1-2018

Assessing the validity of emotional intelligence measures

Christopher T. H. MINERS
Queen's University

Stéphane COTE
University of Toronto

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg
DOI: https://doi.org/10.1177/1754073917693688

Published in Emotional Review, Vol. 10 No. 1 (Jan 2018) 87-95
Published version available at: http://doi.org/10.1177/1754073917693688

Assessing the Validity of Emotional Intelligence Measures

Christopher T. H. Miners
Stephen J. R. Smith School of Business, Queen's University, Canada

Stéphane Côté
Joseph L. Rotman School of Management, University of Toronto, Canada

Filip Lievens
Department of Personnel Management and Work and Organizational Psychology, Ghent University, Belgium

Abstract

We describe an approach that enables a more complete evaluation of the validity of emotional intelligence measures. We argue that a source of evidence for validity is often overlooked by researchers and test developers, namely, evidence based on response processes. This evidence can be obtained through (a) a definition of the ability, (b) a description of the mental processes that operate when a person uses the ability, (c) the development of a theory of response behaviour that links variation in the construct with variation on the responses to the items of a measure, and (d) a test of the theory of response behaviour through one or more strategies that we describe.

Keywords
emotional intelligence, validity

Emotional intelligence (EI) is the ability to reason accurately about emotions, and to use emotions and emotional knowledge to enhance thought (Mayer, Roberts, & Barsade, 2008). It was introduced just over 25 years ago (Salovey & Mayer, 1990), and popularized just over 20 years ago (Goleman, 1995). EI captured people's imaginations because of the intuitive appeal of the argument that it is an important determinant of success. The results of empirical research, however, are often inconsistent with the common intuitions about its role in various forms of success.[1] Some studies found that EI predicts important outcomes, such as the quality of interpersonal relationships (Lopes, Salovey, Côté, & Beers, 2005), leadership emergence (Emery, 2012), and transformational leadership (Rubin, Munz, & Bommer, 2005), but other studies did not find these relationships (Føllesdal & Hagtvet, 2013; Sosik & Megerian, 1999). Still other studies reported relationships between EI and other important outcomes, but only under some circumstances (Côté & Miners, 2006; Fahr, Seo, & Tesluk, 2012).

EI remains controversial with respect to its validity—and therefore its utility—due, in part, to these mixed results. The equivocal results beg the question "Why are the research findings not more consistent with our intuitions about the role that EI plays in various forms of success?" It is possible that people's intuitions are wrong, and that EI is at best a weak and inconsistent determinant of success (Grant, 2014). Alternatively, it is possible that EI is an important determinant of success, but research on it is hampered. In particular, researchers still use multiple definitions and conceptualizations of what is ostensibly EI. Furthermore, theories of the relationships between EI and criteria are still underdeveloped (Côté, 2014; Ybarra, Kross, & Sanchez-Burks, 2014).

Here we focus on another way in which research on EI is hampered. Several researchers have asserted that the measurement of EI is one of the major limitations that remains (Conte, 2005; Maul, 2012a, 2012b; Roberts et al., 2006). The evidence presented by these authors (and others) casts doubt on whether

all measures of EI are equally valid, and confirms that its measurement is a limitation of the extant research. Our concern, however, is wider than measurement alone. Our concern is also with the psychometric *theory* that underlies the validity of measures, as well as the empirical processes that can be used to determine whether the psychometric theory is supported for a particular measure.

We believe that one of the reasons a wide variety of measures continues to be used is that researchers and test developers have not heeded calls to conceptualize validity in theoretical terms, and to evaluate theories about validity more directly (Borsboom, Mellenbergh, & van Heerden, 2004; Maul, 2012b). In response to these calls, we adopt an approach to understand the concept of validity and to assess validity that draws on the innovative work of Bornstein (2011), Borsboom et al. (2004), Embretson (1983), and Messick (1989). We focus on the theoretical processes that link variation in a construct with variation on the responses to the items of a measure, and on the research methods that directly evaluate the validity of those theoretical processes. This approach to understanding and to evaluating validity is consistent with modern psychometric theories of validity, but it is often inconsistent with the practice of validity research. The approach that we describe should therefore be viewed as a complement to what, historically, has been the focus of validity research, namely, test content, internal structure, and relationships with other variables (Bornstein, 2011).

The goal of this article is to describe an approach that enables a more complete evaluation of the validity of EI measures. We first review the dominant psychometric theory of validity before we turn our attention to a source of evidence for validity that is often overlooked by researchers and test developers, namely, evidence based on response processes. We then present a sequence of steps for how to obtain this source of evidence and, therefore, for how to evaluate extant measures of EI more carefully and to improve the development of new measures of EI.

