Journal of Emotion and Psychopathology
2025, Vol 1, Issue 2, 440-477
https://doi.org/10.55913/joep.v1i2.51

The Inter-Personal Affect Regulation Test (IPART): Development and Validation of a Performance-Based Assessment of the Ability to Improve the Emotions of Others

Ryan Smith[1*†], Claire A. Lavalley[1†], Michelle R. P. Durham[2,3], Jeff Skalamera[2], Meltem Ozcan[2], Adam Raikes[2], Sarah Berryhill[2], Jacqueline Marquez[2], Courtney Smith[2], Anna Sanova[2], Anna Alkozei[2], & William D. S. Killgore[2]

[1]*Laureate Institute for Brain Research, United States*

[2]*University of Arizona, Department of Psychiatry, United States*

[3]*Western Kentucky University, United States*

Abstract

The ability to intentionally improve the emotional states of others, or inter-personal affect regulation ability, is considered an important component of emotional intelligence. While multiple self-report scales have been developed to assess self-perceived capacities within this domain, few objective performance measures exist. In this article we describe two studies which outline the creation and validation of a performance-based assessment of this ability: The Inter-Personal Affect Regulation Test (IPART). Study 1 demonstrates that the IPART shows good internal reliability and that test items appear to primarily load onto a single latent factor. It also provides evidence of convergent validity by demonstrating significant relationships with existing self-report scales in expected directions. Study 2A replicates results of Study 1 and extends them by showing relationships to performance-based measures. Study 2B examines longitudinal data to show test-retest reliability, assess relationships with change in other measures over time, and demonstrate how IPART scores are appropriately sensitive to emotional intelligence training. The IPART represents a novel, reliable, and valid tool for assessing individual differences in this important social-cognitive skill – including both the ability to recognize optimal responses and how often one in fact chooses those responses.

Keywords: Emotion; Emotion Regulation; Social Skills; Managing Others' Emotions; Emotional Intelligence.

Introduction

Emotional intelligence (EI) is a well-established construct that refers to a set of interconnected cognitive skills or competencies, many of which involve either recognizing or regulating the emotions of self or others (Webb et al., 2013). While performance-based measures currently exist for several of these skills (e.g., the ability to recognize emotions in faces, voices, etc.; (Banziger et al., 2009)), others are more often assessed via self-report. One such skill involves regulating or "managing" the emotions of others. In the last several years, a range of self-report scales have been developed to assess individual differences in this ability, including the Managing the Emotions of Others Scale (MEOS; (Austin & O'Donnell, 2013; Austin et al., 2014)) and the Emotion Regulation of Others and Self Scale (EROS;

†Indicates co-first authorship based on equal contribution.
*Corresponding Author: Ryan Smith ✉ rsmith@laureateinstitute.org
Received: 22 April 2023 | Revision Received: 13 December 2023 | Accepted: 29 December 2023
Published by Black Swan Psychological Assessments Pty Ltd
www.emotionandpsychopathology.org | ISSN: 2653-7583

(Niven et al., 2011)). Existing self-report measures of EI as a whole, such as the Bar-On Emotional Quotient Inventory (EQ-i; (Bar-On, 2002)), the Trait Emotional Intelligence Questionnaire (TEIQue; (Mikolajczak et al., 2007; Petrides et al., 2007), and the Self-Rated Emotional Intelligence Scale (SREIS; (Brackett et al., 2006)) each also contain subscales that pertain to managing others' emotions.

These self-report measures offer an efficient means of assessing general emotion management tendencies, such as the extent to which people aim to improve or worsen the emotions of others, and whether they do so based on selfish or prosocial motives. Yet, this measurement approach also has limitations. For example, differences in factors such as imagination (Nozaki & Mikolajczak, 2020), prosocial motivation (Austin & O'Donnell, 2013), and biased self-perception (Robinson & Sedikides, 2020) can influence ratings. In addition, the generic nature of many test items (e.g., "If someone has a problem I offer to help if they need it") can preclude incorporation of moderating contextual details (Aldao & Nolen-Hoeksema, 2012). Differences in short- vs. long-term efficacy are also not assessed. For example, the MEOS subscale assessing distraction-based (Diversion) strategies is associated with tendencies to improve emotions (i.e., the MEOS Enhance subscale; r = .67) and agreeableness in personality (r = .35), and such strategies can be effective in some cases (Austin et al., 2014). However, diversion is short-term in focus and often does not address the underlying source of negative affect. This differs from situation- or cognition-focused approaches, such as empathic acceptance, reappraisal, and collaborative problem-solving, which more directly address the problem and focus on long-term improvement (Gross, 2015). If one aims to assess the relative use of such concrete emotion management strategies, other approaches may therefore be required.

Complementary performance-based instruments designed to incorporate situational detail and response specificity could therefore be useful in studying interpersonal emotion management. To date, however, this approach has received less attention. Existing work in this area has been motivated by the "ability model" of EI (Mayer et al., 2016), which focuses on testing responses that can be evaluated for accuracy. For example, the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; (Mayer et al., 2003)) includes 3 test items related to managing others' emotions (i.e., 3 social scenarios described with multiple response options). These are combined with other self-directed emotion regulation test items to derive scores for a broader "managing emotions" subscale. However, various limitations of the MSCEIT have also been highlighted in prior work (Petrides, 2011).

To our knowledge, only one full scale performance-based measure has been created to assess interpersonal emotion management – the Situational Test of Emotion Management (STEM; (Allen et al., 2015)). The STEM presents written descriptions of several emotional situations and asks individuals to select the course of action in each situation, out of four response options, that would be most effective at managing "both the emotions the person is feeling and the problems they face in that situation." The correct answer for each item is based on expert consensus. Higher scores therefore reflect how well individuals can recognize the response that experts agree would be most effective out of the options given. Crucially, individuals are explicitly told that "you are not necessarily choosing what you would do, or the nicest thing to do, but choosing the most effective response for that situation." Based on this instruction, a person could obtain a high score on the STEM, for example, even if they did not believe they would choose the most effective response if placed in that situation.

To our knowledge, no performance-based measure currently exists to assess whether individuals would themselves choose effective courses of action for managing others' emotions in difficult social situations, or how this might diverge from their ability to recognize them. However, similar approaches have shown promise in the literature on Situational Judgement Tests (SJTs) – often used to assess maximal vs. typical performance in career-related situations (Lievens et al., 2008). These tests, which effectively simulate work environments and require the test-taker to make behavioral decisions in context, have shown significant associations with actual workplace performance (McDaniel et al., 2007). Relative to some self-report instruments, they also demonstrate less susceptibility to effects of social desirability (Hooper et al., 2006), and socially desirable responses are only weakly associated with optimal responses (Kaminski et al., 2019). The affective/clinical sciences could benefit from this

approach, as choices made in such contexts could impact how well individuals navigate socio-emotional situations. Given the above considerations, we sought to develop and validate a complementary measure to assess both an individual's ability to identify effective ways of managing others' emotions when presented with specific situations and whether they would themselves adopt those same courses of action: the Inter-Personal Affect Regulation Test (IPART). Below we describe a series of studies carried out to develop and validate the IPART.

Development of this test was also motivated, in part, by the fact that self-report and performance-based assessments tend to show weak correlations – both in general (Dunning et al., 2004) and for EI in particular (Maroti et al., 2018; Webb et al., 2013). It therefore remains unclear whether existing self-report measures of emotion management predict concrete strategy selection. It is possible, for example, that beliefs about one's own emotion management abilities influence social decision-making in a distinct manner from the objective ability to identify effective responses (i.e., both could be relevant to mental health and well-being in different ways; e.g., see Berking and Wupperman (2012) and Hu et al. (2014)). This is consistent with the idea that self-rating scales of EI assess constructs associated with personality (Matthews et al., 2004), while ability-based EI measures are more closely associated with cognitive performance (O'Connor Jr & Little, 2003). Developing the IPART therefore also allowed us to assess how strongly performance was associated with existing self-report measures of emotion management. Thus, after the initial development stage, Study 1 assessed the IPART's relationships to leading self-report measures. Study 2A sought to replicate the results of Study 1 and then further assessed relationships to several performance-based measures (including the STEM). Study 2B analyzed longitudinal data to: 1) confirm test-retest reliability, 2) assess relationships to changes over time in other measures, and 3) examine whether IPART scores are responsive to emotional intelligence training.

Study 1

Methods

Participants

Three-hundred United States residents were recruited via Amazon Mechanical Turk (MTurk) to complete the IPART and other study measures remotely. Participants were required to be 18 to 35 years old, speak English as their primary language and demonstrate an 8th grade reading level or better, and not have any history of neurological illness or other conditions that could impact performance (e.g., those related to substances, medications, or injury). After removing individuals who incorrectly answered two or more attention check questions (N = 38), or otherwise showed signs of careless responding (N = 1), the final sample included 261 participants (M_{age} = 28.78, SD = 3.7; 49.2% female; 65.5% White). Participants were paid eight dollars for their time.

Procedure

Assessments

Participants were first given an initial demographic questionnaire where they provided information about age, sex, ethnicity, education level, and income level. Next, they were asked to complete the IPART, followed by several existing self-report measures:

Managing the Emotions of Others Scale (MEOS)

The MEOS is a 58-item self-report measure of the typical strategies a person uses to manage the emotions of others through mood-improving and mood-worsening behaviors. The measure also encompasses prosocial and non-prosocial aspects of managing the emotions of others. Scores on the MEOS show good consistency across time, and the internal reliabilities (Cronbach's α) for the MEOS factors range from .68 to .91 (Austin & O'Donnell, 2013). For the purposes of this study, we were primarily interested in the "Enhance" and "Worsen" subscales, reflecting tendencies to improve (e.g., providing understanding, help, reassurance) or worsen (e.g., criticism) the emotions of others,

respectively. However, for the interested reader, supplementary analyses are also reported for the remaining subscales. In the present sample for Study 1, internal consistency for the MEOS covered a similar range of values (α = .66 to .93).

Emotion Regulation of Others and Self Scale (EROS)

The EROS is a self-report measure of the strategies one uses to improve and worsen the emotions of self and others. It has been shown to have acceptable internal reliability for each of its factors (Cronbach's α from .74 to .82; (Niven et al., 2011)). For the purposes of this study, we were primarily interested in the "Extrinsic Affect Improving" and "Extrinsic Affect Worsening" subscales, reflecting the selection of behaviors that would improve or worsen the mood of others. However, we also examined the "Intrinsic Affect Improving" and "Intrinsic Affect Worsening" subscales, which reflect the selection of behaviors that would improve or worsen one's own mood. These further subscales might be expected to correlate with the IPART (positively and negatively, respectively) if people perhaps tend to use similar strategies for regulating the emotions of self and others. Internal consistency results in the Study 1 sample were good (α = .80 to .92).

Self-Rated Emotional Intelligence Scale (SREIS)

The SREIS is a 19-item self-report measure of emotional intelligence that evaluates one's abilities to perceive, understand, and manage the emotions of self and others, and to use emotions in decision making. The full scale has been shown to have poor to good reliability, with Cronbach's α ranging from .55 to .85 across multiple studies (Brackett et al., 2006). For this study, we were primarily interested in the SREIS "Social Management" subscale, which measures the self-reported ability to effectively manage others' emotions. However, we also examined the other SREIS scales (i.e., reflecting the ability to perceive emotions, to understand emotions, to manage one's own emotions, and to use emotions in decision-making), as some of these other abilities could plausibly contribute to inter-personal affect regulation ability. In the sample for Study 1, internal consistency was good (α = .87).

Emotion Regulation Questionnaire (ERQ)

The ERQ (Gross & John, 2003) is a 10-item scale measuring intrapersonal emotion regulation tendencies; namely, the tendency to regulate one's own emotions through cognitive reappraisal or suppression. The ERQ has demonstrated good to excellent internal reliability (Cronbach's α for reappraisal: .89 to .90; Cronbach's α for suppression: .76 to .80; (Preece et al., 2020)). For the purposes of this study, we were interested in examining whether higher IPART scores predicted more frequent use of adaptive strategies (i.e., reappraisal) and less frequent use of maladaptive strategies (i.e., suppression). In the sample for Study 1, Cronbach's α demonstrated good consistency for reappraisal (α = .89) and suppression (α = .79).

IPART Development

Item Generation

The initial strategy for developing the IPART included the creation of an item-pool of 42 written social scenarios. These items were developed to capture a range of socially difficult situations in which someone has an explicit goal of helping another individual feel better after some kind of hardship (e.g., losing a job, losing a loved one, ending a long-term relationship, etc.). In a few sentences, each item describes a situation and the thoughts and feelings of an individual within it. It then states that the goal is to choose a response that will best help that individual feel better about the situation. Four response options are then provided, each of which is a potential verbal response. Individuals are asked to provide two answers to each item. First, they are asked to indicate what they believe the best response would be (the "Best Choice" [BC] response). This question assesses an individual's ability to *recognize* responses that would be most effective on average given the situation and the individual's current thoughts and feelings. Second, they are asked to indicate which response they would actually choose if they were in the situation (the "Personal Choice" [PC] response). This question was included to assess an individual's tendency to respond in more or less effective ways, independent of their ability to recognize the best

response. PC scores therefore reflect a type of self-report about anticipated behavior in specific situations, but also differ from other self-report measures in important ways (i.e., suggesting they may carry novel information about an individual). First, they pertain to context-specific responses in multiple-choice format, as opposed to rating levels of agreement with broad statements. Second, they are directly contrasted with BC responses, meaning that individuals assess whether they would choose the option they previously indicated was best.

The four verbal response options for each item were generated based on current empirical and clinical findings regarding the relative effectiveness of different emotion regulation strategies (Barlow et al., 2016; Gross, 1998; Gross & John, 2003; Gross & Levenson, 1997; Shiota & Kalat, 2012; Webb et al., 2012). Particular approaches (e.g., empathizing with and legitimizing concerns, offering encouragement) were motivated in part by prior work demonstrating the benefits and efficacy of emotional and social support (Burleson, 2003; Dunkel Schetter & Brooks, 2009; Reblin & Uchino, 2008). Specifically, the best response option (worth 2 points) always conveyed nonjudgmental validation and then focused on constructive reappraisal and/or collaborative problem solving (we refer to this as the "Improve" option below). The two intermediate response options (worth 1 point) were less validating, and instead focused primarily on suggesting avoidant strategies (e.g., ignoring the problem, suppressing emotions, finding a distraction, etc.), which tend to be less effective (especially in the long-term; we refer to these as the "Divert" options below). Importantly, the stated goal in the IPART is to help the individual feel better "about the situation", which is relevant here because Divert options focus on temporarily avoiding the situation, as opposed to addressing it productively to help them feel better about it (but note that, in an effort to avoid being overly leading, and to get at the "default" strategy an individual would tend to identify, we chose not to include more explicit statements about timeframe [e.g., "in the long-term"]). Finally, the worst response option (worth 0 points) involved judgmental and/or otherwise insensitive statements that would tend to make a person feel worse (we refer to this as the "Worsen" option below). Here is an example item:

Situation: A group of Felix's friends played a prank on him. The prank unintentionally ruined some of his belongings. His friends paid for the damage, but Felix is still upset.

Felix's thoughts and feelings: Felix is agitated. He thinks his friends should have known that he wouldn't find the prank funny. He feels disrespected and thinks that no one understands that it's more than just the damage he's upset about.

Goal: Say something to Felix to improve how he feels about the situation.

