


Overarching Principles for the Organization of Socioemotional Constructs

Sally Olderbak  and Oliver Wilhelm

Institute of Psychology and Education, Ulm University

Current Directions in Psychological Science
2020, Vol. 29(1) 63–70
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0963721419884317
www.psychologicalscience.org/CDPS



Abstract

Psychological scientists have intensively studied how people handle emotions and navigate social situations for more than a century. However, advancements in our understanding of socioemotional constructs have been hampered because of challenges in assessment. Several measurement problems have been identified; however, we want to bring attention to a potentially larger problem. Many operationalizations and measures of socioemotional constructs are poorly embedded within the larger body of psychological research, hampered by jingle and jangle fallacies. *Jingle fallacies* occur when assessment tools are assumed to measure the same construct but in practice measure different constructs. *Jangle fallacies* occur when assessment tools are assumed to measure different constructs but in practice measure the same construct. Both fallacies are primarily due to a qualitative divide between a construct's definition and how it was measured. We discuss this issue, identify examples of jingle and jangle fallacies, and conclude with recommendations.

Keywords

jingle fallacy, jangle fallacy, individual differences, socioemotional constructs, measurement

Scientific interest in the ways that people differ in their personal tendencies and capacities regarding the handling of emotion and navigation of social situations has been examined for more than a century (e.g., Moss, Hunt, Omwake, & Ronning, 1927). Terms used in this field of research include *inter-* and *intrapersonal intelligence* (Gardner, 2000), *emotional intelligence*, *personal intelligence* (Mayer, Caruso, & Salovey, 2016), and *social intelligence* (Thorndike, 1920). We refer to such a set of constructs more generally as *socioemotional constructs*.

We provide an overview of some socioemotional constructs in Figure 1, focusing on those involving the processing of emotion. Subdimensions within these constructs, such as specific emotion-regulation strategies, are excluded from the figure. Constructs are grouped by commonly used descriptions.

To create a taxonomy of these constructs, we begin by applying a prevailing classification of individual-differences construct types, specifically (a) *typical behavior*, representing participants' everyday behavior or preferences, and (b) *maximal effort* (or *abilities*), representing participants' best performance under ideal circumstances and motivation (Cronbach, 1949). Typical behavior is usually assessed by participants via self-report questionnaires or by external raters (e.g., peers)

who rate their agreement with several statements or adjectives describing the participant. Maximal effort is usually assessed after participants complete several trials, and their performance is compared with an objectively correct set standard (e.g., a cognitive test). Trials vary in difficulty level, allowing researchers to differentiate persons with a range of abilities. The extent to which individual measures assess either construct type varies. We provide examples of each type in the sections below.

This distinction is essential because it affects what is measured. For example, self-report measures of emotional intelligence capture typical-behavior emotional intelligence, or one's perspective about one's emotional abilities, and cannot capture objective abilities.

Challenges

There are several specific challenges to the study of socioemotional constructs, which have been thoroughly reviewed elsewhere¹ and which we highly recommend.

Corresponding Author:

Sally Olderbak, Ulm University, Institute of Psychology and Education, Albert-Einstein-Allee 47, 89081 Ulm, Germany
E-mail: sally.olderbak@uni-ulm.de



Fig. 1. Some of the many postulated socioemotional constructs (white and green ovals) organized around commonly used descriptions (brown ovals). These constructs are postulated or measured as representing typical behavior (white), maximal effort (dark green), or both (light green) and can be divided into lower-level facets or factors (e.g., specific emotion-regulation strategies). This is only a higher-order overview.

We focus here on an often neglected but similarly existential set of problems referred to as *jingle fallacies* and *jangle fallacies*. These fallacies indicate a disconnect between what proponents of a measure think they are measuring and what validity evidence is showing. A jingle fallacy is when the same label is used but, in practice, different constructs are measured. The “unthinking acceptance of verbal equality [is treated] as proof of real equality” (Thorndike, 1904, p. 11). The opposite, when different labels are used but, in practice, the same construct is measured, is termed a jangle fallacy. As originally defined, this was “two separate words or expressions covering in fact the same basic situation, but sounding different, as though they were in truth different” (Kelley, 1927, p. 64).

