ ) were 59.2% female, the majority were White (74.3%), and there was good representation across several age groups ([Table A.35](#)). These data were collected to ensure the basic functionality of the EQ 360 2.0 (e.g., instructions, response options, administration time) was adequate.

## Normative Phase

The second phase of data collection that included the collection of data for the normative sample, as well as reliability and validity data, took place between March 2010 and August 2010. Data were gathered from all 50 U.S. states and the District of Columbia, as well as from all 10 Canadian provinces. Raters were sent an email invitation to participate in the EQ 360 2.0 data collection process. The data collection and authentication procedures were identical to those used for the EQ-i 2.0 (see *EQ-i 2.0 – Data Collection – Normative Phase* in the [EQ-i 2.0 Pilot Study and Standardization](#) section. The following section focuses on a description of the normative samples; see the [Reliability](#) and [Validity](#) sections on this page for more information on the reliability and validity samples.

In order to create representative normative samples, specific demographic (i.e., age, gender, race/ethnicity, and geographic targets), guided by recent Canadian and U.S. Census information (i.e., Statistics Canada, 2006; U.S. Bureau of the Census, 2008), were utilized during the data collection procedure. Information was collected on each ratee’s gender, age, race/ethnicity (Asian/Pacific Islander, Black/African-American/African-Canadian, Hispanic/Latino, White, Multiracial, and Other), employment status (employed/self-employed, unemployed, retired, and other), and geographic location (state/province and country). For ease of presentation, race/ethnicity groups are referred to in this manual as follows: Black, Hispanic/Latino, White, and Other. For the EQ 360 2.0, this information was provided about the ratee (i.e., the person being rated) by the rater (i.e., the person completing the assessment). Information about the type and strength of the rater-ratee relationship was also collected.

## Standardization

The standardization process for the EQ 360 2.0 was similar to that of the EQ-i 2.0. A second normative dataset was collected for the EQ 360 2.0, requiring separate norms and statistical analyses.

## Normative Sample

Normative data for the EQ 360 2.0 were collected concurrently with the EQ-i 2.0, during March 2010 and April 2010. Data for the EQ 360 2.0 required raters to rate an individual (“the ratee”) on the EQ 360 2.0 (including the collection of various demographic information about both themselves and the ratees). During this time period, 3,413 participants provided EQ 360 2.0 data for standardization purposes. From these data, a demographically and geographically representative database of 3,200 ratees was selected as the EQ 360 2.0 normative sample. Statistical analyses showed no strong differences between U.S. and Canadian participants in EQ 360 2.0 scores ([Table A.36](#)); therefore, data from both countries were included in the normative sample.

**Rater Description.** The sample of 3,200 raters (i.e., the participants providing the ratings) was 59.2% female, with a mean age of 46.8 years, ( $SD = 13.5$  years). The sample was primarily

White (81.2%), 5.2% were Black, 3.7% were Hispanic/Latino, and 9.9% were of other races/ethnicities. Approximately one-third of the sample was from the U.S. South (33.7%), while 22.0% was from the U.S. West, 20.5% was from the U.S. Midwest, 16.1% was from the U.S. Northeast, 5.6% was from Central Canada, 0.9% was from the Canadian West and Prairies, and 0.3% was from the Canadian East. More than half of the raters had at least a college/university education (54.7%), 27.8% had some college/university education, and 17.6% had a high school diploma or less. The majority (90.4%) of raters knew the ratee for over a year (see [Table A.37](#)) and over half of the raters stated that they knew the ratee “Well” or “Very Well” on a four-point scale ranging from Not Very Well (0) to Very Well (3; see [Table A.38](#)). Therefore, the raters knew the ratees for long enough, and well enough, to provide valid EQ 360 2.0 ratings.