Responses:
a. "Well, if they've paid you back then there's nothing else you can do about it. Let's go hang out with some other friends instead to take your mind off of it." *(1 point; i.e., avoidant, no validation)*

b. "I agree. That was really disrespectful. It sounds like they're trying to fix things by paying you. Have you told them why you're still upset?" *(2 points; i.e., validating, constructive)*

c. "Take it easy. They paid for the things they broke so what else could they possibly do? I think you're being too hard on them." *(0 points; invalidating, judgmental)*

d. "I would just take the money and let this all blow over. They probably didn't mean to hurt you or break your things with the prank." *(1 point; avoidant, no validation)*

Which of the responses above do you think would be the most effective response for achieving the stated goal? *(Best Choice Question)*

Which of the responses above are you most likely to actually choose if you were in the situation? *(Personal Choice Question)*

Separate BC and PC Total scores were derived by summing point values across all items for the responses chosen on the Best Choice and Personal Choice questions, respectively. For additional insights regarding whether higher total scores on the final measure were better accounted for by a greater number of "Improve" choices or a lower number of "Worsen" choices (and as a secondary measure of people's specific response tendencies), we also calculated the total number of items for which the "Improve", "Divert", and "Worsen" response types were selected.

Statistical Analysis

Reliability and Factor Structure Analyses

Internal consistency of the initial pool of potential IPART items (described below) was assessed using Cronbach's α. Following heuristics for good inter-item correlations (Piedmont, 2014), items with correlation coefficients lower than .20 were removed. We also examined item-test correlations for each item, and how Cronbach's α would change with the removal of different items. Eight items were eliminated through this process to arrive at a final test with the highest possible level of internal reliability in the current sample. After identifying the final set of items to be included, we examined its split-half reliability (*splitHalf* function in R; *psych* package (v2.2.9; (Revelle, 2022)) using Guttman's λ^2 (Guttman, 1945).

For both BC and PC scores, exploratory factor analyses (EFA) were conducted (with *factanal* function in the R stats package (R Core Team, 2022)) following the suggestions of Howard (2016). We used the *promax* rotation argument to allow factors to be correlated. These analyses were conducted using R 4.2.0 and are described in further detail within the results. As preliminary steps, we also first assessed whether relationships between the final set of IPART items were strong enough to perform an EFA using the *KMO* function (also in the *psych* package) and confirmed whether a one-factor solution was sufficient, as expected (based on recommendations in (Costello & Osborne, 2005)).

Convergent Validity

We first checked the reliability of each measure. Then, to assess convergent validity, we examined correlations between IPART scores and each of the self-report measures (of intra- and inter-personal affect regulation) listed above. A complete-linkage clustering algorithm was also used to determine the structure of these associations. The optimal number of clusters (based on average silhouette width) was identified using the *fviz_nbclust* function in R (v1.0.7; factoextra package; (Kassambara & Mundt, 2020)). We expected that higher IPART scores would be positively correlated with the MEOS-Enhance, EROS-Extrinsic Affect Improving, and SREIS-Social Management subscales, and negatively correlated with the other MEOS subscales (i.e., Worsen, Conceal, Inauthentic, Divert, and Poor Skills) and EROS-Extrinsic Affect Worsening. Under the assumption that emotion regulation skills directed at the self might overlap with those applied to others, we also examined whether self-reported tendencies to improve (i.e., EROS-Intrinsic Affect Improving, ERQ-Reappraisal) or worsen (i.e., EROS-Intrinsic Affect Worsening, ERQ-Suppression) one's own emotions might also be positively and negatively associated with IPART scores, respectively. Next, given that other aspects of emotional intelligence, such as the ability to perceive, understand, and use emotions, might also indirectly contribute to emotion management ability, we analyzed correlations between IPART scores and the other SREIS subscales. Post-hoc power analyses were also conducted in R using *pwr* (Champely, 2020; Faul et al., 2007) for correlations to ensure enough statistical power to draw conclusions from tests.

Given previously observed sex differences in scores on a range of other performance-based socio-emotional measures (e.g., see (Hall & Matsumoto, 2004; Wright et al., 2017)), we also expected that female participants would tend to outperform male participants on IPART measures, which was evaluated using two-sample t-tests.

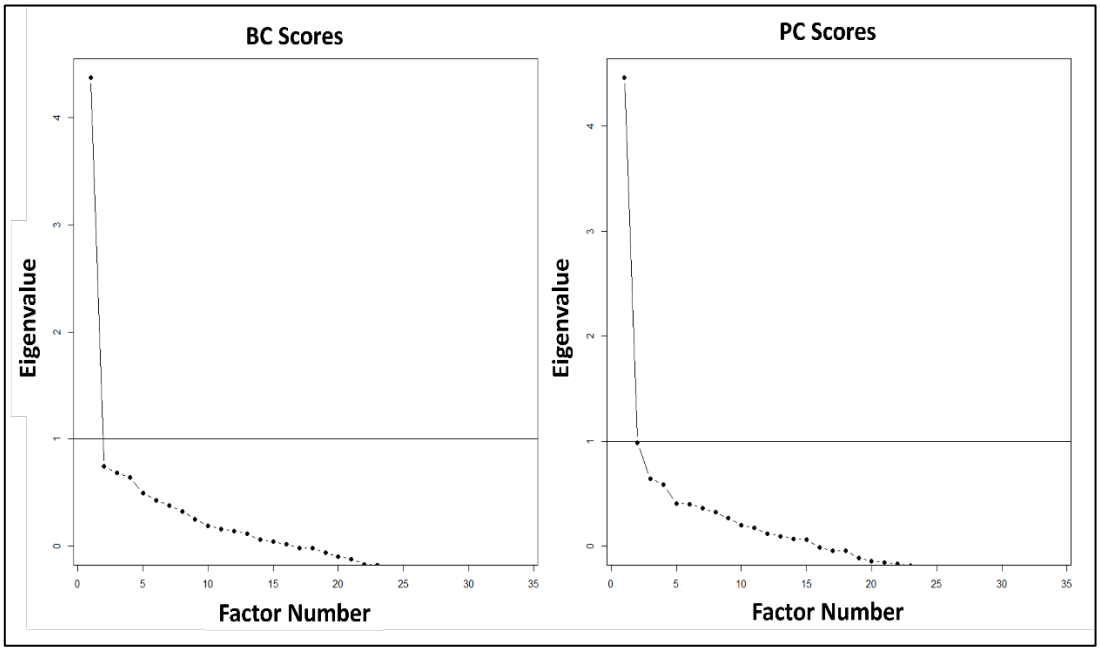
Results

IPART Reliability, Item Selection, and Factor Structure

A 34-item version of the IPART was found to have good internal consistency (Cronbach's α = .82 for BC scores; Cronbach's α = .82 for PC scores). A subsequent split-half reliability analysis also indicated good reliability for this version (Guttman's λ^2 = .83 for BC scores; Guttman's λ^2 = .83 for PC scores).

Based on recommendations for measures of sampling adequacy (MSA) by Kaiser and Rice (1974), preliminary factor analytic results indicated adequate structure for both BC (MSA = .77) and PC scores (MSA = .80). A one-factor solution was found to be sufficient in both cases ($ps \leq .001$). Specifically, BC scores loaded onto a single factor that explained 12.9% of the common variance, while PC scores loaded onto a single factor that explained 13.1% of the common variance (see **Figure 1**).

Figure 1. Scree plots of IPART items separated by BC and PC scores, indicating the eigenvalues associated with factors identified in an exploratory factor analysis (EFA).



Preliminary Analyses

Descriptive statistics for the IPART and all other measures are reported in **Table 1**. Both IPART BC and PC scores were normally distributed (skewness < |2|) (Byrne, 2016; Hair Jr et al., 2010). Cronbach's α demonstrated good internal consistency for both BC and PC (.82 in both cases). As hypothesized, women scored significantly higher than men on both BC and PC Total scores (BC: $t(259) = -4.14$, $p < .001$; PC: $t(259) = -4.38$, $p < .001$). There were no significant correlations between age and either BC or PC scores ($r_{BC} = .09$ and $r_{PC} = -.03$), and, while BC Total scores correlated positively with education level (Spearman's *rho* = .14, $p = .026$), PC Total did not (*rho* = .09, $p = .147$). Additionally, a paired t-test showed that BC Total scores were significantly higher than PC Total scores ($t(260) = 10.83$, $p < .001$), indicating that individuals often chose personal responses other than the ones they believed were best. For the interested reader, **Table 1** also compares male and female participants on other study measures (these post-hoc comparisons were not a priori hypotheses about IPART scores and were not corrected for multiple comparisons).

Table 1. Descriptive statistics of measures in Study 1 (*N* = 261)

Measure	Female	Male	Statistics
IPART			
BC Total	53.74 (7.90)	49.73 (7.73)	$t(259) = -4.14$, $\mathbf{p < .001}^{***}$
PC Total	49.09 (9.05)	44.52 (7.74)	$t(259) = -4.39$, $\mathbf{p < .001}^{***}$
BC Improve	22.29 (6.00)	19.09 (5.82)	$t(259) = -4.37$, $\mathbf{p < .001}^{***}$
PC Improve	18.78 (6.64)	15.53 (5.61)	$t(259) = -4.27$, $\mathbf{p < .001}^{***}$
BC Divert	9.16 (4.63)	11.55 (4.57)	$t(259) = 4.19$, $\mathbf{p < .001}^{***}$
PC Divert	11.54 (4.82)	14.46 (4.47)	$t(259) = 3.34$, $\mathbf{p < .001}^{***}$
BC Worsen	2.55 (2.44)	3.36 (2.55)	$t(259) = 2.63$, $\mathbf{p = .009}^{**}$
PC Worsen	3.69 (2.90)	5.02 (2.91)	$t(259) = 3.69$, $\mathbf{p < .001}^{***}$
SREIS			
Total Score	70.09 (10.07)	68.41 (10.56)	$t(259) = -1.31$, $p = .191$
Perceiving Emotions	15.74 (2.49)	15.07 (2.82)	$t(259) = -2.06$, $\mathbf{p = .041}^{*}$
Use of Emotions	10.05 (1.41)	9.46 (1.47)	$t(259) = -3.32$, $\mathbf{p = .001}^{**}$
Understanding Emotions	14.40 (3.64)	14.02 (3.76)	$t(259) = -0.83$, $p = .407$
Social Management	15.54 (2.94)	14.74 (3.08)	$t(259) = -2.15$, $\mathbf{p = .033}^{*}$
Managing Emotions (Self)	14.35 (3.38)	15.12 (3.10)	$t(259) = 1.93$, $p = .055$
EROS			
Extrinsic Affect Improving	3.89 (0.87)	3.59 (0.93)	$t(259) = -2.71$, $\mathbf{p = .007}^{**}$
Extrinsic Affect Worsening	1.58 (0.75)	1.62 (0.90)	$t(259) = 0.41$, $p = .681$
Intrinsic Affect Improving	3.54 (0.91)	3.32 (0.97)	$t(259) = -1.92$, $p = .056$
Intrinsic Affect Worsening	1.57 (0.76)	1.68 (0.86)	$t(259) = 1.16$, $p = .248$
MEOS			
Enhance	62.00 (10.33)	60.36 (8.80)	$t(259) = -1.38$, $p = .169$
Worsen	23.92 (9.08)	26.84 (9.63)	$t(259) = 2.53$, $\mathbf{p = .012}^{*}$
Conceal	22.02 (5.18)	22.00 (5.15)	$t(259) = -0.02$, $p = .981$
Inauthentic	26.72 (8.02)	27.34 (7.89)	$t(259) = 0.63$, $p = .529$
Divert	24.02 (4.17)	22.63 (3.67)	$t(259) = -2.86$, $\mathbf{p = .005}^{**}$
Poor Skills	10.95 (3.68)	11.35 (3.91)	$t(259) = 0.86$, $p = .392$
ERQ			
Reappraisal	31.75 (7.21)	30.24 (7.26)	$t(259) = -1.69$, $p = .092$
Suppression	14.71 (5.35)	16.14 (5.58)	$t(259) = 2.53$, $\mathbf{p = .036}^{*}$

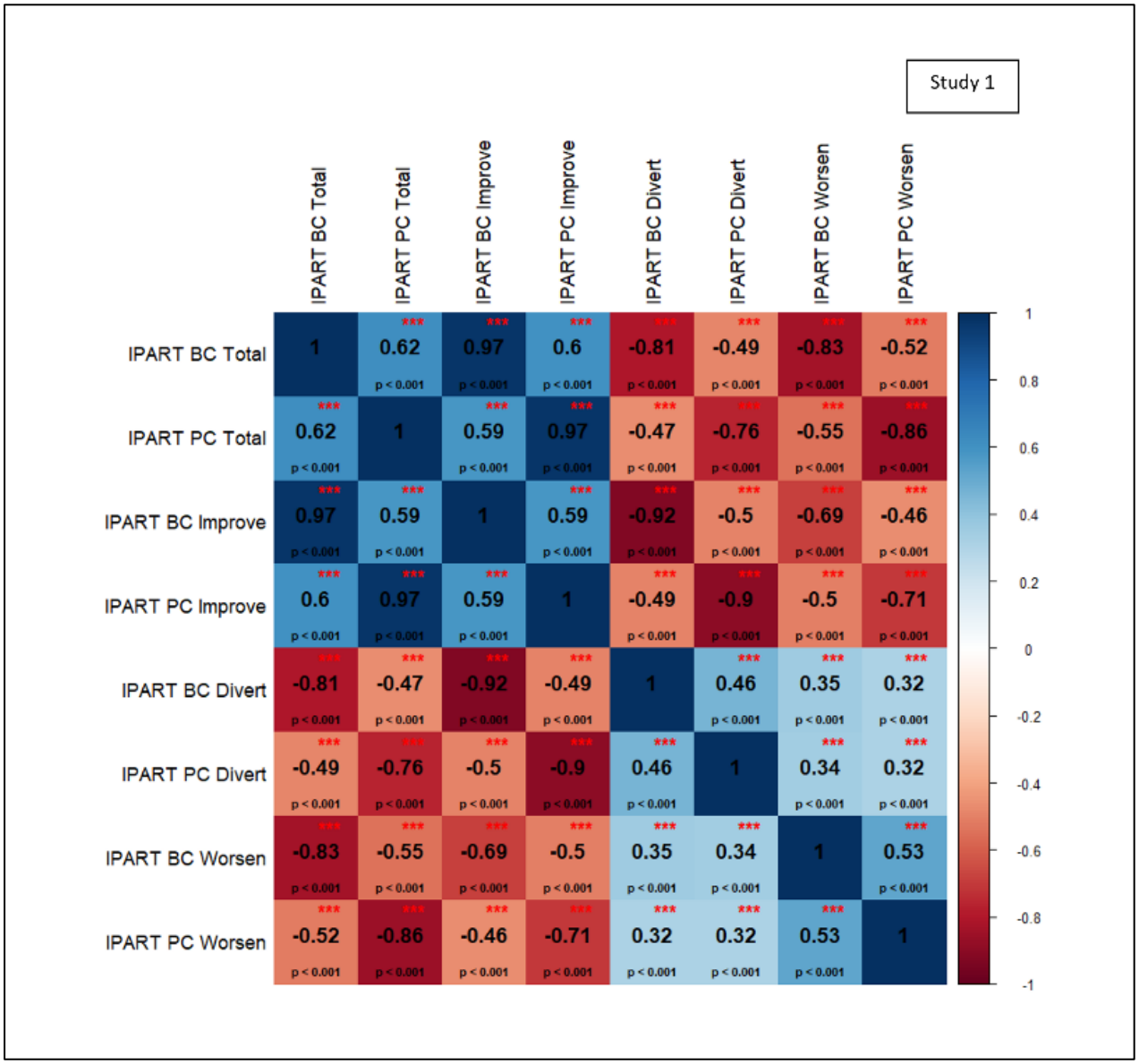
Note. *p<.05, **p<.01, ***p<.001

Convergent Validity Analyses

Inter-correlations between IPART Total and subscale scores are shown in **Figure 2**. BC and PC Total scores were significantly correlated (r = .62, p < .001), yet appeared distinct (i.e., only 38% shared variance). Relationships between IPART scores and the self-report measures in Study 1 are shown in **Figure 3**. Assuming a threshold of 0.8, post-hoc power analyses indicated that a sample size of N = 261 would allow detection of correlations with $r \geq .173$. In **Figure 3**, measures shown closest to the IPART scales (MEOS-Worsen, MEOS-Inauthentic, EROS-Extrinsic Affect Worsening, and EROS-Intrinsic Affect Worsening) clustered with the IPART in the optimal 2-cluster complete-linkage solution, while all other measures formed a second cluster. Consistent with several of our hypotheses, BC scores showed: 1) significantly positive relationships with the MEOS-Enhance and SREIS-Social Management subscales (as well as with the SREIS-Perceiving Emotions subscale and SREIS-Total scores), and 2) significant negative correlations with the MEOS-Worsen and EROS-Extrinsic Affective Worsening (as well as with the EROS-Intrinsic Affective Worsening and ERQ-Suppression subscales). Contrary to hypotheses, BC scores were not correlated with EROS-Extrinsic or Intrinsic Affect Improving, ERQ-Reappraisal, or other SREIS subscales. PC scores showed all the same significant