We have written about these fallacies for specific socioemotional constructs in regard to trait and ability emotional intelligence (Wilhelm, 2005) and in the evaluation of theories proposing a connection between

emotion perception and empathy (Olderbak & Wilhelm, 2017). Here, we discuss the larger problem of jingle and jangle fallacies affecting socioemotional constructs generally, discuss causes of these fallacies, provide more examples, and offer more recommendations for addressing the problem.

Jingle Fallacy: Same Label but Different Constructs Assessed

To illustrate jingle fallacies, we focus on work done in the field of emotion perception, highlighting three common causes. First, jingle fallacies occur when researchers do not consider their measurement approach. For example, measures of emotion perception include typical-behavior self-report questionnaires (see Fig. 2a) and maximal-effort tests (see Figs. 2b and 2c). The differences in measurement approach, such as asking participants to self-report their abilities (Fig. 2a) or asking

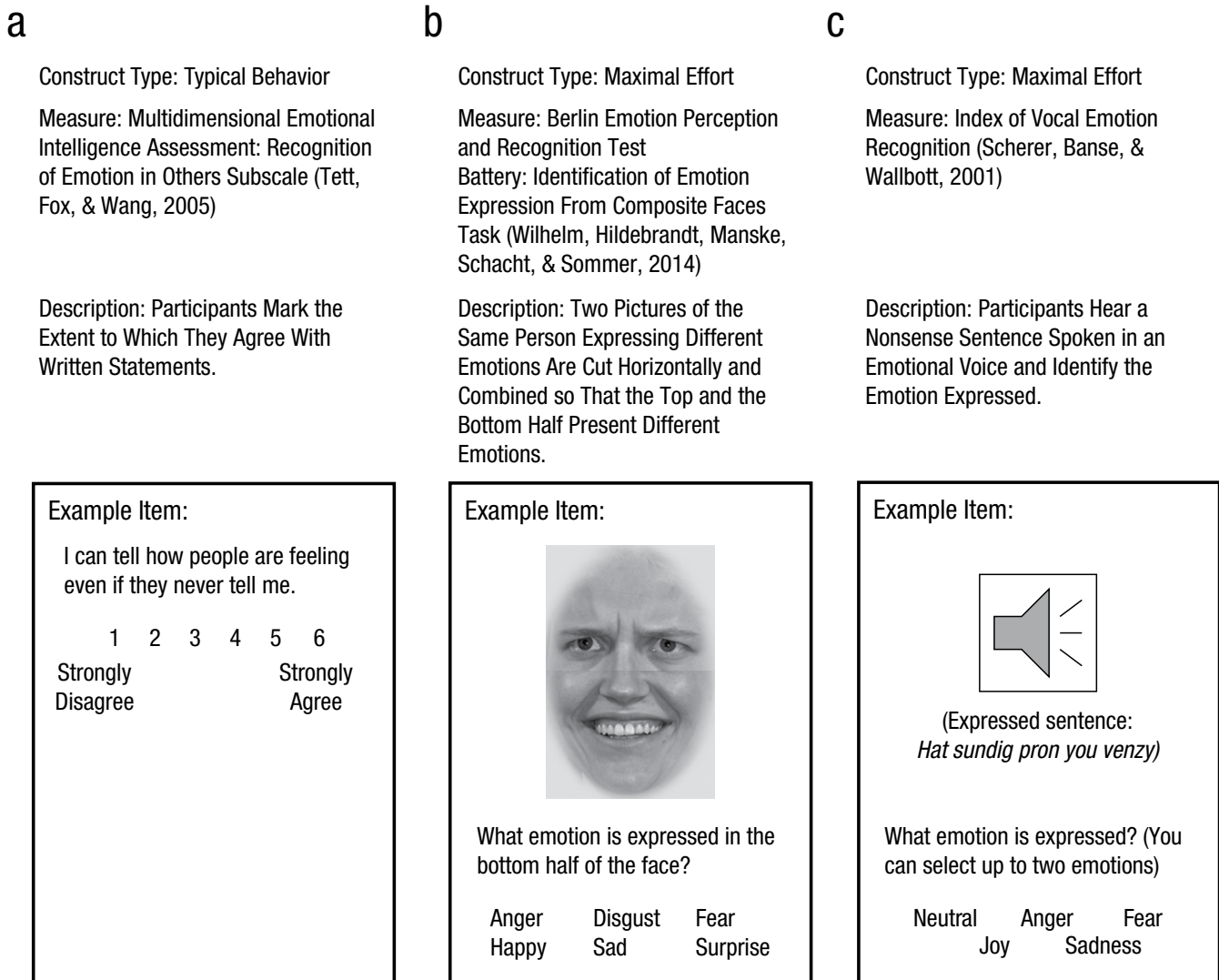


Fig. 2. Example of a jingle fallacy, which occurs when several assessment tools use the same label but measure different constructs. In this example, each of the three instruments is intended to measure emotion perception. However, the different measurement approaches mean that the measure shown in (a) is assessing typical behavior, whereas the measures shown in (b) and (c) are assessing maximal effort. Additionally, because the measure shown in (b) uses only facial stimuli, it has a different test design than the measure shown in (c), which uses only voices. The weak correlations between these measures indicate that they do not measure the same construct.

participants to identify the emotion expressed by a face (Fig. 2b), mean that researchers assess different constructs, in this case self-perceived emotion-perception abilities (i.e., typical behavior; Fig. 2a) and the ability to identify emotion expressed in the face (i.e., maximal effort; Fig. 2b). If the same construct were being assessed, then these measures would be highly correlated with each other. However, they are weakly related (i.e., correlations less than .30), which tells us that very different constructs are being assessed (e.g., Elfenbein, Jang, Sharma, & Sanchez-Burks, 2017).

When the construct type is the same, jingle fallacies can still occur because tests differ in their design. For example, tests from the Berlin Emotion Perception and Recognition Test battery use only facial stimuli (Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014),