Validity

According to Messick (1989) and to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the interpretation of test scores for proposed uses of a test. From this perspective, there are not multiple types of validity but, instead, validity is a unitary concept on which different sources of evidence should be brought to bear. The *Standards* identifies five sources of evidence for validity, namely, evidence based on (1) test content, (2) item response processes, (3) internal structure, (4) relationships with other variables, and (5) the consequences of testing. The first four sources of evidence, together, constitute evidential (research-based) validity, or a measure's potential to provide accurate, useful, and unbiased information (Bornstein, 2011). The fifth source of evidence constitutes consequential (impact-based) validity, or the extent to which a measure realizes this potential in practice (Bornstein, 2011). We limit our discussion to evidential validity to preserve the focus on the interpretation of test scores independent of the consequences of their use, and hereafter *validity* should be understood as such.

The validity of measures has often been assessed in terms of the content validity of their items (i.e., *Standards*, Source 1), the psychometric properties of the measure (i.e., *Standards*, Source 3), and the relationships that are hypothesized to exist between the focal construct and the other constructs in its nomological network (i.e., *Standards*, Source 4; Bornstein, 2011). In other words, the validity of measures has rarely been assessed in terms of response processes (i.e., *Standards*, Source 2; Bornstein, 2011). We therefore believe that research designed to understand response processes should be a helpful addition to validity research in general, and to validity research in the contentious domain of EI in particular.

Response Processes

Response processes are the theoretical (mental) processes that link variation in a construct with variation on the responses to the items of a measure. Evidence for validity based on response processes is therefore obtained through identification of the theoretical processes that intervene between the presentation of the items of a measure and the responses to those items, and measurement or manipulation of the theoretical processes to evaluate their effect on the measurement outcomes (Bornstein, 2011; Embretson, 1983). If a measure captures an ability that underlies EI, then it should be possible to demonstrate that people who complete the measure engage in the mental processes that are associated with the operation of the ability. If the relevant mental processes are operative and have the theorized effect on the responses to the items (e.g., a positive relationship exists between the sophistication or the speed of their operation and the level of performance on the measure), then it increases confidence that the ability is responsible for the choice of responses to the items.

Embretson (1983) introduced the concept of construct representation, the focus of which is on response processes, to complement the way in which validity was often assessed (i.e., through an examination of nomological networks). Her work was primarily in the domain of cognitive modeling, in which she advocated for the development of theoretical models that specify the mental processes that operate during the completion of a measure, the creation of items that vary in the extent to which they tax the operation of specific mental processes, and an evaluation of the theoretical model's capacity to (accurately) predict the level of performance on each item. Researchers have since explored the response processes that are involved in the completion of traditional intelligence tests and other similar tests of abilities and aptitudes, and introduced additional methods to facilitate the evaluation of response processes (Embretson & Gorin, 2001; Kane, 2001; Mislevy, 2007).

The work of Borsboom et al. (2004) encouraged further attention to response processes. The crux of validity is whether

a measure serves as a faithful indicator of the construct that it is designed to measure or, to be cruder, whether it measures what it is supposed to measure (Kelley, 1927). Borsboom and colleagues argued that evidence for validity cannot reside in the relationships between two or more measures or, relatedly, in the relationships between the constructs in a nomological network. The pattern of relationships that defines a nomological network can provide only indirect evidence for validity, because it does not speak to the crux of validity. From their perspective, evidence for validity must reside in the relationship between a construct and the measurement outcomes. In particular, Borsboom and colleagues argued that validity must be established through substantive theory about response processes or, in other words, about the causal role of a construct in determining the value of measurement outcomes and in direct tests of the theory (see Podsakoff, Podsakoff, MacKenzie, & Klinger, 2013, for an application to the domain of personality traits).

Bornstein (2011) also focused on response processes, or what he termed a process-focused model of test score validity. He identified six categories of tests, each of which requires the operation of a different class of mental processes. For example, the scores on performance tests are derived from on-line responding to structured tasks (i.e., problem-solving), whereas the scores on self-attribution tests are derived from reflective processes that are used to determine the extent to which a person attributes personality traits, feelings, thoughts, etcetera to him/herself. He also identified four context-based influences as potential moderators (e.g., affective state) that can help to demonstrate the operation of a class of mental processes. For example, a strong affective reaction that creates a cognitive load should affect the scores on a performance test to a greater extent than the scores on a self-attribution test (thus helping to elucidate the class of mental processes in operation).

Our approach to a more complete evaluation of the validity of EI measures is grounded in the work that we have described. It uses construct representation, the refinement to the concept of validity that appears in the work of Embretson (1983), as its foundation. It embraces the work of Borsboom et al. (2004) by advocating for the measurement of the theoretical processes that link variation in a construct with variation on the responses to the items of a measure, and the work of Bornstein (2011) by advocating for the experimental manipulation of those processes. Our approach is thus, we hope, a logical synthesis and extension of the extant literature.

Our approach is also consistent with Fiori's (2009) dual-process framework for the conscious and the automatic (mental) processes associated with EI. The majority of research on EI has explicitly or tacitly focused on conscious processes. Some of the most emotionally intelligent behaviours, however, are likely to be automatic: automatic processes consume fewer cognitive resources than their conscious counterparts and therefore leave more cognitive resources to respond to the particular features of a situation. Fiori (2009) encouraged researchers to use experimental paradigms to explore the mental processes that underlie the operation of EI, in combination with the more traditional correlational analyses that are often used to understand the construct of EI. Our approach also calls for an exploration of the mental processes that underlie the operation of EI and for the use of experimental paradigms to aid in the exploration, but for a different application (i.e., to evaluate validity). The specific examples that Fiori (2009) provided to test the automatic processes that underlie EI might prove to be just as useful within the context of validity research.