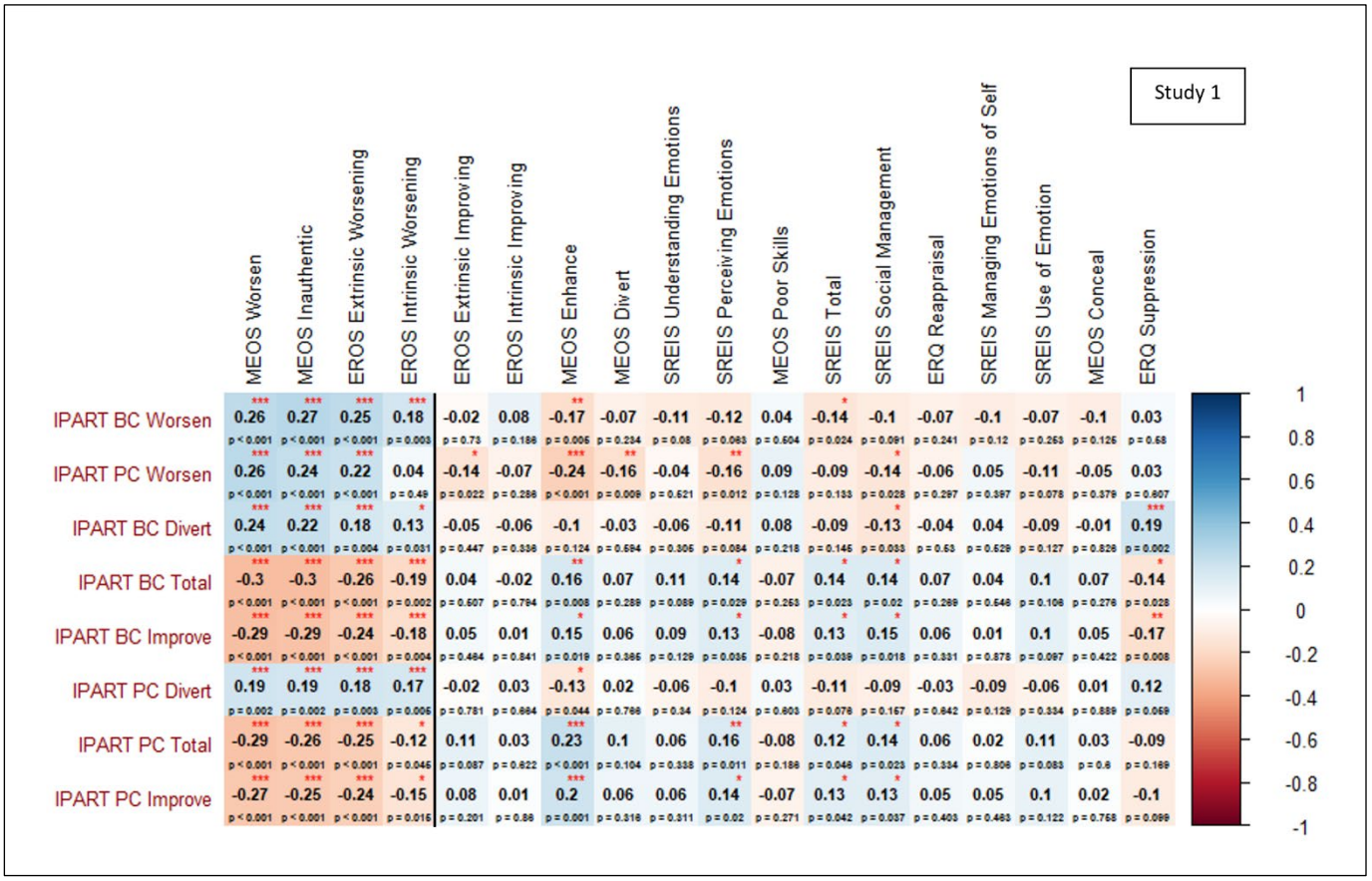
relationships (apart from that with the ERQ-Suppression subscale). Interestingly, MEOS-Inauthentic also correlated negatively with BC and PC Total and Improve scores, and positively with BC and PC Divert and Worsen, while MEOS-Divert was only correlated with PC Worsen scores.

Figure 2. Inter-correlations between IPART Total and subscale scores in Study 1.



Note. *p<.05,**p<.01,***p<.001

Figure 3. Notable correlations between IPART and self-report measures in Study 1. The vertical black line separates the 2 clusters of measures in the optimal clustering solution.



Note. **p*<.05,***p*<.01,****p*<.001

Also noteworthy was the finding that IPART Divert and Worsen sub-scores were positively correlated (r_{BC} = .36, r_{PC} = .32), while both were negatively correlated with IPART Improve sub-scores (Divert: r_{BC} = -.92 and r_{PC} = -.90; Worsen: r_{BC} = -.69 and r_{PC} = -.71), suggesting that the optimal Improve choice options were distinct from the two types of suboptimal choice options, and that Divert options were more often chosen in place of Improve options compared to Worsen options (i.e., they were more strongly negatively correlated).

Summary of Study 1

In summary, a 34-item version of the IPART showed good internal consistency and split-half reliability, and all items loaded onto a single latent factor. Correlations between IPART measures suggested BC and PC scores were not redundant (62% unshared variance). The two suboptimal response options (Divert and Worsen) were positively correlated (and negatively correlated with the Improve option), consistent with them being distinct from the best option. People also more often chose Divert options in place of Improve options (i.e., these two responses had a higher negative correlation than the Worsen and Improve options), as would be expected given a genuine desire to make someone feel better, when combined with less adaptive, avoidant emotion regulation strategies. Many, but not all, hypothesized relationships were found. Relatively speaking, IPART scores showed the strongest (negative) relationships with self-report scales reflecting how much people tend to worsen the emotions of others or show other antisocial tendencies (MEOS-Worsen and Inauthentic, and EROS-Extrinsic Worsening). The consistent pattern of significant correlations between measures shown in **Figure 3** was supportive

of convergent validity. However, these relationships tended to be weak, suggesting the IPART is not redundant with these other measures.

Study 2A

Study 2A examined a new sample of cross-sectional data to accomplish the following aims:

1) Replicate relationships with measures found in Study 1.
2) Examine relationships with additional self-report and performance-based measures not available in Study 1.

Methods

Participants

Participants for Study 2 were drawn from a sample of 448 individuals recruited into a larger study testing the efficacy of an emotional intelligence training program (for details, see (Durham et al., 2023; Persich et al., 2021; R. Smith et al., 2023)). Participants were required to be 18-40 years of age, read at or above an eighth-grade level, and have no neurological, medical, or substance-related conditions that could impact performance on tasks. Recruitment was conducted at the University of Arizona and the surrounding community of Tucson, Arizona (M_{age} = 23.70, *SD* = 5.58, 72.10% female, 60.9% White; 313 university students, 135 community members). Only 427 of these participants had IPART data due to logistical issues (M_{age} = 23.67, *SD* = 5.56, 72.37% female, 60.9% White); therefore, all analyses below were restricted to this subset. No additional participants were removed from data-quality checks similar to those performed in Study 1 (e.g., checking for repeated or alternating response patterns throughout a survey). **Supplemental Table S1** contains demographic and measure information separated by those with and without (*N* = 21) IPART scores. The only significant differences between the two samples were in STEM-B, LEAS Total, and LEAS Other scores, which were higher in those who completed the IPART; this may therefore limit inference regarding those with the lowest scores on these variables.

Procedure

Participants completed the same self-report measures reported from Study 1 (internal consistencies in Study 2A were comparable to those in Study 1: Cronbach's α = .55 to .89). In addition, they completed the following self-report and performance-based measures of related socio-emotional skills, as well as general cognitive measures to assess whether IPART scores are associated with greater general cognitive and reflective capacities. We reasoned that if greater cognitive abilities and general reflectiveness tendencies were positively associated with the frequency of "Improve" choices on the IPART, this would provide added validation for assigning these responses the highest scores. A 30-minute lunch break was taken at the halfway point of the day and two additional breaks (5 minutes on average) were provided between different survey blocks. Overall, responding to all surveys took around 3.5 hours across participants.

Self-Report Measures

Coping Inventory for Stressful Situations: Adult

Developed by Endler & Parker (1990), the CISS is a 48-item measure intended to assess the coping style of an individual at three different levels: emotion-oriented, task-oriented, and avoidant coping. Each scale of this measure has shown adequate to good reliability (α = .66 to .84) (McWilliams et al., 2003). Inclusion of this assessment in our analyses was motivated by positive relationships previously demonstrated between coping styles, stress management, and emotional intelligence (Moradi et al., 2011; Noorbakhsh et al., 2010), and because specific coping styles are associated with IPART response types (e.g., avoidant coping and IPART Divert responses). The CISS therefore offers tests of both convergent and face validity. Internal consistency for this measure in the Study 2A sample was good to excellent (α = .80 to .91).

Trait Emotional Intelligence Questionnaire (TEIQue)

The TEIQue is a 153-item self-report measure that assesses 15 facets, four factors, and one total score of emotional intelligence (EI). The TEIQue has demonstrated poor to excellent internal reliability across all subscales (α = .59 to .91) (Mikolajczak et al., 2007; Petrides et al., 2007). We expected that higher scores on the TEIQue would be positively related to IPART scores, as high-EI individuals would be expected to recognize the best response to emotionally complex situations. Cronbach's α for this measure in Study 2A showed acceptable internal consistency (α = .74).

Multidimensional Assessment of Interoceptive Awareness (MAIA)

The MAIA (version 2) is a self-report measure with 32 items that assess eight aspects of internal/emotional awareness (Mehling et al., 2012). The scale (and subscales) demonstrated nearly acceptable to good internal consistency at initial testing (α = .66 to .82). We expected IPART scores would be positively associated with MAIA emotional awareness scores. In the Study 2A sample, internal consistency was good (α = .82).

Difficulties in Emotion Regulation Scale (DERS)

The DERS is a self-report scale with six subscales and 36 items. Higher scores indicate greater emotion dysregulation (Gratz & Roemer, 2004). Previous work has shown high internal reliability (α = .93) for this measure, significant associations with other self-report measures in expected directions, and good test-retest reliability over a period of 4-8 weeks. The DERS was expected to show negative associations with the IPART, under the assumption that those with greater inter-personal affect regulation abilities would also show less emotion dysregulation. Internal consistency for this measure in the Study 2A sample was excellent (Cronbach's α = .95).

Short Dark Triad scale (SD3)

As the IPART explicitly measures prosocial tendencies, we hypothesized that it should show inverse relationships with antisocial tendencies. To assess this, we included the SD3, a commonly used measure of three antisocial traits: Psychopathy, Narcissism, and Machiavellianism (Jones & Paulhus, 2014). Each subscale has 9 items and demonstrated acceptable internal reliability (α's from .71 to .77; Jones & Paulhus, 2014). In the present sample for Study 2A, Cronbach's α demonstrated good internal consistency (α = .80).

Symptom and Well-Being Measures

Beck Depression Inventory (BDI-II)

The BDI-II is a 21-item scale that measures symptoms of depression (Beck et al., 1996). Each item is scored from 0 to 3, with higher scores consistent with depressive symptomatology. Prior work has demonstrated a relationship between emotion recognition abilities and depression (Collin et al., 2013) as well as vulnerabilities for depression in adolescents with emotion recognition deficits (Nyquist & Luebbe, 2020). Therefore, IPART scores were anticipated to be negatively related to BDI scores.

State-Trait Anxiety Inventory (STAI)

The STAI (Spielberger et al., 1970) is a 40-item scale that measures anxiety symptoms in the present moment (state) and in life more generally (trait). Because deficits in emotional awareness abilities have been found in depression and anxiety disorders (Kranzler et al., 2016), STAI scores were expected to correlate negatively with IPART (particularly BC) scores.

Flourishing Scale (FS)

The Flourishing Scale is an 8-item scale where each item is a positively framed statement about life functioning (e.g., relationships, feelings of competence, purpose). Participants respond on a 7-point Likert scale how much they agree with each statement and receive an overall score signifying how positively they view themselves in these areas deemed important for life functioning. This measure demonstrated very good reliability (α = .87) and good temporal stability (Diener et al., 2010). Because flourishing can be a measure of life-satisfaction and well-being, and this measure explicitly asks about

relationship satisfaction, scores were expected to correlate positively with IPART BC and PC Total and Improve scores. We reasoned that people who can not only recognize the best choice answer but also choose that option in real social situations may have more positive relationships and greater Flourishing scores. Cronbach's α in the Study 2A subsample with Flourishing scores (n = 231) demonstrated excellent internal consistency (α = .90).

Performance-Based Socio-Emotional Measures

Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

This 141-item performance-based scale assesses four branches of emotional intelligence: perceiving emotions, facilitating thought, understanding emotions, and managing emotions (Mayer et al., 2002). Previous analyses have demonstrated acceptable to excellent reliability (α = .76 to .91) for the overall MSCEIT as well as its subscales (Mayer et al., 2003). Given that the MSCEIT is a performance-based metric, we expected that relevant components of this test might show stronger positive relationships with the IPART than those observed with self-report measures. In particular, we expected stronger relationships with its Social Management subscale. However, all sub-scales scores were included due to plausible interactions between socio-emotional skills (e.g., better inter-personal affect regulation could benefit from better emotion recognition skills, etc.). Cronbach's α in Study 2A demonstrated good internal consistency (α = .83).

Situational Test of Emotional Management: Brief (STEM-B)

This short-form performance-based measure of inter-personal affect regulation has 18 items (Allen et al., 2015) and was developed based on the 44-item STEM (MacCann & Roberts, 2008). The STEM-B has good internal reliability (α = .84) and was included in our analyses because it is conceptually closest to the construct assessed by the IPART (Allen et al., 2015). As described above, the STEM-B assesses the ability to recognize the most effective response in a multiple-choice format. So, we expected positive relationships between STEM-B and IPART BC scores in particular, and we were interested to see how this might differ when compared to IPART PC scores. Internal consistency for this measure in the Study 2A sample was very poor (Cronbach's α = .46).