thus relying on perceptual skills. On the other hand, the Index of Vocal Emotion Recognition uses only voices (Scherer, Banse, & Wallbott, 2001), thus relying on auditory skills. As has been supported empirically, these tests assess different abilities (Schlegel, Grandjean, & Scherer, 2012).

Jingle fallacies also occur because researchers have idiosyncratic definitions of a construct. For example, two typical behavioral measures of emotion perception—the Emotion Perception subscale of the Trait Emotional Intelligence Questionnaire and the Recognition of Emotion in Self and Recognition of Emotion in Others subscales of the Multidimensional Emotional Intelligence Assessment—reflect differences in definitions. In the former, emotion perception is defined as being clear about one’s own feelings and those of another (Petrides,

2009). In the latter, emotion perception is defined as having two separate dimensions: (a) being in touch with one's own feelings and (b) attending to the nonverbal emotional cues of other people (Tett, Fox, & Wang, 2005).

Jingle fallacies affect the interpretation of validity evidence. For example, depending on the maximal-effort emotion-perception tests administered, one finds either no relation or strong relations of emotion perception with fluid or crystallized intelligence. Specifically, tests that use person stimuli, such as emotional faces or voices, are strongly correlated with fluid and crystallized intelligence, whereas tests with pictures of inanimate objects, such as stones or a picture of scenery, are weakly related (Olderbak, Semmler, & Doebler, 2018). This is troubling and disconcerting given that these tests are frequently used in clinical settings in which poor performance is an indicator of certain disorders, such as autism spectrum disorder (American Psychiatric Association, 2013).

Jingle Fallacy: Different Labels but Same Construct Assessed

Let us now turn to the jangle fallacy, which can occur for two reasons. First, jangle fallacies occur when new constructs and measures are not adequately compared with related or established constructs. For example, emotion perception, empathic accuracy, and theory of mind have distinct definitions and are proposed to describe distinct psychological processes (see Fig. 3 for construct definitions). However, in practice, tests of individual differences usually use the same test design: a maximal-effort test in which participants are presented with an emotional face of an unfamiliar person, usually displayed on screen. Then they are asked to judge what that person is thinking or feeling (see Fig. 3 for examples of common tests). As illustrated in Figure 3, the overall approach and procedure for assessing these constructs is essentially the same. Thus, the question instead becomes, Why should the assessed constructs differ from one another?