**Ratee Sample.** The normative sample was stratified to match the Census based on the ratee’s (i.e., the person being rated) demographic characteristics. The sample included an equal ratio of males to females, stratified equally across four rater types: direct report (i.e., the ratee is the rater’s manager), manager (i.e., the ratee is the rater’s direct report), work peer, and friend/family member ([Table A.39](#)). Participants were proportioned similarly across most of the age groups, although there were relatively fewer at the lower age range, as an attempt was not made to collect direct-report data for managers under the age of 25 ([Table A.40](#)) as they are relatively rare in the population. Race/ethnicity was stratified by Census figures within rater type, given that these distributions differed slightly across rater type ([Table A.41](#)). The normative sample met each of these targets within 3%, and was within 1% in most cases. Finally, there was good representation from all U.S. and Canadian geographic regions ([Table A.42](#)).

**Focus on Effects Size.** The effects of gender, age, and rater type were examined in the EQ 360 2.0 normative data. As with the EQ-i 2.0 data, the large EQ 360 2.0 normative sample size dictates that effect sizes should be considered more strongly than significance tests (see the [Effect Size](#) section). Cohen’s *d* values are reported to describe the size of gender effects, and partial eta-squared (partial  $\eta^2$ ) values are used to describe the effects of age and rater type.

## Norming Procedures

Similar to the EQ-i 2.0, the first step in the EQ 360 2.0 norming procedure was to determine if any demographic trends existed in the data. Demographic effects were examined using an analysis of covariance (ANCOVA) for the EQ 360 2.0 Total EI score and two separate multivariate analyses of covariance (MANCOVA) for the composite scales and subscales. Rater type (direct report, manager, work peer, family/friend), gender, and age group were examined using race/ethnicity (White vs. non-White) as a covariate. In an attempt to control for Type I errors that might occur with multiple analyses, a more conservative criterion of  $p < .01$  was used for all *F*-tests. Results at the multivariate level revealed significant effects of gender, age, and rater type for both the composites and the subscales ([Table A.43](#)); the only significant interaction at the multivariate level was for the interaction of age and rater type for the subscales. Given these results, the univariate effects are described in detail next.

Overall, gender and age effects were less pronounced in the EQ 360 2.0 normative sample than they were in the EQ-i 2.0 sample (see [Table A.44](#) for effect sizes and [Tables A.45 through A.47](#) for descriptive statistics and significance test results). There were no gender differences that reached even a small effect size for the Total EI score or for any of the composite scales. At the subscale level, only Emotional Expression reached a small effect size, with females being rated

higher than males. With respect to age, Independence, Social Responsibility, Impulse Control, and Flexibility reached small effect sizes. For Independence, Social Responsibility, and Impulse Control, the effect was attributable to lower scores among 18–29-year-olds. For Flexibility, scores decreased in the older age groups. Very few meaningful differences were found across rater types. No meaningful differences were found across rater types for the EQ 360 2.0 Total EI score (i.e., partial  $\eta^2 = .00$ ). Some minor differences were found across rater types for the composite scales and subscales, but all were small effect sizes (i.e., partial  $\eta^2$  lower than .06). None of the age  $\times$  rater type interactions reached significance at the univariate level, with the exception of Problem Solving ( $F [12, 555.80] = 2.55, p = .002$ ); however, the effect size was very small (partial  $\eta^2 = .01$ ).

Overall, the lack of meaningful demographic effects suggested it was unnecessary to create specific rater type-, age-, or gender-based norms for the EQ 360 2.0. Therefore, only overall (General Population) norms are available for the EQ 360 2.0. These norms were created using the same procedure as the EQ-i 2.0 General norms, but without the smoothing process (given no age groups were utilized). Standard scores (with a mean of 100 and standard deviation of 15) were computed for all scales. Skewness and kurtosis statistics (e.g., -0.42 and -0.34, respectively, for Total EI; see Figure B.2) were not large enough to suggest a normalizing transformation was necessary for the EQ 360 2.0 scores.