Levels of Emotional Awareness Scale (LEAS)

This online assessment (https://eleastest.net/) of emotional awareness asks participants to imagine and describe how they think they and another person would feel in response to 10 hypothetical social scenarios (Barchard et al., 2010; Lane et al., 1990). Responses are typed into text boxes and scored automatically by computer software that assesses the specificity of language used in the descriptions (e.g., higher scores are given for use of specific emotion terms like "sad" than coarser-grained valence terms like "bad"). The 10-item version of this measure has shown acceptable internal reliability (α = .78; (Lane & Smith, 2021)). This performance-based metric was expected to positively correlate with IPART scores as people with higher emotional awareness appear to reason more thoroughly/specifically about the emotions individuals would feel in socially/emotionally difficult situations. In particular, it was hypothesized that scores on the "other" subscale (i.e., scores based on descriptions of others' emotions) would be associated with IPART scores, whereas "self" scores (i.e., scores based on descriptions of one's own emotions) might show weaker associations. Internal consistency for the present sample in Study 2A was acceptable (Cronbach's α = .75).

Cognitive/Reflectiveness Measures

Cognitive Reflectiveness Test (CRT-7)

The CRT-7 (Toplak et al., 2013) is made up of seven short questions designed such that there is an immediately intuitive but incorrect answer, and a correct answer that, while not logically difficult, requires the individual to devote effortful cognitive resources instead of immediately choosing the intuitively appealing response. Example item:

"If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?" (intuitive answer: 100 minutes; correct answer: 5 minutes)

Thus, the CRT-7 tests the tendency to "stop and think" before immediately trusting one's intuition.

Actively Open-Minded Thinking Scale (AOMTS)

The AOMTS (Toplak et al., 2014) is a self-report scale which asks individuals to rate 30 statements, from 1 (strongly disagree) to 6 (strongly agree), which describe more or less reflective or "rational" attitudes. Example item:

"I like to gather many different types of evidence before I decide what to do." Higher scores indicate more open-minded, reflective attitudes.

Comprehensive Assessment of Rational Thinking (CART)

The CART (Stanovich et al., 2016) assesses vulnerability to various common reasoning biases that arise (in part) from insufficient engagement of reflective capacities. The 2-subscale version administered here includes statistical and scientific reasoning problems. Example item:

"Dice game: Even numbers win and odd numbers lose on a die throw. The fair die has six sides, with three even and three odd numbers. Jan has thrown seven odd numbers in a row. What are her chances of throwing an even number on her next throw?" (correct answer: 3/6).

Higher scores indicate a greater tendency to engage effortful cognition and avoid common reasoning biases during problem-solving, which we reasoned could apply to the social problem solving assessed by the IPART.

When assessing differences on these reflective cognition measures, it is also important to account for individual differences in general intelligence. To do so, we asked participants to complete the 2-subscale Wechsler Abbreviated Scale of Intelligence (WASI-II;(Wechsler, 2011)), a common measure of IQ.

Wechsler Abbreviated Scale of Intelligence (WASI-II)

The 2-subscale Wechsler abbreviated scale of intelligence (WASI-II; (Wechsler, 2011)) is a common measure of IQ and includes subtests of both vocabulary and visuospatial reasoning. This scale was administered by a trained and experienced research technician according to the procedures in the WASI-II technical manual (Wechsler, 2011).

Statistical Analyses

To replicate the results of Study 1, the same reliability and convergent validity analyses were conducted in the baseline sample collected in Study 2. We also examined relationships with demographic variables as in Study 1 to assess the relevance of these characteristics in strategy selection on the IPART. We again conducted post-hoc power analyses to ensure enough statistical power to draw conclusions from tests. Correlations were also assessed between IPART scores and the additional self-report and performance-based measures not available in Study 1, some of which had theoretical relationships to the construct assessed by the IPART.

Results

Preliminary Analyses

Descriptive statistics for IPART sub-scores and the measures collected in Study 1 are shown in **Table 2A**. Descriptive statistics for additional self-report and performance-based measures are shown in **Table 2B**. As in Study 1, IPART scores demonstrated an approximately normal distribution at baseline (skewness < |2|) with good internal consistency (BC: $\alpha = .83$; PC: $\alpha = .79$). As hypothesized, and in support of the findings in Study 1, BC and PC scores were significantly higher in women than in men (BC: $t(425) = 3.26$, PC: $t(425) = 3.77$; $p = .001$ and $p < .001$, respectively). For the interested reader, we also show post-hoc comparisons of other study variables separated by sex (not correcting for multiple comparisons), which tended to show a similar/consistent pattern of differences (e.g., greater mood improvement scores in women, greater mood worsening scores in men). Unlike in Study 1, BC scores showed a significant relationship with age ($r = .13$, $p = .010$), while PC scores only showed a marginal relationship ($r = .10$, $p = .050$). As in Study 1, BC scores were positively correlated with

education level (r = .12, p = .014). However, unlike Study 1, PC scores also showed this relationship (r = .17, p < .001). BC scores, again, were significantly higher than PC scores at baseline (t(426) = 22.47, p < .001).

Table 2A. Descriptive statistics of self-report measures in Study 2A at baseline that were included in Study 1.

Measure	Female	Male	Statistic
IPART (N = 427)			
BC Total	59.63 (6.48)	57.34 (6.61)	t(425) = 3.26, $\mathbf{p = .001}^{**}$
PC Total	51.24 (7.38)	48.03 (9.04)	t(425) = 3.77, $\mathbf{p < .001}^{***}$
BC Improve	27.06 (5.09)	25.57 (4.83)	t(425) = 2.74, $\mathbf{p = .006}^{**}$
PC Improve	20.29 (5.70)	18.77 (6.27)	t(425) = 2.39, $\mathbf{p = .017}^{*}$
BC Divert	5.52 (4.02)	6.20 (3.60)	t(425) = -1.62, p = .106
PC Divert	10.67 (4.57)	10.48 (4.35)	t(425) = 0.38, p = .707
BC Worsen	1.43 (1.74)	2.23 (2.23)	t(425) = -3.93, $\mathbf{p < .001}^{***}$
PC Worsen	3.05 (2.27)	4.75 (3.32)	t(425) = -6.03, $\mathbf{p < .001}^{***}$
SREIS (N = 426)			
Total	3.59 (0.45)	3.57 (0.43)	t(424) = 0.40, p = .688
Perceiving Emotions	3.87 (0.63)	3.73 (0.69)	t(424) = 1.89, p = .060
Use of Emotions	3.31 (0.44)	3.16 (0.51)	t(424) = 2.95, $\mathbf{p = .003}^{*}$
Understanding Emotions	3.23 (0.92)	3.27 (0.82)	t(424) = -0.38, p = .707
Social Management	3.77 (0.76)	3.61 (0.70)	t(424) = 1.99, $\mathbf{p = .048}^{*}$
Managing Emotions of Self	3.71 (0.67)	3.99 (0.62)	t(424) = -3.88, $\mathbf{p < .001}^{***}$
EROS (N = 426)			
Extrinsic Affect Improving	4.03 (0.76)	3.84 (0.75)	t(423) = 2.29, $\mathbf{p = .022}^{*}$
Extrinsic Affect Worsening	1.50 (0.55)	1.40 (0.46)	t(424) = 1.75, p = .082
Intrinsic Affect Improving	3.71 (0.83)	3.44 (0.90)	t(424) = 2.93, $\mathbf{p = .004}^{**}$
Intrinsic Affect Worsening	1.68 (0.69)	1.57 (0.80)	t(421) = 1.39, p = .167
MEOS (N = 427)			
Enhance	63.79 (6.84)	61.61 (7.37)	t(425) = 2.89, $\mathbf{p = .004}^{**}$
Worsen	25.69 (7.27)	28.43 (7.13)	t(425) = -3.51, $\mathbf{p < .001}^{***}$
Conceal	23.36 (5.74)	23.50 (5.21)	t(425) = -0.23, p = .820
Inauthentic	30.16 (7.26)	27.12 (7.00)	t(425) = 3.91, $\mathbf{p < .001}^{***}$
Divert	28.24 (3.17)	27.99 (2.95)	t(425) = 0.73, p = .468
Poor Skills	10.71 (2.77)	11.14 (3.06)	t(425) = -1.42, p = .156
ERQ (N = 427)			
Reappraisal	31.59 (5.50)	31.03 (6.53)	t(425) = 0.90, p = .370
Suppression	13.55 (5.69)	14.91 (5.10)	t(425) = -2.27, $\mathbf{p = .024}^{*}$

Note. *p<.05, **p<.01, ***p<.001

Table 2B. Descriptive statistics of additional measures in Study 2A at baseline that were not assessed in Study 1.

Measure	Female	Male	Statistic
LEAS (N = 388)			
Total	39.47 (4.72)	36.52 (4.68)	$t(386) = 5.49$, $\mathbf{p < .001}^{***}$
Other	31.86 (4.05)	30.09 (4.52)	$t(386) = 3.71$, $\mathbf{p < .001}^{***}$
Self	33.44 (4.10)	30.50 (4.93)	$t(386) = 5.95$, $\mathbf{p < .001}^{***}$
STEM-B (N = 427)			
Total	0.63 (0.09)	0.62 (0.09)	$t(425) = 1.25$, $p = .214$
MSCEIT (N = 426)			
Total	0.53 (0.04)	0.54 (0.05)	$t(423) = -2.73$, $\mathbf{p = .007}^{**}$
Faces Task	0.59 (0.10)	0.61 (0.12)	$t(424) = -2.15$, $\mathbf{p = .032}^{*}$
Facilitation Task	0.49 (0.07)	0.50 (0.07)	$t(424) = -1.51$, $p = .132$
Changes Task	0.64 (0.09)	0.66 (0.08)	$t(424) = -1.87$, $p = .062$
Emotion Management Task	0.44 (0.07)	0.46 (0.06)	$t(424) = -3.31$, $\mathbf{p = .001}^{**}$
Pictures Task	0.59 (0.07)	0.60 (0.07)	$t(424) = -1.75$, $p = .081$
Sensations Task	0.53 (0.08)	0.55 (0.10)	$t(423) = -2.05$, $\mathbf{p = .041}^{*}$
Blends Task	0.57 (0.07)	0.58 (0.07)	$t(424) = -1.25$, $p = .213$
Social Management Task	0.45 (0.09)	0.47 (0.09)	$t(424) = -2.20$, $\mathbf{p = .028}^{*}$
TEIQue (N = 427)			
Total	5.03 (0.62)	5.11 (0.61)	$t(425) = -1.23$, $p = .221$
Emotion Expression	4.92 (1.39)	4.77 (1.24)	$t(425) = 1.06$, $p = .291$
Emotion Regulation	4.64 (0.94)	5.08 (0.89)	$t(425) = -4.35$, $\mathbf{p < .001}^{***}$
Empathy	5.42 (0.87)	5.13 (0.83)	$t(425) = 3.05$, $\mathbf{p = .002}^{**}$
Emotion Perception	4.98 (0.99)	4.90 (0.90)	$t(425) = 0.72$, $p = .471$
Emotion Management	4.62 (0.87)	4.84 (0.86)	$t(425) = -2.33$, $\mathbf{p = .020}^{*}$
Emotionality	5.29 (0.79)	5.11 (0.69)	$t(425) = 2.13$, $\mathbf{p = .034}^{*}$
Social Awareness	4.91 (0.96)	5.14 (0.91)	$t(425) = -2.23$, $\mathbf{p = .027}^{*}$
Sociability	4.69 (0.78)	5.00 (0.74)	$t(425) = -3.81$, $\mathbf{p < .001}^{***}$
DERS (N = 427)			
Total	78.98 (23.38)	72.92 (20.85)	$t(425) = 2.47$, $\mathbf{p = .014}^{*}$
Nonacceptance of Emotion Responses	13.68 (6.53)	12.07 (5.78)	$t(425) = 2.36$, $\mathbf{p = .019}^{*}$
Difficulty Engaging in Goal-Directed Behavior	14.15 (4.76)	12.61 (4.87)	$t(425) = 2.96$, $\mathbf{p = .003}^{**}$
Impulse Control Difficulties	10.47 (4.11)	9.51 (3.38)	$t(425) = 2.27$, $\mathbf{p = .024}^{**}$
Lack of Emotional Awareness	13.65 (4.87)	14.62 (4.60)	$t(425) = -1.87$, $p = .063$
Limited Access to Emotion Regulation Strategies	16.52 (6.68)	14.39 (6.48)	$t(425) = 2.97$, $\mathbf{p = .003}^{**}$
Lack of Emotional Clarity	10.52 (3.78)	9.72 (3.33)	$t(425) = 2.01$, $\mathbf{p = .046}^{*}$
CISS (N = 426)			
Emotions	42.08 (11.37)	36.82 (10.30)	$t(424) = 4.37$, $\mathbf{p < .001}^{***}$
Task	61.09 (9.65)	61.95 (9.13)	$t(424) = -0.83$, $p = .406$
Avoidance	51.69 (9.95)	44.96 (9.67)	$t(424) = 6.28$, $\mathbf{p < .001}^{***}$
Distraction	24.57 (5.99)	20.36 (5.99)	$t(424) = 6.47$, $\mathbf{p < .001}^{***}$
Social Diversion	18.47 (4.87)	16.10 (4.43)	$t(424) = 4.60$, $\mathbf{p < .001}^{***}$
MAIA (N = 427)			
Emotional Awareness	3.41 (1.02)	3.32 (1.06)	$t(425) = 0.75$, $p = .457$

Note. *p<.05, **p<.01, ***p<.001

Sample differences in the two studies' demographic data are presented in **Supplemental Table S2**. The average age of participants in Study 1 (M = 28.79, SD = 3.71) was significantly higher than that of individuals in Study 2A (M = 23.67, SD = 5.57; $t(686) = -13.18$, $p < .001$). The two samples also differed significantly in their sex ratios ($\chi^2 = 35.87$, $p < .001$; higher ratio of female to male participants in Study 2A) and ethnic compositions ($\chi^2 = 18.6$, $p = .005$; more White and Latinx individuals in Study 2A).