By using a similar test design, one should not be surprised that performance on these tests is strongly related (e.g., Olderbak et al., 2015) and that the same brain regions are activated when these tasks are being completed (Mitchell & Phillips, 2015). Other things being equal, one also finds the same predictive validity for such measures. For example, researchers have found that the hormone oxytocin improves performance for each of the tests presented in Figure 3, separately concluding that oxytocin improves empathic accuracy (Bartz et al., 2010), emotion perception (Shahrestani, Kemp, & Guastella, 2013), and theory of mind (Domes, Heinrichs, Michel, Berger, & Herpertz, 2007).

We argue that all three approaches allow one to successfully complete the trials with basic emotion-perception skills (i.e., perceiving what is expressed). It is not evident why empathic accuracy or theory of mind are additionally needed or required to succeed in the tests. Thus, we conclude that studies using these tests are most likely investigating the ability to perceive facially displayed emotions. Thus, findings with oxytocin could instead be summarized to show that the hormone oxytocin improves emotion perception.

Jangle fallacies can also occur when relations between constructs are investigated without controlling for measurement error. When measurement error is properly handled—for example, by modeling constructs as latent variables (Kline, 2005)—these correlations should approach 1.0, indicating that the same construct is measured.

Recommendations

Because of jingle and jangle fallacies, scientific progress in understanding socioemotional processes has been limited. The proliferation of constructs subject to either jingle or jangle fallacies can be attributed to several causes, including incentive structures in research to discover something new, insufficient knowledge of one's field and previous work in related fields, and insufficient training in measurement. In order to bring resolution to jingle and jangle fallacies, and also to solve known problems for the assessment of socioemotional constructs, we offer some general recommendations.

Recommendation 1: embed socioemotional constructs within the larger network of psychological constructs

Existing and newly proposed socioemotional constructs should always be tested against known and established psychological constructs. A useful methodological approach is the application of a multitrait, multimethod framework, in which new constructs are assessed with multiple measures in relation to similar and dissimilar constructs (Campbell & Fiske, 1959). This framework allows researchers to evaluate agreement between measures of the same construct, the uniqueness of that construct from other established constructs, the distinction from constructs that are assumed to be dissimilar, and whether or not methodological aspects of measurement drive relations. For example, a multitrait, multimethod framework would show low convergent validity for jingle fallacies (same label but different constructs assessed) and poor discriminant validity for jangle fallacies (different labels but same construct assessed).