Emotional Intelligence and Validity

The previous arguments lead us to believe that to help determine the validity of extant measures of EI and to improve the development of new measures of EI, it is necessary to describe the theoretical processes that intervene between the construct of EI and the responses to the items of a measure. We now present a four-step, systematic approach to help determine how variation in the construct of EI causes variation in measurement outcomes.

We focus on a subdimension of the first branch of EI (i.e., the perceiving emotion branch), emotion recognition ability. Our focus on a subdimension of EI is consistent with the need for future theoretical development to identify the subdimensions of branches that are relevant to a phenomenon (e.g., Yip & Côté, 2013). If a theory is overly inclusive with respect to the specific subdimensions, then the effects of the relevant ones will be weakened or remain undetected.

Adopting our approach highlights the need to precisely identify the mechanism(s) that link variation in an ability with variation on the responses to the items of a measure. The movement to examine the specific abilities that underlie the branches of EI, however, is nascent. There is limited theoretical development and empirical evidence to help researchers determine how, exactly, some of the specific abilities operate (see the Discussion section for further details). This is a significant gap in the EI literature that, itself, warrants more attention (Fiori, 2009; Maul, 2012b). It is even difficult to identify all of the relevant mechanisms within the relatively well-developed emotion recognition ability literature. The arguments that we present for emotion recognition ability are consistent with our understanding of the extant literature, but they should nonetheless be viewed as illustrative and, at this point in time, as plausible hypotheses rather than known facts.

Step 1: Define the Ability

Researchers who aim to establish and to test the theoretical processes that link variation in a construct with variation on the responses to the items of a measure must first establish a clear, concise, and complete definition of that construct. A definition of this sort facilitates the articulation of the theoretical processes that are involved, because it invites consideration of the processes that are likely to be relevant to the link and of the processes that are unlikely to be relevant to the link (Locke, 2003; Whetten, 1989). An unclear or incomplete definition, by contrast, invites the inappropriate inclusion or exclusion of theoretical processes. We define emotion recognition ability (ERA) as follows: "Emotion

recognition ability is the capacity of a person to accurately identify the emotions that other people express" (Ickes, Stinson, Bissonette, & Garcia, 1990; Levenson & Ruef, 1992).

Step 2: Describe the Operation of the Ability

With a clear, concise, and complete definition in place, researchers could then describe the general process that operates when a person uses the ability. First, in the case of ERA, a person devotes attention to information about the emotions of another person. The information could be verbalized (i.e., "I am angry"), or encoded in speech (e.g., vocal tone, volume, and prosody), facial expressions (e.g., position or movement of the eyebrows, direction or movement of the gaze, and contour of the mouth), bodily posture (e.g., position of the head and shoulders, placement of the arms, and angle of the torso), or bodily movement (e.g., displacement of the aforementioned anatomy; Bänziger, Mortillaro, & Scherer, 2012; Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016; Matsumoto, Hwang, & Frank, 2016).

Second, the person must compare this information to stored information (i.e., concepts or templates) about which emotion, and which level of intensity of the emotion, are associated with a particular configuration of speech, facial expression, bodily posture, and bodily movement (Adolphs, 2002). The accuracy of the stored information, and the speed with which it is retrieved, assembled, and compared, will vary from person to person as a function of an ability to discern information about emotions from social interactions (Nisbett et al., 2012), and as a function of formal and informal instruction about the role of emotions in interpersonal relationships (Elfenbein, 2006; Saarni, 1999).

Third, the person must interpret the emotion and the intensity of the emotion within the specific context in which the emotion is unfolding (Trope, 1986). The context is likely to be viewed and understood through multiple, complementary lenses: from an understanding of the norms of a culture to the voice climate of a group to the closeness of our relationship with the other person (Clark & Finkel, 2005; Matsumoto et al., 2016; Morrison, Wheeler-Smith, & Kamdar, 2011).

Step 3: Develop a Theory of Response Behaviour

A description of the general process that operates when a person employs an ability allows researchers to hypothesize the specific theoretical mechanisms that link variation in a construct with variation on the responses to the items of a measure, and thus develop a theory of response behaviour.

The description of the process that operates when a person employs his or her ERA explicitly or implicitly suggests that pattern recognition, speed of recall from memory, working memory capacity, and the depth and the breadth of knowledge about the display of emotions should all contribute to accurate observations during a social interaction and to choosing the correct responses to the items of a measure (Adolphs, 2002; Lynn et al., 2016; Nook, Lindquist, & Zaki, 2015; Zaki, Weber, Bolger, & Ochsner, 2009). It therefore stands to reason, for example, that people with a high level of ERA should engage in more of the behaviours that enable accurate pattern recognition than people with a lower level of this ability (or in the same behaviours as them, but with greater efficiency). The increase in this behaviour or in its efficiency, in turn, should lead to more accurate observations during a social interaction and faster (accurate) observations and, similarly, to the choice of more accurate responses to the items of a measure and faster (accurate) responses.