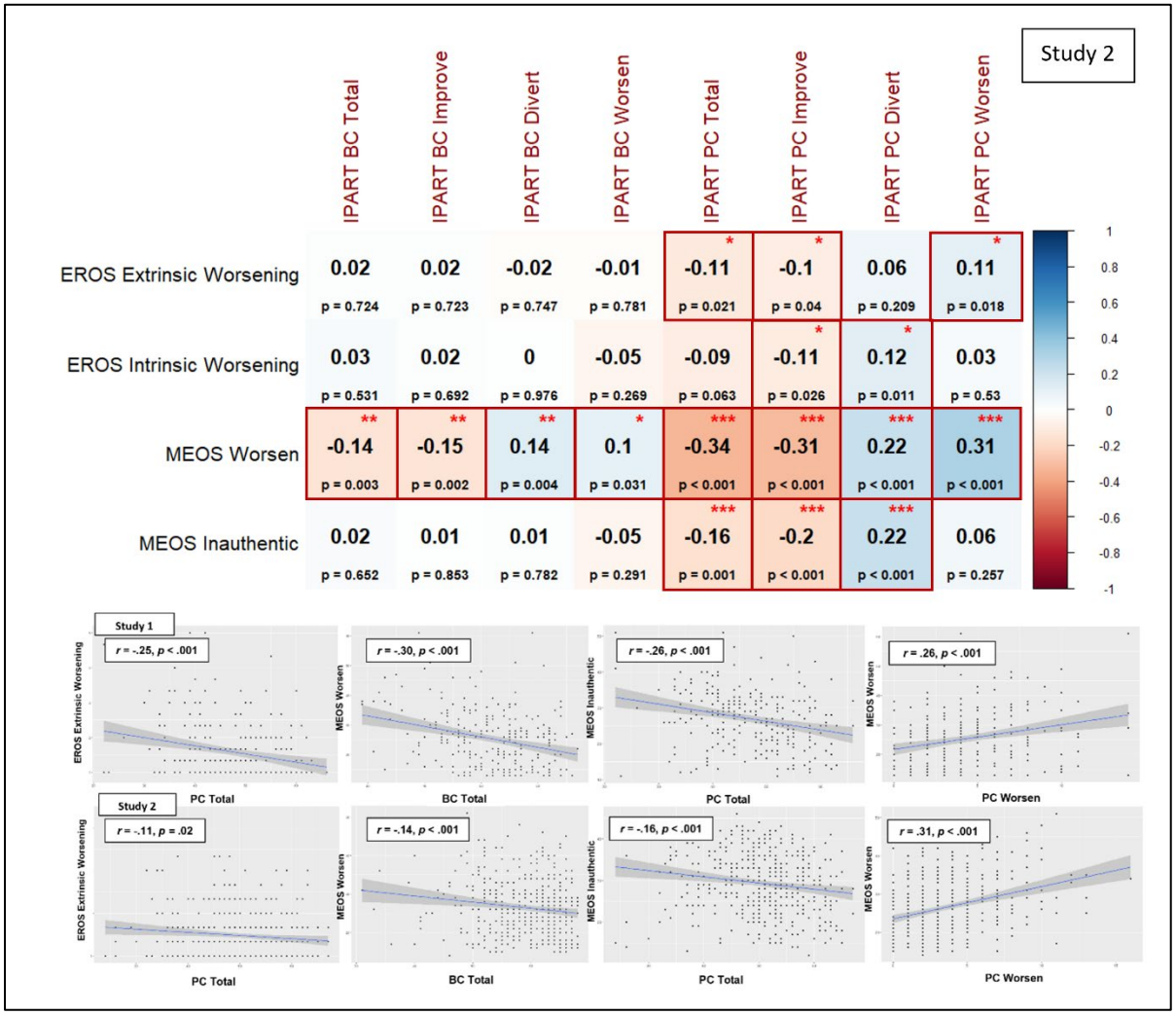
IPART Reliability

Reliability analyses of the 34-item measure administered in Study 2A validated the previous findings of good internal consistency (Cronbach's α: BC = .83, PC = .79). Results of split-half analyses also confirmed good reliability (Guttman $\lambda^2 = .85$ for BC; Guttman $\lambda^2 = .81$ for PC).

Convergent Validity Analyses

Correlations between IPART scores and the measures significantly correlated with IPART scores in Study 1 are shown in the top panel of **Figure 4**. Scatterplots visualizing the strongest relationships that replicated from Study 1 to Study 2 are shown in the bottom panel of this figure. Correlations between IPART scores and all self-report measures included in Study 1 are shown in **Supplemental Figure S1**. Correlations between IPART scores and all measures included in Study 2A are shown in Supplemental **Figure S2A and S2B** (separated by those that clustered with IPART in the optimal clustering solution and those that did not). We note here that, assuming a threshold of 0.8, post-hoc power analyses indicate the full sample size would allow detection of correlations with $r \geq .135$.

Figure 4. ***Top Panel*: Correlation matrix showing relationships in Study 2 between IPART and self-report measures (top panel), all of which were significant in Study 1 (apart from that between PC Worsen and EROS-Intrinsic Affect Worsening).** ***Bottom Panel*: Scatterplots of select relationships that replicated between samples. Boxes with red outline reflect associations that replicated from Study 1 to Study 2A.**



Note. *p<.05, **p<.01, ***p<.001

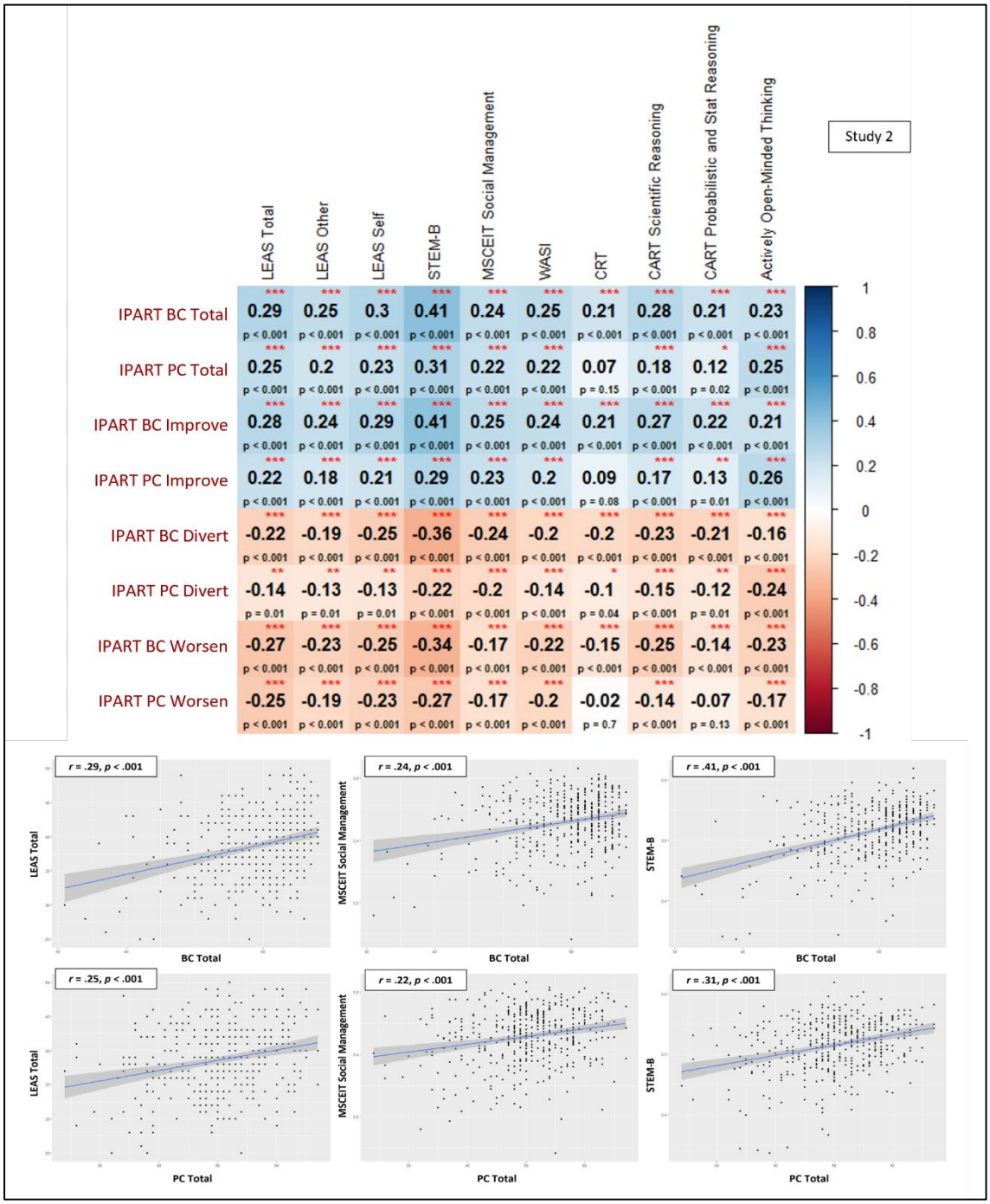
As in Study 1, BC and PC Total scores showed a significant positive association (r = .41, p < .001), but had considerable non-shared variance. Consistent with the results from Study 1, PC Total scores again showed significant positive relationships with MEOS Enhance, SREIS Social Management, SREIS Understanding Emotions, SREIS Managing One's Own Emotions, SREIS Using Emotions, SREIS Total, and EROS-Intrinsic Affect Improving. Total scores for PC were also negatively related to MEOS Worsen and EROS-Extrinsic Affect Worsening, as hypothesized. These significant relationships that replicated from Study 1 to Study 2A are visualized in the bottom panel of **Figure 4**. However, note that many other relationships shown in the top panel of **Figure 4** did not clearly replicate the results from Study 1, particularly those with EROS. Unlike Study 1, PC scores in this sample also showed a significant negative relationship with ERQ Suppression (r = -.22, p < .001), as was initially predicted. Consistent with the results of Study 1, BC Total scores showed significant negative relationships with MEOS Worsen (r = -.14, p < .001). In the present sample, BC scores also showed significant relationships with SREIS Use of Emotion ($p \leq .038$); however, no other relationships with BC scores found in Study 1 were also significant in Study 2A. Note that, given the apparent sparsity of BC and PC Total scores at the low end of the distribution in some cases, we also confirmed that all significant

results remained unchanged after removing any data points identified using iterative Grubbs with a threshold of $p = .01$ (2 outliers for BC Total scores, 0 for PC Total scores).

Analyses also revealed expected inverse relationships between IPART and antisocial behavioral tendencies on two subscales of the short dark triad (SD3). Specifically, Psychopathy and Machiavellianism were negatively associated with IPART PC Total scores (rs = -.34, ps < .001) and positively associated with PC Worsen scores (r = .34, r = .32, respectively; ps < .001). However, relationships between BC and PC scores and the Narcissism scale were weak or nonsignificant (ps ≥ .03). Note that, because this measure was added part way through data collection (n = 169), the SD3 could not be included in the full correlation plots shown in **Supplemental Figures S2A and S2B**. Post-hoc power analyses indicated that this smaller sample size would allow detection of correlations with r ≥ .215. A separate visualization of IPART and SD3 relationships is shown in **Supplemental Figure S3**.

As expected, IPART scores showed significant relationships with several performance-based measures (see **Figure 5**). Among others, both BC and PC scores showed significant positive associations with LEAS scores (rs between .20 and .30, ps < .001), several MSCEIT scores (including Social Management; r_{BC} = .24, r_{PC} = .22; ps < .001), and STEM-B scores (r_{BC} = .41, r_{PC} = .31; ps < .001), as hypothesized. While numerically larger for BC scores, correlation comparisons (based on r-to-z transforms) indicated the relationship between IPART and MSCEIT Social Management scores was not significantly different from any significant relationships observed between IPART and self-report measures. However, the relationships between IPART BC scores and the STEM-B were significantly larger than any relationships with self-report measures (though the same was not true for PC scores).

Figure 5. ***Top Panel*: Correlation matrix showing notable relationships in Study 2 between IPART and performance-based measures.** ***Bottom Panel*: Scatterplots depicting select relationships.**



Note. *p<.05, **p<.01, ***p<.001

Given evidence in the current study (and previous studies) for relationships between sex and both emotion regulation and emotional awareness (Gardener et al., 2013; Hall & Matsumoto, 2004; Wright et al., 2017; Zlomke & Hahn, 2010) as well as relationships between a range of other cognitive abilities and emotion regulation skills (Gyurak et al., 2009; McRae et al., 2012), we also confirmed whether the IPART could predict scores on the previously mentioned performance-based measures after accounting for interactions with sex and cognitive ability. To this end, we ran linear models (LMs) including IPART BC or PC Total scores at baseline and their interactions with sex and WASI-II. Results showed that PC scores significantly predicted all three performance-based measures (ps < .001 to .020), while BC scores

only predicted MSCEIT Social Management (p = .010). **Supplemental Table S3** contains the full results of these LMs.

As shown in **Supplemental Figure S2A**, IPART scores also clustered with measures of general cognitive and reflective abilities (WASI-II, CART, CRT, and AOMTS). Partial correlations between BC and PC Total scores and each measure of cognitive reflectiveness (while controlling for WASI-II) gave further insight into what was driving these relationships. Specifically, when accounting for WASI-II scores, BC Total scores maintained positive relationships with each cognitive measure except for CRT (r = .08, p = .090). On the other hand, PC Total scores only maintained the relationship with AOMTS (r = .20, p < .001), while all others became non-significant. This finding – that BC scores correlate more highly with measures of cognitive ability than PC scores – mirrors what has been found in the literature on SJTs, in which maximal performance scores tend to show stronger relationships with cognitive ability measures than typical performance scores (McDaniel et al., 2007).

Note that, due to a large number of missing scores in the scale at baseline, Flourishing (n = 231) was not included in the clustered plots in **Supplemental Figures S2A and S2B**. IPART PC Total and Improve scores showed weak positive relationships with Flourishing scores (r_{Total} = .13, p = .040; $r_{Improve}$ = .14, p = .030), while PC Divert scores showed a negative relationship with this measure (r_{Divert} = -.14, p = .040). On the other hand, BC scores showed no significant relationships with Flourishing scores, contrary to hypotheses. However, these results should be interpreted cautiously, as post-hoc power analyses indicated that the smaller sample size would only allow reliable (0.8) detection of rs ≥ .185.

A similar trend of weak negative relationships with PC Total and/or Improve scores was also seen with BDI ($r_{Improve}$ = -.10, p = .040) and STAI State (r_{Total} = -.10, p = .040; $r_{Improve}$ = -.12, p = .010), while positive relationships were seen with PC Divert in BDI (r_{Divert} = .11, p = .030), STAI State (r_{Divert} = .14, p < .001) and STAI Trait (r_{Divert} = .12, p = .010). However, as mentioned above, the sample size was underpowered for reliable detection of rs < .135.

Summary of Study 2A

In summary, Study 2A replicated many results in Study 1, including consistency and reliability levels, sex differences, and correlations with the same self-report measures. However, only PC scores showed clear replication for many relationships across both studies, while BC scores were less consistent. Study 2A also demonstrated significant relationships between IPART and multiple performance-based measures of emotional intelligence and awareness not included in Study 1, as well as with measures of general cognitive ability and tendencies to engage in reflective cognition. This further supports the idea that the IPART measures a socio-cognitive ability that overlaps with existing measures, but also appears to capture a distinct and potentially important determinant of social functioning not fully assessed by these other measures (Robinson & Sedikides, 2020).

Study 2B

Study 2B examined longitudinal data collected in a subset of participants in Study 2A that completed an emotional intelligence training program or an emotionally neutral placebo training program to accomplish the following aims:

1) Examine test-retest reliability within the group that took the placebo program with content unrelated to emotional intelligence.
2) Examine whether, relative to the placebo program, IPART scores were appropriately sensitive to emotional intelligence training.
3) Examine whether changes in IPART scores over time showed relationships with change scores over time in other socio-emotional measures analyzed in Study 2A.

Methods

Participants

Data for Study 2B were gathered as part of a larger longitudinal study designed to validate the effectiveness of an emotional intelligence training program. Previous studies have shown that this program is effective in improving scores on multiple standard measures of emotional intelligence and other socio-cognitive abilities (Durham et al., 2023; Persich et al., 2021; Smith et al., 2023). Thus, as a further test of the face validity of the IPART, we took the opportunity to examine whether scores might be sensitive to this training program (i.e., over and above any test-retest effects observed pre- and post-completion of a placebo training program) and how IPART scores covaried with other EI measures over time.

The participants described in Study 2A were randomly assigned to one of two groups and asked to complete either: 1) the emotional intelligence training program, where they completed a 10 to 12-hour training course over either 1 or 3 weeks that was designed to improve socio-emotional skills, or 2) a placebo training condition, which was a matched program that also included a 10 to 12-hour training course over 1 or 3 weeks, but that did not focus on socio-emotional skills (instead, teaching skills such as scientific classification of plants and animals). Out of the participants with available IPART data, 225 were assigned to the emotional intelligence program and 202 were assigned to the placebo training program. After finishing the training or placebo program, participants completed the same measures from the baseline visit in a post-program visit and a subset of these measures at a 6-month follow-up. Out of the 427 individuals in Study 2A, 311 returned for the post-training visit (Training group: N = 159; Placebo group: N = 152). For reasons explained in previous reports assessing effects of this training program (including the COVID-19 lockdown, among other factors), only 89 participants participated in the 6-month follow-up (Training group: N = 53; Placebo group: N = 36).