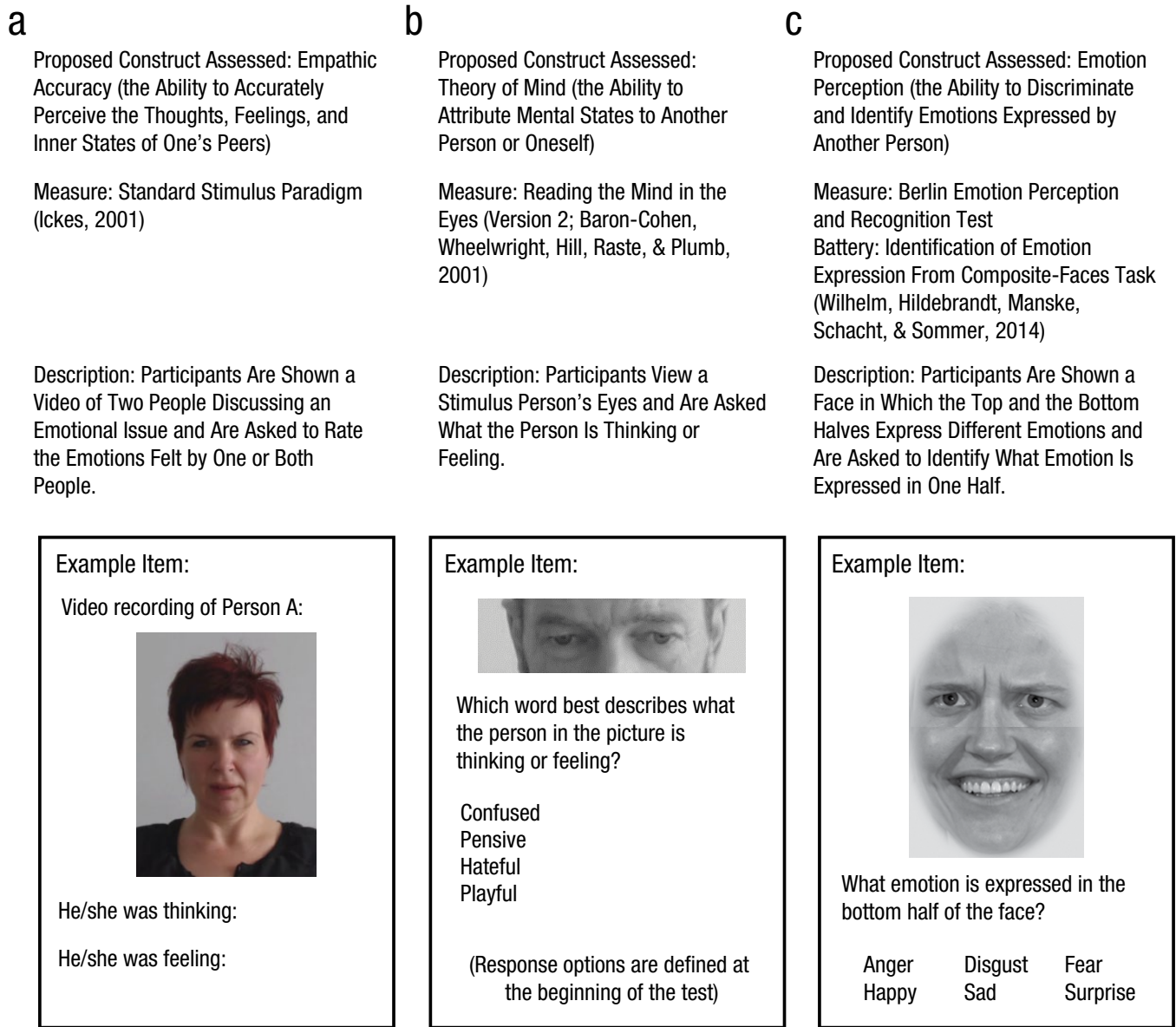


Fig. 3. Example of a jangle fallacy, which occurs when different construct labels are used, but because the measurement approach is the same, the same construct is assessed. In this example, the three measures are intended to assess empathic accuracy (a), theory of mind (b), and emotion perception (c), but because of the strong similarity in measurement design, performance on all three measures is generally strongly correlated, indicating that all three measure the same construct. The images used in (a) and (b) are similar to the types of stimuli presented in these tasks; the image in (c) is an actual picture from that task.

Recommendation 2: sanitize terminology and embrace a measurement approach

It is important that researchers use more descriptive labels for their constructs. This involves thinking carefully about the psychological process elicited by the measure and whether that construct qualifies as maximal effort or typical behavior. For example, constructs that are assessed as typical behavior should not include the terms *intelligence*, *capacity*, or *ability* in their label. Likewise, constructs that are assessed as maximal effort

should not include terms such as *motivation*, *volition*, or *preference*.

Constructs should be measured and conceptualized as either typical behavior or maximal effort, not both. Likewise, researchers should be careful that their definition of a construct matches how that construct is measured. For example, affective empathy is sometimes defined as the ability to feel what another person feels. Because it is impossible to get an objective external assessment of how a participant feels, responses cannot be compared with an objectively correct response standard. Thus, affective empathy cannot be assessed as

maximal effort, and researchers should not claim to assess it as an ability.

Recommendation 3: have a strong connection between theory and measurement

Stronger measures result from a clearer connection with a theoretical model. Models with more specific descriptions of individual-differences constructs, from which we can build better measures, are needed (Landy, 2006). For example, the branch of ability emotional intelligence that investigates the facilitation of thought using emotion includes vague descriptions of several types of reasoning involving the use of emotions to enhance thought. For example, one type of reasoning is leveraging mood swings to generate different perspectives (Mayer et al., 2016). For this type of reasoning, it is unclear what is meant by leveraging mood swings or how one would build a cognitive test for this type of reasoning.

Likewise, jingle and jangle fallacies could be minimized if test designs better matched construct descriptions. For example, the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) is purported to assess the ability to manage emotions. However, participants are asked to respond to situational judgments, and at no point do participants actually need to manage their own emotions. Thus, the connection between theory and measurement in this test is insufficient.