Step 4: Test the Theory of Response Behaviour

The identification and articulation of the mechanisms that underlie an ability would enable researchers to generate a list of measurable mental activities, attributes, and processes. This list would then facilitate tests of whether variation in the ability results in the predicted variation on responses to the items of a measure through a particular mechanism (i.e., ERA and pattern recognition in the running example).

We outline three broad, complementary strategies that can be used to determine whether a particular mechanism links variation in a construct with variation on responses to the items of a measure: measurement of mediation, moderation of process, and evaluation of alternatives. We also provide examples of the application of the strategies to one of the most popular measures of EI (and therefore of ERA within the domain of EI research), the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002), to illustrate their application.

Step 4a: Identification of Multiple Levels of the Focal Construct

Before we proceed, we discuss a requirement that is central to each of the strategies; namely, the identification of multiple levels of the focal construct (in our case, ERA). Researchers can identify multiple levels of the focal construct through an initial assessment that identifies people who differ with respect to their level of the ability. After identifying multiple levels of the focal construct, it is then possible to administer the focal measure and, in turn, to determine whether variation in the level of the focal construct results in the predicted variation on responses to the items of the focal measure (i.e., whether people with a high level of the ability outperform people with a lower level of the ability on the focal measure).

The initial assessment could involve the administration of one or more extant measures of the focal construct, if such measures are available (e.g., Diagnostic Analysis of Nonverbal Accuracy: Nowicki, & Duke, 1994; Geneva Emotion Recognition Test: Schlegel, Grandjean, & Scherer, 2014). While it requires empirical validation, it might also be possible to administer extant measures of the more basic abilities that underlie the focal construct (i.e., Stratum 1 factors in Carroll's [1993] model of intelligence). For example, it is difficult to imagine that some of the basic abilities that underlie the broad ability of visual perception (e.g., spatial relations) and the broad ability of auditory

perception (e.g., speech sound discrimination) do not also underlie ERA. The initial assessment could also involve interviews or other procedures conducted by subject matter experts (e.g., clinical identification of individuals with and without disorders that are associated with emotional processing deficits, such as schizophrenia; Derntl et al., 2009). The use of more than one approach for the initial assessment, if possible, and of structural equation modeling would be desirable. It would prevent reliance on any one measure to capture the focal construct, and allow for the control of random and systematic measurement error.

Step 4b: Strategies to Test a Theory of Response Behaviour

We now describe the three strategies that can be used to determine whether a particular mechanism links variation in a construct with variation on the responses to the items of a measure.

Strategy 1: Measurement of mediation. The first strategy requires researchers to identify multiple levels of the focal construct and to measure the theorized mechanism(s) that link variation in the construct with variation on responses to the items of a measure (Spencer, Zanna, & Fong, 2005). The identification of multiple levels of the focal construct could be accomplished through one or more of the approaches described in Step 4a. A measure of the mechanism(s) and, in turn, the focal measure could then be administered. Mediation could be formally tested using the procedures advocated by Preacher and Hayes (2004) or MacKinnon, Lockwood, Hoffman, West, and Sheets (2002). This strategy therefore leads to confirmation that a high level of an ability leads to better performance on the focal measure than a lower level of the ability, and that the theorized mechanism(s) help us to understand why a high level of the ability results in better performance on the focal measure.

For the Faces Task of the MSCEIT (Mayer, Salovey, & Caruso, 2002), eye-tracking could be used as an index of the mechanism in the running example, pattern recognition. The hypothesis is that people with a high level of ERA will focus on the relevant features of the stimulus that is presented in each item of the measure (i.e., on the specific muscle movements captured in a photograph of a facial expression that, together, characterize the emotion that is present) and neglect the irrelevant features of the stimulus. It is also possible that people focus on the same (relevant) features of the stimulus regardless of their ERA, but people with a high level of ERA enact the behaviour faster (i.e., more efficiently). According to the theorized mechanism, the more precise (or faster) pattern of eye movements of people with a high level of ERA will enable them to choose the correct responses to the items of the focal measure more often (or more quickly). The hypothesized relationships between ERA, pattern recognition (as indexed by eye-tracking), and performance on the Faces Task would then be tested for mediation.

Strategy 2: Moderation of process. The second strategy involves the identification of multiple levels of the focal construct and the subsequent manipulation of a moderator to determine if the moderator changes the relationship between the construct and the responses to the items of a measure (Spencer et al., 2005). The second strategy does not require the mechanism to be measured. The second strategy, instead, requires the identification of a moderator that enhances or diminishes the operation of the mechanism and, importantly, only the mechanism (otherwise the interpretation of any effect of the moderator becomes ambiguous). This strategy is helpful when the mechanism is difficult or even impossible to measure in the context of a particular study (Spencer et al., 2005). If the moderator enhances (or diminishes) the effect of the mechanism, then there should be a weaker relationship between the focal construct and the measurement outcomes, because the mechanism will be active (or inactive) regardless of the level of the focal construct. Moderation could be formally tested using the procedures advocated by Aiken and West (1991). The application of this strategy therefore leads to confirmation that a moderator can weaken the relationship between the ability and the level of performance on the focal measure through its capacity to facilitate or interfere with the operation of the theorized mechanism(s).