### **Standardization Summary**

More than 4,000 assessments were collected between 2009 and 2010 in the standardization of the EQ 360 2.0. A sample of 3,200 participants was chosen as the EQ 360 2.0 normative sample. The sample was evenly distributed by gender and rater type, and matched to the census based on race/ethnicity. Statistical analyses revealed a lack of meaningful differences in EQ 360 2.0 scores across gender, age group, or rater type. Therefore, a single normative group was created. The norming process resulted in standard scores with means of 100 and standard deviations of 15 for the Total EI score, composite scales, and subscales. The following sections describe the psychometric properties (i.e., reliability and validity) of the EQ 360 2.0.

## **EQ 360 2.0 Reliability**

Similar to the EQ-i 2.0, reliability analyses were conducted for the EQ 360 2.0. Specifically, internal consistency and test-retest reliability analyses were performed. A practical application of these analyses is to detect discrepancies between self (EQ-i 2.0) and rater (EQ 360 2.0) scores.

### **Internal Consistency**

Internal consistency conveys the degree to which a set of items are associated with one another. High levels of internal consistency suggest that the items are measuring a single, cohesive construct. Internal consistency is typically measured using Cronbach's alpha (Cronbach, 1951), which ranges from 0.0 to 1.0 with higher values reflecting higher internal consistency. □Cronbach's alpha values for the EQ 360 2.0 normative sample are displayed in [Table A.48](#). Similar to results found in the EQ-i 2.0 normative samples, most of these values ranged from good to excellent for the Total EI score, composite scales, and subscales, with all but one value reaching at least .82.

## Self-To-Rater Gaps: Comparing Eq-I 2.0 To Eq 360 2.0 Scores

The EQ 360 2.0 report includes a section that compares scores from self-ratings to scores from the rater groups. The process used to calculate confidence intervals and gaps between subscales for the EQ-i 2.0 report was again used to determine if the self to rater comparison revealed similar scores, or gaps between scores. Considering statistical results as well as practical functionality, results revealed that a critical value of 10 points was appropriate as the criterion for identifying self-to-rater gaps. This value is actually slightly smaller than those suggested by the statistical analyses, but was selected so that the user can be confident they are identifying any potentially important discrepancies between self and observer ratings of EI abilities. For example, self-report and rater-group subscale scores less than 10 points apart will be reported as being similar, while subscale scores that are 10 or more points apart will be reported as having a gap.

### Test-Retest Reliability and Stability

Similar to the EQ-i 2.0, test-retest reliability and stability were evaluated for the EQ 360 2.0. Test-retest reliability was calculated by examining the correlation between an individual's scores in two assessments, separated by a meaningful amount of time. Test-retest stability analyses were performed by calculating the difference between Time 1 and Time 2 standard scores for each individual in the test-retest sample.

For the EQ 360 2.0 sample, test-retest data was available for 203 individuals who were assessed roughly three weeks apart (mean interval = 19.30 days,  $SD = 2.44$  days, range = 14–23 days). Demographic characteristics of the ratees (i.e., the people being rated) in the retest sample are displayed in [Table A.49](#). Test-retest correlations ([Table A.50](#)) were high for the EQ 360 2.0 Total EI score, composite scales, and subscales, ranging from  $r = .76$  to  $.89$ .

Similar to the EQ-i 2.0, EQ 360 2.0 test-retest stability values were calculated as the difference between Time 1 and Time 2 standard scores. [Table A.51](#) displays the frequencies of these differences (positive differences indicate that scores increased over time whereas negative differences indicate that scores decreased over time), as well as the mean differences (i.e., the difference between Time 1 and Time 2 ratings for each individual averaged across the samples) and the 95% confidence intervals surrounding the mean differences. The results suggest scores were similar to those found in the EQ-i 2.0 samples: for all subscales, roughly 90% or more of individuals' scores did not change by more than one normative standard deviation (i.e., 15 standard score points) over time. For instance, 95.1% of EQ 360 2.0 Total EI scores deviated by less than one standard deviation over time. The mean difference was  $-0.18$  standard score units, and the 95% confidence interval ( $-0.93; 0.57$ ) contained zero. These results provide support that the EQ 360 2.0 captures the temporal stability of emotional intelligence, even when rated by outside observers.