Recommendation 4: choose truth over consensus

Part of the challenge of developing better maximal-effect tests is identifying objectively correct responses. More work is needed to identify and validate objectively correct responses for socioemotional maximal-effort constructs. Many of these constructs use a situational-judgment test format in which participants are given a scenario and asked which behavior is the most appropriate response (Motowidlo, Hanson, & Crafts, 1997). With this format, the behavior selected by the majority of the population or a majority of experts on that topic is deemed the correct response. However, this procedure presupposes that the population or experts are correct. In fact, because individuals have their own biases and abilities, lay or expert opinions are a poor standard for declaring the objective truth of responses.

Conclusions

We argue that the study of socioemotional constructs is plagued with frequent and prominent jingle and

jangle fallacies. Here, we focused on socioemotional constructs specifically associated with the processing of emotion. However, the logic applied here extends to other constructs, such as personal and social intelligence. Bringing resolution to jingle and jangle fallacies would most likely lead to a dramatic reduction in socioemotional constructs, a culling of theories in the field, and the development of broadly accepted measures. However, the field could then focus on deepening our understanding of surviving constructs. To further support these efforts, we think the open-science movement and recent efforts to improve the reproducibility of psychological research should be extended, with more theoretical demands concerning (a) a sound elaboration of the expected convergent and discriminant correlations of a construct and (b) a discussion of how and why newly developed tasks constitute an improvement over existing measures.

Recommended Reading

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. doi:10.1017/CBO9780511490026. Discusses the concept of test validity.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230–253. doi:10.1037/a0013219. Outlines a statistical approach that can be applied to identify the existence of jingle and jangle fallacies.
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507–536. doi:10.1146/annurev.psych.59.103006.093646. Reviews the four-branch model of ability emotional intelligence.
- Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*, 267–295. Investigates jingle and jangle fallacies in emotion assessment.
- Wilhelm, O. (2005). (See References). Reviews psychometric challenges in the study of emotional intelligence.

Action Editor

Randall W. Engle served as action editor for this article.

ORCID iD

Sally Olderbak  <https://orcid.org/0000-0001-6133-0458>

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The analysis of these results and preparation of this article were supported by grants from the German Research Foundation (Deutsche Forschungsgemeinschaft; OL 452/3-1) and the Margarete von Wrangell-Habilitationsprogramm für Frauen, both awarded to S. Olderbak.

Note

1. See the following special issues or special sections on problems in emotional-intelligence research: Barchard, Brackett, and Mestre (2016); "Emotional Intelligence" (2004); McCauley (2010); Murphy (2006); Scherer & Davidson (2001); and Spector (2005). Likewise, see the book *A Critique of Emotional Intelligence: What Are the Problems and How Can They Be Fixed?* (Murphy, 2006). For problems specifically related to the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), see Maul (2012b) and the associated commentaries and reply by Mayer, Salovey, and Caruso (2012); MacCann, Matthews, and Roberts (2012); and Maul (2012a).