For example, in a control condition, the photographs of facial expressions from the Faces Task of the MSCEIT would be presented as normal. There should be a significant difference between the performance of people with a high level of ERA and people with a lower level of ERA. In an experimental condition, the photographs of facial expressions from the Faces Task would be presented upside down. Past research has shown that this manipulation interferes with the processing of facial features—the hypothesized mechanism in this example (Derntl, Seidel, Kainz, & Carbon, 2009). There should be a smaller difference between the performance of people with a high level of ERA and people with a lower level of ERA in the experimental condition (i.e., when the moderator is introduced). The scores of the former are likely to drop substantially, whereas the ordinarily low scores of the latter are unlikely to drop as far. People with a high level of ERA and people with a low level of ERA are therefore likely to obtain similar scores. The hypothesized relationships between ERA, the moderator (i.e., experimental condition), and performance on the Faces Task would then be tested for moderation. If the anticipated pattern of results is obtained, then it would suggest that pattern recognition is an operative mechanism, as theorized.

Strategy 3: Evaluation of alternatives. The third strategy requires researchers to identify multiple levels of the focal construct, and to evaluate alternative explanations for a particular mechanism that links variation in a construct with variation on the responses to the items of a measure. There are likely to be a number of mechanisms that compete with or complement the focal mechanism (i.e., the mechanism that was identified through the development of a theory of response behaviour) with respect to its role of carrying (at least some of) the variation in a construct to variation on the responses to the items of a measure. It is therefore important to demonstrate that the focal mechanism is operative even when the alternative mechanisms are incorporated into an experimental paradigm or controlled

for in a statistical model. Researchers would need to identify multiple levels of the focal construct, administer an experimental paradigm that incorporates the focal mechanism *and* the alternative mechanism(s) (or measures of the focal mechanism a*nd* the alternative mechanism[s]), and finally administer the focal measure. Complex mediation models that can accommodate more than one mediator could be formally tested using the procedures advocated by Williams and MacKinnon (2008). Researchers would then be able to determine whether the focal mechanism is still operative even in the presence of the alternative mechanism(s). In other words, the application of this strategy requires confirmation that a high level of an ability results in better performance on the focal measure than a lower level of the ability through the theorized mechanism(s), and not exclusively through the operation of different mechanisms.[2]

For example, in the case of pattern recognition, it could be argued that the operative mechanism is, instead, increased attention to or concentration on the stimulus. This would suggest that the eye movements of people with a high or a low level of ERA are indistinguishable; it is just that the former spends more time doing it than the latter and this explains their superior performance on the Faces Task of the MSCEIT. This alternative explanation could be ruled out by a direct comparison of the pattern of eye movements exhibited by people who differ with respect to their level of ERA to verify that the patterns are different. The pattern of people with a high level of ERA should be more precise than the pattern of people with a lower level of it, and the average response latency of the former should be shorter (as per the theorized mechanism), not longer (as per the alternative explanation), than the latter.

Discussion

To improve the validity of measures of EI, we put a strong emphasis on developing a substantive theory about the theoretical processes that link variation in a construct with variation on the responses to the items of a measure and direct tests of the theory. Drawing on the innovative work of Embretson (1983), Messick (1989), Borsboom et al. (2004), and Bornstein (2011), we argued that research on response processes is an important but often neglected aspect of validity research. To promote research of this sort, we described a four-step approach that should help to determine whether extant measures of EI are valid and to improve the development of new measures of EI. We summarize the four steps that constitute the approach in Table 1.

The approach that we described should be helpful in a number of ways. It will help researchers and test developers to assess the validity of extant measures of EI, as noted, and to create (a) variations of extant measures of EI that are designed to facilitate their administration (e.g., short forms), (b) the next generation of extant measures of EI, and (c) new measures of new abilities that are proposed to underlie EI. All of this should have a knock-on effect, in so far as the introduction of new, improved measures of EI will build the confidence of researchers and invite further exploration of its role in various forms of success (and failure). The approach, ultimately, should facilitate an improvement to the quality of research on EI through an improvement of its measurement.