### Reliability Summary

Overall, the EQ 360 2.0 demonstrates sound reliability. Internal consistency (alpha) values were generally high for the overall normative group, suggesting that the items cohesively measure Total EI as well as the constructs measured by the composite scales and subscales. Test-retest reliability and stability values were also high, reflecting a level of temporal stability that would be expected for emotional intelligence. Users of the EQ 360 2.0 can be confident that the scores generated by these assessments will be consistent and reliable.

## EQ 360 2.0 Validity

EQ 360 2.0 validity analyses were performed to ensure that the validity of the observer-rated version of the EQ-i 2.0 is comparable to the self-report version. These analyses are summarized in the following section. Specifically,

- the factor structure of the EQ 360 2.0 was examined through correlations among the composite scales and subscales;
- EQ 360 2.0 scores were compared to EQ-i 2.0 scores to evaluate self-other agreement (correlations) and self-other consistency (differences between EQ 360 2.0 and EQ-i 2.0 scores); and
- multiple regression analyses were conducted to determine the ability of the EQ 360 2.0 to correlate with a measure of social adjustment—the Social Adjustment Scale–Self-Report (Weismann, 1999)—independently of EQ-i 2.0 scores (i.e., to examine the added value of the EQ 360 2.0).

Following these results, analyses examining potential bias of raters in relation to the race/ethnicity of ratees will be illustrated, (that is, the degree to which White and non-White observers rate Black, Hispanic/Latino, and White ratees similarly). Based on these results, practical implications of the validity analyses include the use of EQ 360 2.0 assessments to inform decisions in occupational settings such as initial hiring or subsequent promotion, or acceptance into academic institutions, or the added value of using multiple sources (i.e., self-report, observer reports) in gathering this information about individuals, with the confidence that ratings are not affected by demographic variables.

### Correlations Among EQ 360 2.0 Composite Scales & Subscales

Correlations among the EQ 360 2.0 composite scales and subscales were examined in the normative sample to determine if the pattern of results found in the EQ-i 2.0 normative sample data would be replicated. [Tables A.52](#) (Composite Scales) and [A.53](#) (Subscales) display these correlations. These correlations were strong, and in most cases stronger than in the EQ-i 2.0 normative sample. Composite scale correlations ranged from  $r = .64$  (Self-Expression/Interpersonal) to  $r = .86$  (Decision Making/Stress Management). For the most part, subscale correlations were especially strong within the same composite, as expected (see shaded cells in [Table A.53](#)). Each of these values exceeded a medium effect size and most exceeded a large effect size, ranging from  $r = .37$  (Emotional Expression/Independence) to  $r = .81$  (Empathy/Interpersonal Relationships). These results suggest the composite scales and subscales share a relevant underlying factor (i.e., emotional intelligence), similar to that found in the EQ-i 2.0.

### Relationship between the EQ 360 2.0 and the EQ-i 2.0

Associations between self- and other-ratings serve as another source of a scale's validity. The EQ 360 2.0 can be validated by finding a strong level of agreement between EQ 360 2.0 and EQ-i 2.0 scores. In order to assess the association between self and observer ratings, a sample of 108 participants rated themselves on the EQ-i 2.0 and were also rated by a rater on the EQ 360 2.0. Most of the EQ 360 2.0 ratings were provided by family members or spouses (65.7%) or a friend (21.3%). Most (97.2%) of the raters knew the person they were rating for at least one year, with

81.5% of the raters stating that they knew the person they were rating “Very Well” (on a 4-point scale ranging from “Not Very Well” to “Very Well”) and 76.9% stating that they interacted with the person “Very Often” in the past month (on a 4-point scale ranging from “Occasionally” to “Very Often”). A breakdown of the sample (i.e., those who provided self-ratings and were also rated by others) is presented in [Table A.54](#).