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Barchard, K. A., Brackett, M. A., & Mestre, J. M. (Eds.). (2016). Taking stock and moving forward: 25 years of emotional intelligence research [Special issue]. *Emotion Review*, 8(4).
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251. doi:10.1111/1469-7610.00715
- Bartz, J. A., Zaki, J., Bolger, N., Hollander, E., Ludwig, N. N., Kolevzon, A., & Ochsner, K. N. (2010). Oxytocin selectively improves empathic accuracy. *Psychological Science*, 21, 1426–1428. doi:10.1177/0956797610383439
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi:10.1037/h0046016
- Cronbach, L. J. (1949). *Essentials of psychological testing*. New York, NY: Harper. doi:10.1002/sce.3730350432
- Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biological Psychiatry*, 61, 731–733. doi:10.1016/j.biopsych.2006.07.015
- Elfenbein, H. A., Jang, D., Sharma, S., & Sanchez-Burks, J. (2017). Validating emotional attention regulation as a component of emotional intelligence: A Stroop approach to individual differences in tuning in to and out of nonverbal cues. *Emotion*, 17, 348–358. doi:10.1037/t42197-000
- Emotional Intelligence. (2004). [Special issue]. *Psychological Inquiry*, 15(3).
- Gardner, H. E. (2000). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic Books.
- Ickes, W. J. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 223–246). Mahwah, NJ: Erlbaum.
- Kelley, T. L. (1927). *Interpretation of educational measurement*. Yonkers, NY: World Book.
- Kline, R. B. (2005). *Methodology in the social sciences. Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Landy, F. J. (2006). The long, frustrating, and fruitless search for social intelligence: A cautionary tale. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 81–123). New York, NY: Erlbaum.
- MacCann, C., Matthews, G., & Roberts, R. D. (2012). Casting the first stone of validity standards: A less critical perspective of the MSCEIT. *Emotion Review*, 4, 409–410.
- Maul, A. (2012a). Higher standards of validity evidence are needed in the measurement of emotional intelligence. *Emotion Review*, 4, 411–412.
- Maul, A. (2012b). The validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) as a measure of emotion intelligence. *Emotion Review*, 4, 394–402.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, 8, 290–300. doi:10.1177/1754073916639667
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) users manual*. Toronto, Canada: Multi-Health Systems.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2012). The validity of the MSCEIT: Additional analyses and evidence. *Emotion Review*, 4, 403–408.
- McCauley, C. D. (Ed.). (2010). Emotional intelligence [Special section]. *Industrial and Organizational Psychology*, 3(2), 109–191.
- Mitchell, R. L. C., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, 70, 1–10. doi:10.1016/j.neuropsychologia.2015.02.018
- Moss, F. A., Hunt, T., Omwake, K. T., & Ronning, M. M. (1927). *Social Intelligence Test*. Washington, DC: Center for Psychological Service.
- Motowidlo, S. J., Hanson, M. A., & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement models in industrial psychology* (pp. 241–260). Palo Alto, CA: Davies-Black Publishing.
- Murphy, K. R. (Ed.). (2006). *A critique of emotional intelligence: What are the problems and how can they be fixed?* New York, NY: Erlbaum.
- Olderbak, S. G., Semmler, M., & Doebler, P. (2018). Four-branch model of ability emotional intelligence with fluid and crystallized intelligence: A meta-analysis of relations. *Emotion Review*, 11, 166–183. doi:10.1177/1754073918776776
- Olderbak, S. G., & Wilhelm, O. (2017). Emotion perception and empathy: An individual differences test of relations. *Emotion*, 17, 1092–1106. doi:10.1037/emo0000308
- Olderbak, S. G., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the Reading the Mind in the Eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, 6, Article 1503. doi:10.3389/fpsyg.2015.01503
- Petrides, K. V. (2009). *Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue)*

- (1st ed.). London, England: London Psychometric Laboratory.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92. doi:10.1177/0022022101032001009
- Scherer, K. R., & Davidson, R. J. (Eds). (2001). Emotional intelligence [Special issue]. *Emotion, 1*(3).
- Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences, 53*, 16–21. doi:10.1016/j.paid.2012.01.026
- Shahrestani, S., Kemp, A. H., & Guastella, A. J. (2013). The impact of a single administration of intranasal oxytocin on the recognition of basic emotions in humans: A meta-analysis. *Neuropsychopharmacology, 38*, 1929–1936. doi:10.1038/npp.2013.86
- Spector, P. E. (Ed). (2005). Emotional intelligence [Special section]. *Journal of Organizational Behavior, 26*, 409–466.
- Tett, R. P., Fox, K. E., & Wang, A. (2005). Development and validation of a self-report measure of emotional intelligence as a multidimensional trait domain. *Personality and Social Psychology Bulletin, 31*, 859–888. doi:10.1177/0146167204272860
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York, NY: Teacher's College, Columbia University.
- Thorndike, E. L. (1920, January). Intelligence and its uses. *Harper's Magazine*, pp. 227–235.
- Wilhelm, O. (2005). Measures of emotional intelligence: Practice and standards. In R. Schulze (Ed.), *International handbook of emotional intelligence* (pp. 131–154). Seattle, WA: Hogrefe & Huber.
- Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology, 5*, Article 404. doi:10.3389/fpsyg.2014.00404