Caveats

Our four-step approach provides a range of possibilities to determine whether a mechanism links variation in a construct with variation on the responses to the items of a measure. For example, consider its application to the ability to select emotion regulation strategies (ASERS) during the development of a new measure. Step 1: ASERS is the capacity of a person to choose strategies that are likely to create desired emotions (Côté, 2014). Imagine that for each item of the new measure, respondents are presented with a description of a situation in which they are (ostensibly) interacting with another person, an emotion regulation goal for that situation, and a short video in which the person with whom they are (ostensibly) interacting addresses them. Respondents must choose the response option that maximizes the likelihood of achieving the emotion regulation goal. Step 2: The operation of ASERS requires mental representation of the range of emotion regulation strategies that exist, consideration of the features of the strategies and the situation to determine the likelihood of each strategy achieving the goal, and awareness of the resources that are available (e.g., mental energy; Aldao, 2013; Gross, 2015; Saunders, Milyavskaya, & Inzlicht, 2015; Sheppes, Scheibe, Suri, Radu, & Gross, 2014). Step 3: Speed of recall from memory, executive function, fluid intelligence, and the depth and the breadth of knowledge about emotion regulation strategies should all contribute to choosing the correct responses to the items of the new measure (Messina, Bianco, Sambin, & Viviani, 2015; Ochsner & Gross, 2008; Optiz, Lee, Gross, & Urry, 2014; Zelazo & Cunningham, 2007). Step 4a: The initial assessment of ASERS could involve, for example, the administration of extant measures of the ability (e.g., Situational Test of Emotion Management; MacCann & Roberts, 2008), and the clinical identification of individuals with and without disorders that are associated with emotion dysregulation (e.g., social anxiety disorder; Gross, 2015). Step 4b, Strategy 1: measurement of mediation: High-frequency heart rate variability (HF-HRV) could be used as an index of one potential underlying mechanism, executive function. HF-HRV is a measure of autonomic contributions to cardiac functioning that decreases when people complete tasks that recruit executive function, including emotion regulation tasks (Elliot, Payen, Brisswalter, Cury, & Thayer, 2011). The decrease, in turn, is associated with a lower level of performance on such tasks (Elliot et al., 2011; Thayer, Hansen, Saus-Rose, & Johnsen, 2009; Thayer & Lane, 2000). We hypothesize that people with a high level of ASERS will exhibit a smaller decrease in HF-HRV than people with a lower level of the ability when the emotion evoked by a video requires regulation. The smaller decrease in HF-HRV of people with a high level of ASERS, which reflects more efficient and effective cognitive functioning, should enable them to choose the correct responses to the items of the new measure more often. The hypothesized relations between ASERS, executive function (as

Table 1. Checklist for the evaluation of validity based on response behaviour.

- □ **Step 1: Define the ability**
- □ **Step 2: Describe the operation of the ability**
  Describe the mental processes that operate when a person uses the ability
- □ **Step 3: Develop a theory of response behaviour**
  Use the description of the mental processes to identify the mechanism(s) through which variations in the ability cause variations in the item responses on the focal measure
- □ **Step 4: Test the theory of response behaviour**
  Test the mechanism(s) through which variations in the ability cause variations in the item responses on the focal measure using one or more of the following strategies:

Step 4a: Identify multiple levels of the focal construct

- □ **Conduct an initial assessment**
  Conduct an initial assessment that identifies people who differ with respect to their level of the ability

Step 4b: Strategies to test a theory of response behaviour

- □ **Measurement of mediation**
  Measure the proposed mechanism(s) and then test whether it mediates the relationship between the ability and the item responses on the focal measure
- □ **Moderation of process**
  Test whether a moderator alters the effect of variation in the ability on variation in the item responses on the focal measure
- □ **Evaluation of alternatives**
  Evaluate whether the focal mechanism continues to serve as a mediator when alternative mechanisms are taken into consideration

indexed by HF-HRV), and performance on the new measure would be tested for mediation. Support for the hypothesized relationships would indicate that there is evidence for the validity of the new measure based on response processes.

At the same time, however, we acknowledge that there is a limit to the range of possibilities offered by our approach. There are some abilities that might require the identification of people who differ with respect to their level of the ability, but for which there are a limited number of (valid) extant measures and a limited number of alternative means to identify them.

Challenges

As previously noted, we acknowledge that it is currently difficult to identify the mechanism(s) that link variation in some of the specific abilities that underlie the branches of EI to the responses on a focal measure. There is limited theoretical development and empirical evidence to help researchers determine how, exactly, some of the specific abilities operate. The approach described by Bornstein (2011) may be helpful in this regard, because the approach does not require the measurement of mental processes. His approach, instead, relies on the nature of the test itself to identify the mental processes that are likely to be operative, and on context-based influences as potential moderators that can help to demonstrate the operation of the mental processes.

Conclusion

We are keenly aware of the challenge that is presented by research of the sort that we have described. This challenge helps to explain why the arguments presented by some scholars (e.g., Bornstein, 2011; Borsboom et al., 2004; Embretson, 1983; Messick, 1989) have shaped modern psychometric theories of validity more than they have shaped the practice of validity assessment. As Bornstein (2011) noted, it is difficult to map the idealized theories of validity advocated by psychometricians onto the messy practice of validity research. We nonetheless believe that it is important to provide evidence of this sort when possible, and to strive to obtain more evidence as the science of EI evolves and more evidence becomes available.