### **Correlations Between EQ 360 2.0 And EQ-I 2.0 (Self-Other Agreement)**

The correlation between the EQ-i 2.0 and EQ 360 2.0 Total EI scores was  $r = .60, p < .01$  ([Table A.55](#)). Correlations for the composite scales and subscales were all significant at  $p < .01$ , and almost every correlation reached the criterion for a large effect size. Specifically, the correlations ranged from  $r = .44$  (Stress Tolerance) to  $r = .72$  (Happiness). These results suggest that self-other agreement for the EQ 360 2.0 (and EQ-i 2.0) is strong. Moreover, this pattern suggests that EI as measured by the EQ-i 2.0 and EQ 360 2.0 is a robust trait that is evaluated similarly via self-report and external observers. However, these correlations are not high enough to suggest redundancy; each measure is assessing unique information about the individual and both types of scores provide important information. Specifically, self-ratings will not always align with observer ratings.

### **Comparing Scores On The EQ 360 2.0 And EQ-I 2.0 (Self-Other Consistency)**

To supplement the correlational results between the EQ 360 2.0 and EQ-i 2.0, standard scores were compared between the two measures. The correlations compare the rank order of individuals on the EQ 360 2.0 and EQ-i 2.0. That is, high correlations between the two measures suggest that individuals who are rated as high in EI by observers (EQ 360 2.0) also have high self-report (EQ-i 2.0) scores, and individuals with low EQ 360 2.0 ratings also have low EQ-i 2.0 scores. However, the EQ 360 2.0 and EQ-i 2.0 ratings themselves may be quite different on an absolute level. For example, scores on the EQ 360 2.0 may be dramatically and uniformly lower than EQ-i 2.0 ratings, but as long as the rank-order of the ratings remains similar across the two measures, the correlation between the two will be high. Examining the degree to which EQ 360 2.0 and EQ-i 2.0 standard scores differ will help determine the nature of the relationship between the EQ-i 2.0 and EQ 360 2.0. These analyses also summarize the consistency of scores between self- and other-ratings.

EQ 360 2.0 and EQ-i 2.0 standard scores were compared by calculating a difference score between the two measures, which consisted of subtracting each EQ 360 2.0 standard score from its corresponding EQ-i 2.0 standard score. Therefore, a positive difference represents higher EQ-i 2.0 scores relative to EQ 360 2.0 scores, and a negative difference represents higher EQ 360 2.0 scores relative to EQ-i 2.0 scores. Recall that the criterion for describing a meaningful difference between self- and other-ratings was determined to be 10 standard score points (see [Planning the EQ 360 2.0 Assessment Process](#)). Difference scores are displayed in [Table A.56](#) to summarize the proportion of difference scores that fall above or below 10 standard score points. Just over half of EQ 360 2.0 and EQ-i 2.0 scores fell within 10 points of each other for the Total EI score and all composite scales and subscales. Overall, the results demonstrate a good degree of consistency between EQ 360 2.0 and EQ-i 2.0 scores; however, the fact that large differences are observed for close to half of the sample demonstrates the importance of collecting both self and observer ratings.

### **Associations Among EQ-I 2.0, EQ 360 2.0, And Sas-Sr**

Emotional intelligence tends to show consistent associations with general adjustment. Social adjustment—as measured by the Social Adjustment Scale – Self-Report (SAS – SR, Weissman, 1999)—should therefore show strong associations with the EQ-i 2.0 and EQ 360 2.0. The SAS-SR is a 54-item self-report scale intended to measure “instrumental and expressive role performance” (p. 1) in six major areas of functioning: work (employed, homemaker, or student); social and leisure activities; relationships with extended family; role as a marital partner; parental role; and role within the family unit. Across these six role areas, SAS-SR questions cover four qualitative categories: performance at expected tasks; the amount of friction with people; finer aspects of interpersonal relations; and feelings and satisfactions. Items are rated on a five-point rating scale with higher scores reflecting higher levels of impairment.