We used the data from 202 participants in the placebo group to assess test-retest reliability (i.e., relating IPART scores at baseline and at the post-placebo program visit). We then explored whether IPART scores improved over time in the emotional skills training program group (i.e., compared to changes in the placebo group) and whether changes in IPART scores were associated with changes in other socio-emotional measures.

Statistical Analyses

To assess test-retest reliability, intraclass correlation coefficients (ICCs) based on mean-rating (k = 2), two-way, agreement models were performed on IPART scores in the placebo group from baseline to post-program completion using the *icc* function from the *irr* package in R (v0.84.1; (Gamer & Lemon, 2019a, 2019b)). All ICC analysis choices follow the guidelines outlined by (Koo & Li, 2016).

To examine the effects of training program condition (training vs. placebo) and timepoint (baseline to post-training) on IPART scores (and their interaction), we ran linear mixed effects analyses (LMEs) with group and timepoint as (treatment-coded) predictors. As LMEs follow the "intention-to-treat" principle by using a likelihood-based method, we did not impute missing values. Power analyses for these LMEs were performed using *simr* (Green & MacLeod, 2016) with *nsim* = 100 simulations. These LMEs were also repeated for IPART scores from post-program to 6-month follow-up to explore potential long-term effects.

To assess whether BC scores were affected differently by the training program than PC scores, we also ran LMEs with program condition (treatment-coded), timepoint (baseline to post-program; treatment-coded), and type of IPART Total score (BC/PC; sum-coded) as predictors.

As a final test of convergent validity, we also examined whether changes over time in IPART scores were correlated with changes in other measures across all participants (after accounting for changes that could be explained by variation in baseline scores; i.e., by correlating the residuals in linear models that used baseline scores as predictors of pre-to-post changes).

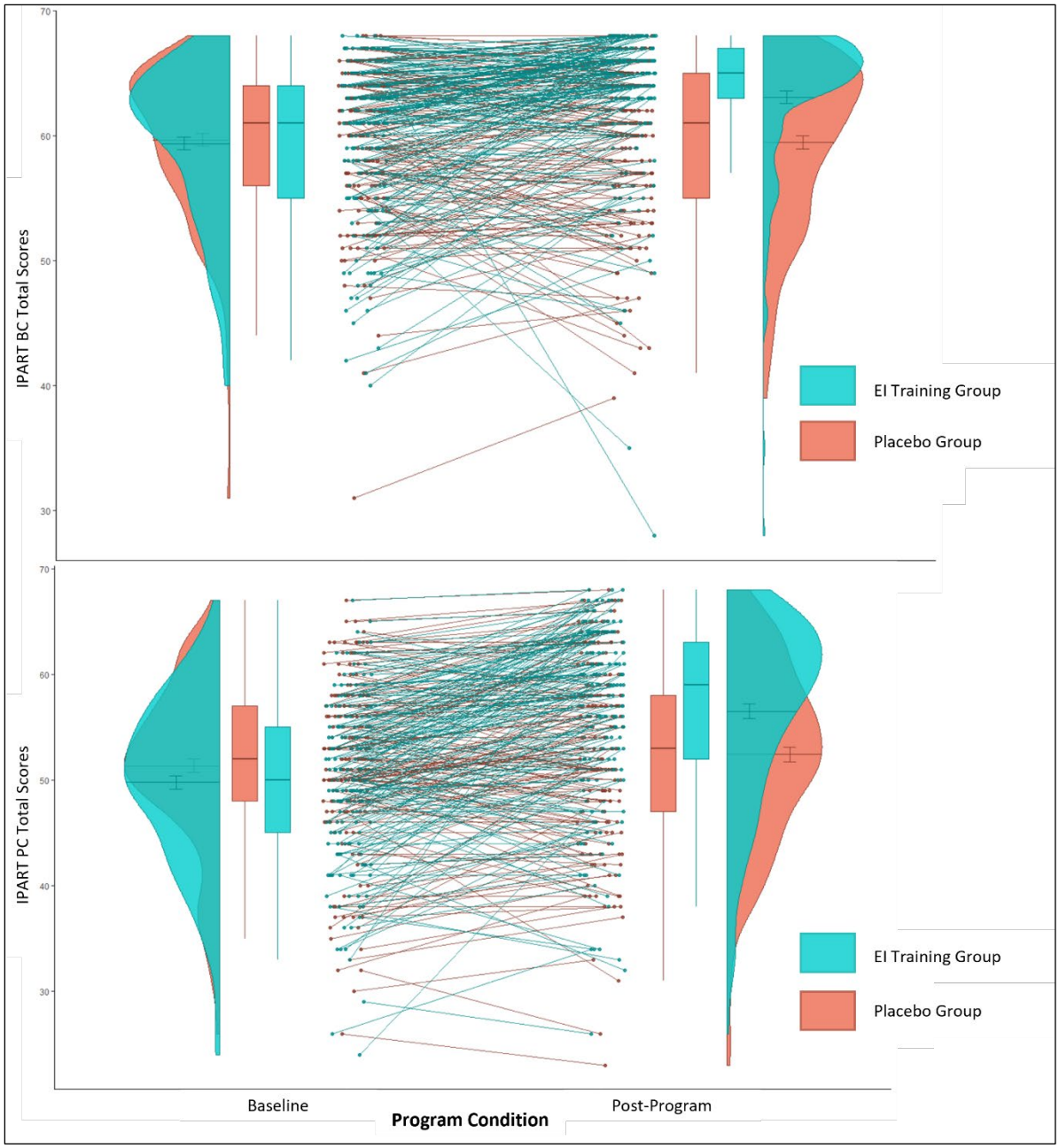
Results

Test-Retest Reliability

Preliminary analyses found that participants who did not come back for the follow-up study visit (N = 116) had significantly lower IPART BC Total scores at baseline than those who did return for follow-up ($t(425) = -2.68$, $p = .008$). However, the two groups did not differ in PC Total scores at baseline ($t(425) = -0.86$, $p = .388$).

Test-retest reliability for those who did return for the follow-up visit was measured from baseline to post-program completion in the placebo group. Following Cicchetti (1994), PC Total and Improve scores demonstrated excellent reliability (ICC = 0.81 for both) across timepoints, while BC Total and Improve scores were slightly lower (ICC = 0.74 for both). Divert and Worsen sub-scores for both BC and PC showed good reliability from baseline to post-program completion (BC_{Divert}: ICC = 0.66; PC_{Divert}: ICC = 0.73; BC_{Worsen}: ICC = 0.60; PC_{Worsen}: ICC = 0.73). To assess potential differences between those who completed placebo training after 1 week vs. 3 weeks, we conducted z-score difference tests on separately calculated ICCs. These analyses revealed that ICCs for most IPART scores were not different between the two groups, apart from BC Total and Improve scores, which were more consistent in the 1-week group from baseline to post-training ($z_{Total} = 2.73$, $p = .006$; $z_{Improve} = 2.94$, $p = .003$). **Figure 6** illustrates BC and PC Total distributions at baseline and post-program in both groups.

Figure 6. Raincloud plots showing IPART BC and PC Total distributions (with mean and SE) at baseline (left) and post-program (right) in both the placebo group and the training program group from Study 2B. Points/lines show change trajectories from baseline to post-program for individuals in each group.



Effects of the Training Program

As reported in **Table 3**, results of LMEs investigating the effects of the training program over time indicated that program condition, timepoint, and their interaction were all highly significant predictors for both BC and PC Total scores. Post-hoc contrasts revealed that, for both BC and PC Total and Improve, scores in the training group were significantly higher after program completion. Divert and Worsen scores, on the other hand, significantly decreased after completion of the training program, as expected. Main effects of training program and time were explained by their interaction, indicating that the training group saw significant improvements in IPART scores after completing the program, while the placebo group showed weaker changes. However, test-retest effects were also observed. Namely,

PC Total and Improve scores increased with repeated testing in the placebo group, and PC Divert scores decreased over time in this group.

Post-hoc power analyses, based on the effect sizes of significant results, revealed that analyses for IPART BC Total, Improve, and Divert scores were 84-94% powered to detect observed effects of program, 100% powered to detect observed effects of visit, and 100% powered to detect the observed interactions. Due to the smaller observed effect sizes, the model predicting BC Worsen had only 57% power to detect the observed effect of program, but 98% power for the effect of visit, and 63% power for the associated interaction. Similarly, models predicting PC scores were less powered due to the low observed effect sizes in some cases: program (64-78% power), visit (100% power), and their interactions (97-100% power).

Table 3. Results of LMEs predicting IPART scores in Study 2B (baseline to post-program)

	F (*df*)	***p***	**η_p^2**	**B [CI]**	**Post-Hoc Contrasts for Program by Visit Interaction**
BC Total					
Program Condition	9.63 (1, 431)	**.002**	0.02	0.34 [-0.91, 1.59]	*Placebo:* T2-T1: 0.05, *t*(347) = 0.12, *p* = .906
Visit	40.08 (1, 349)	**< .001**	0.10	0.05 [-0.83, 0.94]	
Program*Visit	35.83 (1, 349)	**< .001**	0.09	3.78 [2.54, 5.02]	*Training:* T2-T1: 3.83, *t*(351) = 8.71, ***p* < .001**
PC Total					
Program Condition	5.84 (1, 429)	**.016**	0.01	-0.25 [-1.84, 1.34]	*Placebo:* T2-T1: 1.27, *t*(338) = 2.48, ***p* = .014**
Visit	126.68 (1, 341)	**< .001**	0.27	1.27 [0.27, 2.27]	
Program*Visit	56.36 (1, 340)	**< .001**	0.14	5.37 [3.96, 6.77]	*Training:* T2-T1: 6.63, *t*(343) = 13.30, ***p* < .001**
BC Improve					
Program Condition	10.49 (1, 431)	**.001**	0.02	0.19 [-0.77, 1.16]	*Placebo:* T2-T1: -0.06, *t*(347) = -0.16, *p* = .876
Visit	43.69 (1, 349)	**< .001**	0.11	-0.06 [-0.75, 0.64]	
Program*Visit	44.08 (1, 349)	**< .001**	0.11	3.27 [2.30, 4.23]	*Training:* T2-T1: 3.22, *t*(351) = 9.37, ***p* < .001**
PC Improve					
Program Condition	6.83 (1, 429)	**.009**	0.02	0.21[-1.39, 0.98]	*Placebo:* T2-T1: 0.99, *t*(338) = 2.60, ***p* = .010**
Visit	145.99 (1, 341)	**< .001**	0.30	0.99 [0.25, 1.74]	
Program*Visit	66.24 (1, 340)	**< .001**	0.16	4.32 [3.28, 5.36]	*Training:* T2-T1: 5.31, *t*(343) = 14.33, ***p* < .001**
BC Divert					
Program Condition	10.42 (1, 430)	**.001**	0.02	-0.05 [-0.80, 0.71]	*Placebo:* T2-T1: 0.14, *t*(352) = 0.47, *p* = .637
Visit	38.94 (1, 355)	**< .001**	0.10	0.14 [-0.44, 0.71]	

Program*Visit	45.13 (1, 354)	**< .001**	0.11	-2.74 [-3.54, -1.94]	*Training:* T2-T1: -2.60, $t(357) = -9.16$, ***p* < .001**
PC Divert					
Program Condition	6.95 (1, 428)	**.009**	0.02	0.16	*Placebo:* T2-T1: -0.73, $t(344) = -1.55$, ***p* = .017** *Training:* T2-T1: -3.99, $t(347) = -13.43$, ***p* < .001**
Visit	127.54 (1, 344)	**< .001**	0.27	-0.73	
Program*Visit	58.52 (1, 344)	**< .001**	0.15	-3.26	
BC Worsen					
Program Condition	4.66 (1, 428)	**.031**	0.01	-0.14	*Placebo:* T2-T1: -0.12, $t(358) = -0.84$, $p = .400$ *Training:* T2-T1: -0.63, $t(363) = -4.53$, ***p* < .001**
Visit	14.72 (1, 361)	**< .001**	0.04	-0.12	
Program*Visit	6.50 (1, 360)	**.011**	0.02	-0.51	
PC Worsen					
Program Condition	2.16 (1, 427)	.143	0.01	0.05	*Placebo:* T2-T1: -0.28, $t(344) = -1.55$, $p = .123$ *Training:* T2-T1: -1.31, $t(350) = -7.42$, ***p* < .001**
Visit	40.87 (1, 347)	**< .001**	0.11	-0.28	
Program*Visit	16.51 (1, 346)	**< .001**	0.05	-1.03	

Additionally, results of LMEs investigating the long-term effects of the training program from post-program to 6-month follow-up are available in **Supplemental Table S4**. Post-hoc contrasts indicated that BC Total and Improve scores in both groups decreased over time and Divert scores increased. In the training group, BC Worsen scores marginally increased after the 6-month period. Interestingly, no PC scores in the placebo group changed over time, while Total and Improve scores decreased and Divert scores increased in the training group. No differences in IPART scores were significant when comparing baseline to 6-month follow-up. These results therefore suggest that, while IPART scores improved from pre- to post-training, these improvements were temporary and tended to return to baseline 6 months later.

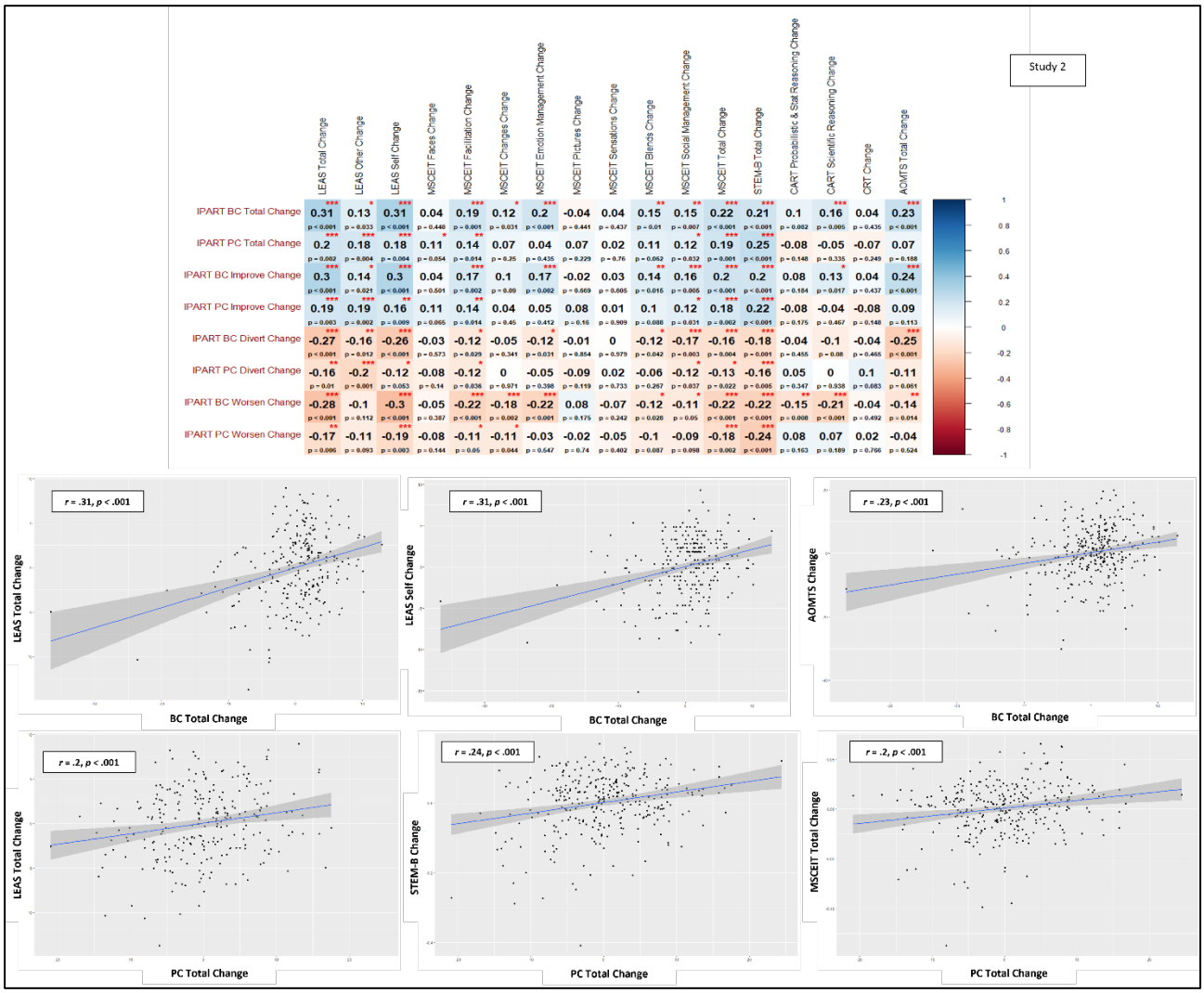
Results of LMEs examining the effects of program condition, timepoint, and BC/PC IPART scores are included in **Supplemental Table S5**. The same pattern of main effects and interactions as in **Table 3** was found here. Post-hoc contrasts revealed that the training program did not target specifically BC or PC scores or any of their sub-scores, as all showed significant improvement from pre- to post-training.