We hope that, above all else, we have raised awareness of an important issue for researchers and test developers in general and within the domain of EI research in particular. The four-step approach that we described and the application of the strategies that appear under Step 4b represent an early attempt to explain how to obtain a "new" source of evidence for validity in the domain of EI. As such, we recognize that they are imperfect. We nonetheless believe that it is important to obtain evidence for validity that is based on response processes, in addition to the evidence that is typically obtained. It will help to elucidate the processes that underlie the operation of EI, and enable more stringent tests of and therefore stronger conclusions about the validity of its measures.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1 We limit our discussion of EI to ability models for reasons that are described elsewhere (e.g., Côté, 2014).

2 Consistent with the arguments presented by Borsboom et al. (2004) about the limitations of nomological networks as a source of evidence for validity, the evaluation of alternative explanations is necessary but insufficient to establish validity. The third strategy therefore complements the first two strategies.

References

Adolphs, R. (2002). Recognizing emotions from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*, 21–62. doi:10.1177/1534582302001001003

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: SAGE.

Aldao, A. (2013). The future of emotion regulation research: Capturing context. *Perspectives on Psychological Science*, *8*, 155–172. doi:10.1177/1745691612459518

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal Expression Corpus for experimental research on emotion perception. *Emotion*, *12*, 1161–1179. doi:10.1037/a0025827

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, *23*, 532–544. doi:10.1037/a0022402

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. doi:10.1037/0033–295X.111.4.1061

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Clark, M. S., & Finkel, E. J. (2005). Willingness to express emotion: The impact of relationship type, communal orientation, and their interaction. *Personal Relationships*, *12*, 169–180.

Conte, J. M. (2005). A review and critique of emotional intelligence measures. *Journal of Organization Behavior*, *26*, 433–440. doi:10.1002/job.319

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, *16*, 117–128.

Côté, S. (2014). Emotional intelligence in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 459–488. doi:10.1146/annurev-orgpsych-031413–091233

Côté, S., & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, *51*, 1–28.

Derntl, B., Finkelmeyer, A., Toygar, T. K., Hülsmann, A., Schneider, F., Falkenberg, D. I., & Habel, U. (2009). Generalized deficit in all core components of empathy in schizophrenia. *Schizophrenia Research*, *108*, 197–206. doi:10.1016/j.schres.2008.11.009

Derntl, B., Seidel, E.-M., Kainz, E., & Carbon, C.-C. (2009). Recognition of emotion expressions is affected by inversion and presentation time. *Perception*, *38*, 1849–1862. doi:10.1068/p6448

Elfenbein, H. A. (2006). Learning in emotion judgments: Training and the cross-cultural understanding of facial expressions. *Journal of Nonverbal Behavior*, *30*, 21–36. doi:10.1007/s10919–005–0002-y

Elliot, A. J., Payen, V., Brisswalter, J., Cury, F., & Thayer, J. F. (2011). A subtle threat cue, heart rate variability, and cognitive performance. *Psychophysiology*, *48*, 1340–1345. doi:10.1111/j.1469–8986.2011.01216.x

Embretson, S. E. (1983). Construct validity: Construct representation v nomothetic span. *Psychological Bulletin*, *93*, 179–197. doi:10.1037/0033–2909.93.1.179

Embretson, S. E., & Gorin, J. (2001). Improving construct validity through cognitive psychology principles. *Journal of Educational Measurement*, *38*, 343–368. doi:10.1111/j.1745–3984.2001.tb01131.x

Emery, C. (2012). Uncovering the role of emotional abilities in leadership emergence. A longitudinal analysis of leadership networks. *Social Networks*, *34*, 429–437. doi:10.1016/j.socnet.2012.02.001

Fahr, C. I. C.C., Seo, M.-G., & Tesluk, P. E. (2012). Emotional intelligence, teamwork effectiveness, and job performance: The moderating role of job context. *Journal of Applied Psychology*, *97*, 890–900. doi:10.1037/a0027377

Fiori, M. (2009). A new look at emotional intelligence: A dual-process framework. *Personality and Social Psychology Review*, *13*, 21–44. doi:10.1177/1088868308326909

Føllesdal, H., & Hagtvet, K. (2013). Does emotional intelligence as ability predict transformational leadership? A multilevel approach. *The Leadership Quarterly*, *24*, 747–762. doi:10.1016/j.leaqua.2013.07.004

Goleman, D. (1995). *Emotional intelligence*. New York, NY: Bantam.

Grant, A. (2014). *Emotional intelligence is overrated*. Retrieved from https://www.linkedin.com/pulse/20140930125543-69244073-emotional-intelligence-is-overrated

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*, 1–26. doi:10.1080/1047840X.2014.940781

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*, 730–742. doi:10.1037/0022–3514.59.4.730

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*, 319–342. doi:10.1111/j.1745–3984.2001.tb01130.x

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: Oxford University Press.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, *63*, 234–246. doi:10.1037/0022–3514.63.2.234

Locke, E. A. (2003). Good definitions: The epistemological foundation of scientific progress. In J. Greenberg (Ed.), *Organizational behavior: The state of the science* (2nd ed., pp. 415–444). Mahwah, NJ: Erlbaum.