The independent associations of the EQ-i 2.0 and EQ 360 2.0 to SAS-SR scores were examined through multiple regression analyses, to shed light on the unique contributions of observer EI ratings in predicting social adjustment over self-report ratings, and vice-versa. The demographic description of the participants in this sample is displayed in [Table A.57](#), and [Table A.58](#) displays the results of the analyses. Correlations with the SAS-SR for both the EQ-i 2.0 and EQ 360 2.0 were mostly strong and in the expected direction (correlations are negative because high SAS-SR scores reflect social maladjustment). Stepwise multiple regression analyses were then performed in two steps. In the first step, only the EQ-i 2.0 scale was entered as a predictor, with SAS-SR scores as the outcome. In the second step, the EQ-i 2.0 scale and the EQ 360 2.0 scale were entered simultaneously. Therefore, it is possible to evaluate the independent associations of each scale with the SAS-SR. This analysis was conducted separately for each composite scale, subscale, and Total EI. For the Total EI score as well as most of the composite scales and subscales, both the EQ-i 2.0 and EQ 360 2.0 scales were independently related to the SAS-SR Total Score at the  $p < .05$  significance level. In other words, self-report and observer ratings were each uniquely informative of SAS-SR scores.

A final set of statistics relevant to these analyses is the  $R^2$  change ([Table A.58](#)). This statistic communicates the amount of explanatory power the EQ 360 2.0 scale adds to the prediction of SAS-SR scores after accounting for its respective EQ-i 2.0 scale. In other words, the incremental validity of the EQ 360 2.0 scores can be quantified. The strongest effects were found for the Empathy and Reality Testing subscales and the Interpersonal composite. Overall, the pattern of results showed expected associations between EI and social adjustment for both the EQ-i 2.0 and the EQ 360 2.0. The EQ-i 2.0 and EQ 360 2.0 subscales and composite scales provided unique and incremental contributions towards social adjustment.

### **Examination of Potential Race/Ethnicity Effects in the Rater-Ratee Relationship in the EQ 360 2.0**

Another important issue related to EQ 360 2.0 ratings is whether a race/ethnicity bias exists. That is, neither the race/ethnicity of the ratee nor the race/ethnicity of the rater should have an effect on EQ 360 2.0 scores. Analysis of covariance (ANCOVA) was used to examine these potential effects in the EQ 360 2.0 Total EI score, using rater race/ethnicity (White vs. non-White) and ratee race/ethnicity (Black vs. Hispanic/Latino vs. White) as independent variables, and ratee gender and age group as covariates. Two separate multivariate analyses of covariance (MANCOVAs) were used to examine the composite scales and subscales. Specifically, a significant or meaningful interaction between the two independent variables would provide

evidence that raters' race/ethnicity is influencing differences in ratings of White, Black and Hispanic/Latino ratees. [Table A.59](#) demonstrates that this was not the case in the EQ 360 2.0 normative sample. The Wilk's lambda values suggested that only a negligible amount of variance could be explained by the interaction between rater and ratee race/ethnicity. The interaction terms were not significant at the  $p < .01$  level, and none of the effect sizes met the minimum requirements for even a small effect size (i.e.,  $\eta^2 = .01$ ). These results illustrate that raters did not show differences in their ratings based on the ethnicity of the ratees.

### **Validity Summary**

Several validity analyses were conducted for the EQ 360 2.0. Support for the scale's factor structure, as identified in the EQ-i 2.0, also emerged in the EQ 360 2.0. The validity of the EQ 360 2.0 was further supported through comparisons with the EQ-i 2.0 (self-other agreement and consistency) and a measure of social adjustment (unique and incremental validity relative to the EQ-i 2.0). There was no evidence of bias in relation to the race/ethnicity of the rater or the ratee. Overall, the analyses suggest the EQ 360 2.0 is a valid measure of EI.