As shown in **Supplemental Figures S4A, S4B, and S5** (which include measures that clustered with IPART (S4A and S4B) and those that did not (S5)), after accounting for variation in baseline scores, improvements in IPART scores from baseline to post-training were significantly associated with improvements in many of the same measures showing significant cross-sectional correlations with the IPART. Interestingly, while not all of the cross-sectional BC relationships from Study 1 were found in Study 2A, changes in BC scores did correlate significantly with changes in some of these measures, such as MEOS Enhance ($r = .19$, $p < .001$), SREIS Total scores ($r = .19$, $p < .001$) and various SREIS sub-scores. Changes in BC Total scores also correlated significantly with changes in TEIQue Relationships scores ($r = .18$, $p < .001$). Changes in PC scores had stronger relationships with changes

in measures such as SREIS Total ($r = .35$, $p < .001$), SREIS Understanding Emotions ($r = .30$, $p < .001$), and DERS Total ($r = -.28$, $p < .001$) (noted relationships are visualized in scatterplots in **Supplemental Figure S6**). Interestingly, while correlations between IPART scores and symptom measures were weak, changes in PC Total and Improve scores were significantly correlated with changes in BDI and STAI State and Trait scores, such that increases in PC scores were associated with decreases in symptomology (separately shown in **Supplemental Figure S7** for ease). Most notable were positive relationships between increases in PC Divert scores and increases in BDI and STAI Trait scores ($r = .20$, $p < .001$ and $r = .25$, $p < .001$, respectively). In other words, depression and anxiety became greater in those who increased their tendency to adopt suboptimal distraction-based emotion management strategies.

As shown in **Figure 6** below, changes in BC and PC Total scores also correlated positively with the three performance-based socio-emotional measures (STEM-B: $r_{BC} = .21$, $p < .001$, $r_{PC} = .24$, $p < .001$; MSCEIT Social Management: $r_{BC} = .15$, $p = .010$, $r_{PC} = .12$, $p = .030$; LEAS Total: $r_{BC} = .31$, $p < .001$, $r_{PC} = .20$, $p < .001$). Finally, changes in BC scores also showed significant relationships with changes in CART Scientific Reasoning ($r = .18$, $p < .001$) and AOMTS ($r = .23$, $p < .001$) while changes in PC scores did not see the same results ($p \geq .150$). Notable significant relationships are shown in scatterplots in the bottom panel of **Figure 6**.

Figure 6. ***Top Panel*: Correlation matrix showing relationships between change over time in IPART scores and change over time in other performance-based and cognitive/reflective measures collected in Study 2 (after accounting for baseline scores).** ***Bottom Panel*: Example scatterplots depicting some significant relationships.**



Note. *p<.05, **p<.01, ***p<.001

Summary of Study 2B

In summary, Study 2B found that the IPART had good test-retest reliability. It further showed that IPART scores could be improved by an emotional intelligence training program relative to placebo. Finally, it showed that changes in IPART scores covaried with several other socio-emotional measures over time. These latter findings thus provide additional support for face validity and convergent validity of the IPART as a measure of a meaningful socio-emotional skill.

Discussion

In the studies described above, we detailed the creation and validation of the Inter-Personal Affect Regulation Test (IPART), a novel performance-based assessment of the ability to improve the emotions of others in a goal-directed manner. In Study 1, the final version of the test (with 34-items) showed good internal consistency and evidence suggested a single factor was sufficient to describe both BC and PC Total scores. Study 1 also provided evidence of convergent validity, demonstrating a consistent

pattern of significant relationships between IPART scores and several widely-used self-report measures of related constructs in expected directions (e.g., greater self-reported EI and reduced tendencies to worsen the emotions of others). This was true for both the "Best Choice" (BC) and "Personal Choice" (PC) scores. As expected, female participants also showed better performance than male participants. Study 2A replicated many results in Study 1, including consistency and reliability levels, sex differences, and correlations with some of the same self-report measures. Other results were less consistent between studies (i.e., significant relationships with EROS-Intrinsic and -Extrinsic Affect Worsening in the first study but not the second). IPART scores in the second study also demonstrated significant relationships with multiple performance-based measures of emotional intelligence and awareness not available in Study 1, as well as positive relationships with measures of general cognitive ability and reflective tendencies – as might be expected if IPART scores measure a specific aspect of social cognition. However, only PC scores showed clear replication for many relationships observed in the first study; BC scores were less consistent.

Study 2B showed that IPART scores had good test-retest reliability, that they covaried over time with other socio-emotional measures, and that they could be improved with training. Crucially, Study 2B also revealed positive relationships between PC scores and Flourishing scores, which, given the strong social component of life satisfaction, further supports the IPART's construct validity. However, these results are preliminary, as the smaller sample size for those who had Flourishing scores was underpowered to confidently detect these relationships. Therefore, these effects should be confirmed in future studies.

The IPART similarly showed negative relationships with measures of anxiety and depression symptoms. In particular, higher PC Improve scores were associated with lower state-anxiety and depression. In longitudinal data, we also observed relationships between increases in PC scores (Total and Improve) and decreases in depression and anxiety (both state and trait) symptoms. As with the relationships to Flourishing scores, this should be seen as preliminary, as the sample was underpowered to detect these effects. However, if confirmed in future work, it suggests studies might utilize the IPART to address clinically relevant questions, such as how inter-personal affect regulation skills relate to well-being.

While significant and in the predicted directions, it is worth highlighting that observed relationships with self-report measures tended to have small effect sizes (e.g., *rs* between .10 and .34 for MEOS and EROS), which suggests that the IPART primarily measures something separable from self-perceived attributes. This finding is somewhat expected, as self-report measures tend to correlate more highly with other self-report measures than with performance measures due to common method variance (Podsakoff et al., 2003). However, the consistent pattern of relationships seen across multiple self-report and performance measures jointly supports convergent validity. Together, these findings therefore suggest that the IPART measures skills that are related to, but largely distinct and separable from, those assessed in existing self-report and performance-based instruments.

Importantly, BC and PC scores were only moderately correlated (r = .62 in Study 1, and r = .42 in Study 2A). These moderate correlations revealed that, even if an individual recognized one response as the best option, they did not always indicate that they would choose that response themselves. This discrepancy between the ability to recognize and choose the best response in a social context represents a novel and under-studied individual difference within inter-personal emotion regulation research that is plausibly of greater relevance than recognition ability alone, and highlights the unique strengths of the IPART, as other performance measures of emotion management (e.g., MSCEIT and STEM) focus primarily on the ability to recognize the best option. Another indicator that PC scores may be important is that BC scores in Study 2 showed potential ceiling effects (as can be seen in **Figure 5**), while PC scores showed more variability. This further suggests that, while people are often able to recognize the best responses to these complex social situations, there are greater individual differences in the tendency to choose those responses. The ceiling effects on BC scores in Study 2 may also help explain the lack of replication of some significant relationships seen in Study 1. One important direction for future

research will be to evaluate whether PC scores can uniquely account for real-world differences in observed patterns of behavior.

A related strength of the IPART is that the response options designated as best, intermediate, and worst for BC and PC scores were motivated and designed based on empirically supported differences in the effectiveness of distinct emotion regulation strategies (i.e., empathy/support addresses the situation and is more effective/adaptive in the long-term than distraction/situation avoidance, which is in turn more effective than invalidating responses). This therefore goes a step beyond standard consensus-based scoring approaches often used in development of existing performance measures.

Because data from Study 2 were part of a larger longitudinal study on EI training, as a further test of face validity we also took the opportunity to assess whether IPART scores were sensitive to this training. Here we found evidence that IPART scores did improve in the short-term after training (relative to a placebo training), but they appeared to return to pre-training levels after several months (at least in a small subset of participants who returned for 6-month follow-up).

Limitations and Future Directions

The present studies have limitations that are important to consider. First, it is worth noting that participants in Study 1 and Study 2 had different demographic compositions. Study 1 was also conducted online which carries with it well-known limitations with respect to control of the testing environment and the previous experiences of participants (Chandler et al., 2019). Study 2 was conducted in-person, had a lower average age, and had a higher ratio of female participants. It could be, for example, that the relationships in Study 1 that did not replicate in Study 2 were influenced by such differences (e.g., in age or sex ratio). Additionally, as our MTurk sample in Study 1 was solely from the United States, and the in-person sample in Study 2 represents a specific region in the United States, these samples only represent a narrow range for age, education, socioeconomic status (SES), and racial diversity. Therefore, examining other cultural groups will be necessary to support the generalizability of our results. The present samples were also taken only from student and community populations. It will be useful for future research to examine the IPART in other populations of interest, such as those with affective disorders.

It is also worth noting that analyses of change scores, as done in our longitudinal data, come along with particular statistical and interpretative challenges, associated with potentially confounding effects such as regression to the mean. This has led to mixed views on whether and how they should be done (Allison, 1990; Cronbach & Furby, 1970; Gardner & Neufeld, 1987; Lord, 1958; McArdle, 2009; Rogosa & Willett, 1983; Thomas & Zumbo, 2012), which should be kept in mind when considering some relationships we observed in Study 2B (noting that baseline scores were accounted for in all change-score analyses).

Another potential limitation in the results of both reported studies is the frequency of small effect sizes, which could be seen to limit convergent validity. However, the sample size provided sufficient power to detect these effects in most cases, and low-magnitude relationships were often expected. The replicable results also provide robust evidence for the presence of these relationships (i.e., that they are not false positives), even if weak, which is the most crucial element for validation purposes in our view. It should also be highlighted that some findings were underpowered and should be considered preliminary. This includes the relationships observed with depression and anxiety symptom measures, and with one measure of subjective well-being. Future studies will need to replicate these results.

Another issue with respect to validity is that, while evidence suggested that one factor was sufficient for the 34 items that comprise the current IPART, only about 13% of the total variance was explained by this factor for both BC and PC scores. Further iterations of the IPART could further examine this factor structure and consider subscales or versions with a reduced number of items. However, as argued in Catano et al. (2012), items in situationally based tests like the IPART should not necessarily be expected to co-vary highly because of the heterogeneous nature of the specific situations presented.

These authors have therefore suggested that a stronger focus should be placed on other metrics in this context, such as test-retest reliability (i.e., which was good for the IPART in our results).

It is also worth considering the IPART in relation to other perspectives on contextualized emotional regulation. For example, some literature suggests optimal responses may vary by the individual and the situation; e.g., positive comments may only make others feel better if a situation is perceived to be controllable (Aldao et al., 2015; Bonanno & Burton, 2013; Doré et al., 2016; Fernandes & Tone, 2021; Shu et al., 2021; Troy et al., 2013; Wenzel et al., 2020). Some responses that would be considered suboptimal for long-term solutions could also be effective in causing short-term improvements in another's moods (e.g., distraction) (Augustine & Hemenover, 2009; Sheppes & Meiran, 2007). The items in the IPART were constructed to minimize some of these issues to the extent possible. For example, the degree to which empathic responses vs. constructive planning were included in optimal responses was sensitive to aspects of the described situations. We also focused on responses that tend to be optimal in the long run across situations of interpersonal conflict and likely of most overall benefit. Responses expected to have only short-term benefit were scored lower, but higher than those expected to worsen the situation, to reflect this difference. Further, the stated goal focused on the situation, and some (e.g., distraction-based) strategies assigned lower scores strategies ignored the situation, even if they could have some short-term efficacy. However, more research will be needed to link IPART scores to objective outcomes, such as peer reports of a person's inter-personal emotion regulation abilities, in order to explore the real-world effects of endorsing responses that are effective long-term versus short-term.

While constructed for feasible/objective scoring, the multiple-choice nature of the IPART may also limit its generalizability to real-life contexts, as people may actually respond outside the given options. However, the current design still allows for participants to demonstrate relative differences in the ability to choose the optimal response, as well as communicate which response they would be most likely to choose. Situationally-based assessments are also thought to be low-fidelity simulations in which the process of responding to the situation reflects similar processes that elicit behavior in real life (Corstjens et al., 2017; Motowidlo et al., 1990). Nonetheless, this ability may differ from the ability to spontaneously generate effective strategies in real interpersonal situations and, thus, future research should investigate this concept and compare it to the skills represented in the IPART. A related limitation may arise from the item format, in which BC and PC questions were asked together for each item. Although BC and PC scores were only moderately correlated, asking participants to give their best choice (BC) response just before their personal choice (PC) response could have nonetheless promoted congruency and limited the separability of their respective scores. One interesting direction for future research could therefore be to separately administer BC and PC versions of the IPART and evaluate whether reductions are observed in the congruence between measures. This could dove-tail with other interesting lines of work showing, for example, that discrepancies between what people think they should do versus would do in social situations are significantly associated with well-being and objective measures of social competence (Robinson et al., 2022).

Practical Implications and Conclusions

One possible use of the IPART could be to help evaluate the efficacy of clinical interventions that focus on improving interpersonal skills/relationships and gaining social support. Improvements in responses on the IPART (e.g., as found in response to the emotional intelligence training in Study 2) could offer evidence of efficacy and provide distinct insights from self-reported changes. There are also interesting questions about life outcomes resulting from differences in inter-personal regulation, such as relationship or career success, which might be assessed by this scale. For example, future longitudinal research using the IPART could construct structural and/or causal models to evaluate paths from skill development to adaptive outcomes.

In conclusion, these studies provide support for the internal reliability and convergent validity of the IPART – replicating and extending many initial results in a second sample. Given the plausible real-world consequences of inter-personal affect regulation ability, we conclude that the IPART represents

an important step toward effectively measuring this ability, as well as its relationship to other cognitive abilities and its consequences for overall life success. One unique and noteworthy strength of the IPART is the ability to separately examine the differences in an individual's ability to identify the most effective emotion management approach and their actual choice patterns (i.e., whether individuals, having identified the best course of action, follow through with it). Such information could have wide-ranging implications for research on emotion and interpersonal relationships, as well as for the development of effective interventions and training programs for emotional intelligence.

Additional Information

Supplementary Material

https://osf.io/cxjfh/

Funding

This study was supported by a USAMRAA grant to WDSK: W81XWH-16-1-0062

Conflict of Interest

None declared.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The protocols for this project were approved by the Institutional Review Board of the University of Arizona (Study 1 and 2) and the U.S. Army Human Research Protection Office (HRPO; Study 2).