Lopes, P. N., Salovey, P., Côté, S., & Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion*, *5*, 113–118. doi:10.1037/1528–3542.5.1.113

Lynn, S. K., Ibagon, C., Bui, E., Palitz, S. A., Simon, N. A., & Barrett, L. F. (2016). Working memory capacity is associated with optimal adaptation of response bias to perceptual sensitivity in emotion perception. *Emotion*, *16*, 155–163. doi:10.1037/emo0000111

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*, 540–551. doi:10.1037/a0012746

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104. doi:10.1037/1082–989X.7.1.83

Matsumoto, D., Hwang, H. C., & Frank, M. G. (Eds.). (2016). *APA handbook of nonverbal communication*. Washington, DC: American Psychological Association.

Maul, A. (2012a). Examining the structure of emotional intelligence at the item level: New perspectives, new conclusions. *Cognition and Emotion*, *26*, 503–520. doi:10.1080/02699931.2011.588690

Maul, A. (2012b). The validity of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) as a measure of emotional intelligence. *Emotion Review*, *4*, 394–402. doi:10.1177/1754073912445811

Mayer, J. D., Roberts, R., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507–536. doi:10.1146/annurev.psych.59.103006.093646

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer–Salovey–Caruso Emotional Intelligence Test*. Toronto, Canada: Multi-Health Systems.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.

Messina, I., Bianco, S., Sambin, M., & Viviani, R. (2015). Executive and semantic processes in the reappraisal of negative stimuli: Insights from a meta-analysis of neuroimaging studies. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00956

Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*, 463–469. doi:10.3102/0013189X07311660

Morrison, E. W., Wheeler-Smith, S. L., & Kamdar, D. (2011). Speaking up in groups: A cross-level study of group voice climate and voice. *Journal of Applied Psychology*, *96*, 183–191. doi:10.1037/a0020744

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159. doi:10.1037/a0026699

Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion*, *15*, 569–578. doi:10.1037/a0039166

Nowicki, S., & Duke, M. P. (1994) Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*, 9–35. doi:10.1007/BF02169077

Ochsner, K. N., & Gross, J. J. (2008). Cognitive emotion regulation. *Current Directions in Psychological Science*, *17*, 153–158. doi:10.1111/j.1467–8721.2008.00566.x

Optiz, P. C., Lee, I. A., Gross, J. J., & Urry, H. L. (2014). Fluid cognitive ability is a resource for successful emotion regulation in older and younger adults. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00609

Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Klinger, R. L. (2013). Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. *Journal of Applied Psychology*, *98*, 99–113. doi:10.1037/a0029570

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*, *36*, 717–731. doi:10.3758/BF03206553

Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the validity of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) with established emotions measures. *Emotion*, *6*, 663–669. doi:10.1037/1528–3542.6.4.663

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, *48*, 845–858. doi:10.5465/AMJ.2005.18803926

Saarni, C. (1999). *The development of emotional competence*. New York, NY: The Guilford Press.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, *9*, 185–211. doi:10.2190/DUGG-P24E-52WK-6CDG

Saunders, B., Milyavskaya, M., & Inzlicht, M. (2015). Variation in cognitive control as emotion regulation. *Psychological Inquiry*, *26*, 108–115. doi:10.1080/1047840X.2015.962396

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, *26*, 666–672. doi:10.1037/a0035246

Sheppes, G., Scheibe, S., Suri, G., Radu, P., & Gross, J. J. (2014). Emotion regulation choice: A conceptual framework and supporting evidence. *Journal of Experimental Psychology: General*, *143*, 163–181. doi:10.1037/a0030831

Sosik, J. J., & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance: The role of self–other agreement on transformational leadership perceptions. *Group & Organization Management*, *24*, 367–390. doi:10.1177/1059601199243006

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845–851. doi:10.1037/0022–3514.89.6.845

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, *37*, 141–153. doi:10.1007/s12160–009–9101-z

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*, 201–216. doi:10.1016/S0165–0327(00)00338–4

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Bulletin*, *93*, 239–257. doi:10.1037/0033–295X.93.3.239

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*, 490–495.

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling: An Interdisciplinary Journal*, *15*, 23–51. doi:10.1080/10705510701758166

Ybarra, O., Kross, E., & Sanchez-Burks, J. (2014). The "big idea" that is yet to be: Toward a more motivated, contextual, and dynamic model of emotional intelligence. *Academy of Management Perspectives*, *28*, 93–107. doi:10.5465/amp.2012.0106

Yip, J., & Côté, S. (2013). The emotionally intelligent decision-maker: Emotion understanding ability reduces the effect of incidental anxiety on risk-taking. *Psychological Science*, *24*, 48–55. doi:10.1177/0956797612450031

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, *106*, 11382–11387. doi:10.1073/pnas.0902666106

Zelazo, P. D., & Cunningham, W. A. (2007). Executive function: Mechanisms underlying emotion regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation* (1st ed., pp. 135–158). New York, NY: Guilford.