Data Availability

The full 34-item IPART is included in the appendix. The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Author CRediT Statement

Ryan Smith took the lead in conceptualizing the IPART, designing Study 1, supervising its execution, and writing the manuscript. Claire A. Lavalley and Michelle R. P. Durham played major roles in carrying out analyses and writing/editing the manuscript. Meltem Ozcan, Courtney Smith, Anna Sanova, Jacqueline Marquez, Sarah Berryhill, and Anna Alkozei aided in test creation and editing the manuscript. Adam Raikes aided in statistical analyses and editing the manuscript. William D. S. Killgore designed Study 2, obtained funding and resources for development and data collection, supervised study execution, and edited the manuscript.

References

Aldao, A., & Nolen-Hoeksema, S. (2012). The influence of context on the implementation of adaptive emotion regulation strategies. *Behaviour research and therapy*, *50*(7-8), 493-501.

Aldao, A., Sheppes, G., & Gross, J. J. (2015). Emotion regulation flexibility. *Cognitive Therapy and Research*, *39*, 263-278.

Allen, V., Rahman, N., Weissman, A., MacCann, C., Lewis, C., & Roberts, R. D. (2015). The Situational Test of Emotional Management–Brief (STEM-B): Development and validation using item response theory and latent class analysis. *Personality and Individual Differences*, *81*, 195-200.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological methodology*, 93-114.

Augustine, A. A., & Hemenover, S. H. (2009). On the relative effectiveness of affect regulation strategies: A meta-analysis. *Cognition and Emotion*, *23*(6), 1181-1220.

Austin, E. J., & O'Donnell, M. M. (2013). Development and preliminary validation of a scale to assess managing the emotions of others. *Personality and Individual Differences*, *55*(7), 834-839.

Austin, E. J., Saklofske, D. H., Smith, M., & Tohver, G. (2014). Associations of the managing the emotions of others (MEOS) scale with personality, the Dark Triad and trait EI. *Personality and Individual Differences*, *65*, 8-13.

Banziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*(5), 691-704. https://doi.org/10.1037/a0017088

Bar-On, R. (2002). EQ-i BarOn emotional quotient inventory: A measure of emotional intelligence: User's Manual.

Barchard, K. A., Lane, R. D., & Watson, B. D. (2010). *Emotional awareness: Computer and hand scoring of an open-ended test.* American Psychological Association convention, San Diego, CA.,

Barlow, D. H., Allen, L. B., & Choate, M. L. (2016). Toward a unified treatment for emotional disorders - republished article. *Behav Ther*, *47*(6), 838-853. https://doi.org/10.1016/j.beth.2016.11.005

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490-498.

Berking, M., & Wupperman, P. (2012). Emotion regulation and mental health: recent findings, current challenges, and future directions. *Current opinion in psychiatry*, *25*(2), 128-134.

Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on psychological science*, *8*(6), 591-612.

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: a comparison of self-report and performance measures of emotional intelligence. *Journal of personality and social psychology*, *91*(4), 780.

Burleson, B. R. (2003). Emotional support skills. In *Handbook of communication and social interaction skills* (pp. 569-612). Routledge.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. routledge.

Catano, V. M., Brochu, A., & Lamerson, C. D. (2012). A ssessing the R eliability of S ituational J udgment T ests U sed in H igh-S takes S ituations. *International Journal of Selection and Assessment*, *20*(3), 333-346.

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. In https://CRAN.R-project.org/package=pwr

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, *51*(5), 2022-2038. https://doi.org/10.3758/s13428-019-01273-7

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: a systematic review. *Research in developmental disabilities*, *34*(5), 1505-1520.

Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgement tests for selection. *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention*, 226-246.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

Cronbach, L. J., & Furby, L. (1970). How we should measure" change": Or should we? *Psychological bulletin*, *74*(1), 68.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social indicators research*, *97*, 143-156.

Doré, B. P., Silvers, J. A., & Ochsner, K. N. (2016). Toward a personalized science of emotion regulation. *Social and Personality Psychology Compass*, *10*(4), 171-187.

Dunkel Schetter, C., & Brooks, K. (2009). The nature of social support. *Encyclopedia of human relationships*, 1565-1570.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest*, *5*(3), 69-106.

Durham, M. R. P., Smith, R., Cloonan, S., Hildebrand, L. L., Woods-Lubert, R., Skalamera, J., Berryhill, S. M., Weihs, K. L., Lane, R. D., & Allen, J. J. (2023). Development and validation of an online emotional intelligence training program. *Frontiers in Psychology*, *14*.

Endler, N., & Parker, J. (1990). Coping inventory for stressful situations. Multi-Health systems Incorporated. In.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fernandes, M. A., & Tone, E. B. (2021). A systematic review and meta-analysis of the association between expressive suppression and positive affect. *Clinical psychology review*, *88*, 102068.

Gamer, M., & Lemon, J. (2019a). *irr: Various Coeffecients of Interrater Reliability and Agreement*. In https://CRAN.R-project.org/package=irr

Gamer, M., & Lemon, J. (2019b). irr: Various coefficients of interrater reliability and agreement. *http://cran. r-project. org/web/packages/irr/irr. pdf*.

Gardener, E. K., Carr, A. R., MacGregor, A., & Felmingham, K. L. (2013). Sex differences and emotion regulation: an event-related potential study. *PloS one*, *8*(10), e73475.

Gardner, R., & Neufeld, R. W. (1987). Use of the simple change score in correlational analyses'. *Educational and psychological measurement*, *47*(4), 849-864.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of psychopathology and behavioral assessment*, *26*, 41-54.

Green, P., & MacLeod, C. J. (2016). simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Gross, J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology*, *2*, 271-299.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological inquiry*, *26*(1), 1-26.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, *85*(2), 348.

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: the acute effects of inhibiting negative and positive emotion. *Journal of abnormal psychology*, *106*(1), 95.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282.

Gyurak, A., Goodkind, M. S., Madan, A., Kramer, J. H., Miller, B. L., & Levenson, R. W. (2009). Do tests of executive functioning predict ability to downregulate emotions spontaneously and when instructed to suppress? *Cognitive, Affective, & Behavioral Neuroscience*, *9*(2), 144-152.

Hair Jr, J., Black, W., Babin, B., & Anderson, R. (2010). Multivariate data analysis a global perspective. Pearson Education Inc. *USR, New Jersey*, *7458*.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201.

Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTs: Faking, coaching, and retesting issues. *Situational judgment tests: Theory, measurement, and application*, 205-232.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International journal of human-computer interaction*, *32*(1), 51-62.

Hu, T., Zhang, D., Wang, J., Mistry, R., Ran, G., & Wang, X. (2014). Relation between emotion regulation and mental health: A meta-analysis review. *Psychological reports*, *114*(2), 341-362.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3) a brief measure of dark personality traits. *Assessment*, *21*(1), 28-41.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, *34*(1), 111-117.

Kaminski, K., Felfe, J., Schäpers, P., & Krumm, S. (2019). A closer look at response options: Is judgment in situational judgment tests a function of the desirability of response options? *International Journal of Selection and Assessment*, *27*(1), 72-82.

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. https://CRAN.R-project.org/package=factoextra

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Kranzler, A., Young, J. F., Hankin, B. L., Abela, J. R., Elias, M. J., & Selby, E. A. (2016). Emotional Awareness: A Transdiagnostic Predictor of Depression and Anxiety for Children and Adolescents. *J Clin Child Adolesc Psychol*, *45*(3), 262-269. https://doi.org/10.1080/15374416.2014.987379

Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of personality assessment*, *55*(1-2), 124-134.

Lane, R. D., & Smith, R. (2021). Levels of emotional awareness: theory and measurement of a socio-emotional skill. *Journal of Intelligence*, *9*(3), 42.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426-441.

Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and psychological measurement*, *18*(3), 437-451.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: theory and data. *Emotion*, *8*(4), 540.

Maroti, D., Lilliengren, P., & Bileviciute-Ljungar, I. (2018). The Relationship Between Alexithymia and Emotional Awareness: A Meta-Analytic Review of the Correlation Between TAS-20 and LEAS. *Front Psychol*, *9*, 453. https://doi.org/10.3389/fpsyg.2018.00453

Matthews, G., Roberts, R. D., & Zeidner, M. (2004). TARGET ARTICLES:" Seven Myths About Emotional Intelligence". *Psychological inquiry*, *15*(3), 179-196.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion review*, *8*(4), 290-300.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) item booklet. *UNH Personality Lab*, 26.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*(1), 97-105. https://doi.org/10.1037/1528-3542.3.1.97

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual review of psychology*, *60*, 577-605.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & GRUBB III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel psychology*, *60*(1), 63-91.

McRae, K., Jacobs, S. E., Ray, R. D., John, O. P., & Gross, J. J. (2012). Individual differences in reappraisal ability: Links to reappraisal frequency, well-being, and cognitive control. *Journal of Research in Personality*, *46*(1), 2-7.

McWilliams, L. A., Cox, B. J., & Enns, M. W. (2003). Use of the Coping Inventory for Stressful Situations in a Clinically Depressed Sample: Factor Structure, Personality Correlates, and Prediction of Distress. *Journal of Clinical Psychology*, *59*(4), 423-437. https://doi.org/10.1002/jclp.10080

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *PloS one*, *7*(11), e48230.

Mikolajczak, M., Luminet, O., Leroy, C., & Roy, E. (2007). Psychometric properties of the Trait Emotional Intelligence Questionnaire: Factor structure, reliability, construct, and incremental validity in a French-speaking population. *Journal of personality assessment*, *88*(3), 338-353.

Moradi, A., Pishva, N., Ehsan, H. B., & Hadadi, P. (2011). The relationship between coping strategies and emotional intelligence. *Procedia-Social and Behavioral Sciences*, *30*, 748-751.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of applied psychology*, *75*(6), 640.

Niven, K., Totterdell, P., Stride, C. B., & Holman, D. (2011). Emotion Regulation of Others and Self (EROS): The development and validation of a new individual difference measure. *Current Psychology*, *30*, 53-73.

Noorbakhsh, S. N., Besharat, M. A., & Zarei, J. (2010). Emotional intelligence and coping styles with stress. *Procedia-Social and Behavioral Sciences*, *5*, 818-822.

Nozaki, Y., & Mikolajczak, M. (2020). Extrinsic emotion regulation. *Emotion*, *20*(1), 10.

Nyquist, A. C., & Luebbe, A. M. (2020). An Emotion Recognition-Awareness Vulnerability Hypothesis for Depression in Adolescence: A Systematic Review. *Clin Child Fam Psychol Rev*, *23*(1), 27-53. https://doi.org/10.1007/s10567-019-00302-3

O'Connor Jr, R. M., & Little, I. S. (2003). Revisiting the predictive validity of emotional intelligence: Self-report versus ability-based measures. *Personality and Individual Differences*, *35*(8), 1893-1902.

Persich, M. R., Smith, R., Cloonan, S. A., Strong, M., & Killgore, W. D. S. (2021). Emotional Intelligence Training as a Protective Factor for Mental Health During the COVID-19 Pandemic. *Depression and anxiety*.

Petrides, K. V. (2011). Ability and trait emotional intelligence.

Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *British journal of psychology*, *98*(2), 273-289.

Piedmont, R. L. (2014). Inter-item correlations. *Encyclopedia of quality of life and well-being research*, 3303-3304.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Preece, D. A., Becerra, R., Robinson, K., & Gross, J. J. (2020). The Emotion Regulation Questionnaire: Psychometric Properties in General Community Samples. *J Pers Assess*, *102*(3), 348-356. https://doi.org/10.1080/00223891.2018.1564319

R Core Team. (2022). *R: A language and environment for statistical computing* In *R Foundation for Statistical Computing*. https://www.R-project.org/

Reblin, M., & Uchino, B. N. (2008). Social and emotional support and its implication for health. *Current opinion in psychiatry*, *21*(2), 201.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. In Northwestern University. https://CRAN.R-project.org/package=psych

Robinson, M. D., Persich, M. R., & Irvin, R. L. (2022). An ego effectiveness perspective of successful self-control: An individual difference and its links to social functioning and well-being. *Journal of Research in Personality*, *97*, 104207.

Robinson, M. D., & Sedikides, C. (2020). Personality and the self. . In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology (2nd ed.).* (pp. 339–351). Cambridge University Press. https://doi.org/https://doi.org/10.1017/9781108264822.031

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of educational measurement*, 335-343.

Self, D. (2004). Drug dependence and addiction: neural substrates. *Am J Psychiatry*, *161*(2), 223. https://doi.org/10.1176/appi.ajp.161.2.223

Sheppes, G., & Meiran, N. (2007). Better late than never? On the dynamics of online regulation of sadness using distraction and cognitive reappraisal. *Personality and Social Psychology Bulletin*, *33*(11), 1518-1532.

Shiota, M., & Kalat, J. (2012). Emotion.

Shu, J., Bolger, N., & Ochsner, K. N. (2021). Social emotion regulation strategies are differentially helpful for anxiety and sadness. *Emotion*, *21*(6), 1144.

Smith, R., Persich, M. R., Chuning, A. E., Cloonan, S., Woods-Lubert, R., Skalamera, J., Berryhill, S. M., Weihs, K. L., Lane, R. D., & Allen, J. J. (2023). Improvements in mindfulness, interoceptive and emotional awareness, emotion regulation, and interpersonal emotion management following completion of an online emotional skills training program. *Emotion*.

Smith, R., Persich, M. R., Cloonan, S., Woods-Lubert, R., Skalamera, J., Berryhill, S. M., Weihs, K. L., Lane, R. D., Allen, J. J. B., Dailey, N. S., Alkozei, A., Vanuk, J. R., & Killgore, W. D. S. (2023). Improvements in Mindfulness, Interoceptive and Emotional Awareness, Emotion Regulation, and Inter-Personal Emotion Management Following Completion of an Online Emotional Skills Training Program. *Emotion*. https://doi.org/https://doi.org/10.1037/emo0001237

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. MIT press.

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and psychological measurement*, *72*(1), 37-43.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & reasoning*, *20*(2), 147-168. https://doi.org/10.1080/13546783.2013.844729

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: development, cognitive abilities, and thinking dispositions. *Dev Psychol*, *50*(4), 1037-1048. https://doi.org/10.1037/a0034910

Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychological science*, *24*(12), 2505-2514.

Webb, C. A., Schwab, Z. J., Weber, M., DelDonno, S., Kipman, M., Weiner, M. R., & Killgore, W. D. S. (2013). Convergent and divergent validity of integrative versus mixed model measures of emotional intelligence. *Intelligence*, *41*(3), 149-156. https://doi.org/10.1016/j.intell.2013.01.004

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychol Bull*, *138*(4), 775-808. https://doi.org/10.1037/a0027600

Wechsler, D. (2011). WASI -II: Wechsler abbreviated scale of intelligence - second edition. *WASI*.

Wenzel, M., Rowland, Z., Weber, H., & Kubiak, T. (2020). A round peg in a square hole: Strategy-situation fit of intra-and interpersonal emotion regulation strategies and controllability. *Cognition and Emotion*, *34*(5), 1003-1009.

Wright, R., Riedel, R., Sechrest, L., Lane, R. D., & Smith, R. (2017). Sex differences in emotion recognition ability: The mediating role of trait emotional awareness. *Motivation and Emotion*, *42*(1), 149-160. https://doi.org/10.1007/s11031-017-9648-0

Zlomke, K. R., & Hahn, K. S. (2010). Cognitive emotion regulation strategies: Gender differences and associations to worry. *Personality and Individual Differences*, *48*(4), 408-413.