ASSESSING SELF-REPORT SURVEY SCALES OF DISCRETE EMOTIONS:

A SYSTEMATIC REVIEW

by

Nicolas Levesque

Submitted in partial fulfilment of the requirements

for the degree of Master of Arts in School Psychology

at

Mount Saint Vincent University, Halifax, NS

September 2020

Table of Contents

Abstract

Research in the field of emotion is characterized by inconsistencies in theory and in empirical measurement of data. This systematic review of the measurement tools we use to capture said data sheds light upon the areas that require more attention to detail. Discussion of the aspects of emotion that prevail regardless of theoretical approach provides a good starting point for the reform of aligning measurement to consistent and theoretically situated definitions of emotion. Results demonstrate that there is a lack of consensus regarding which emotions are considered discrete and a vacancy in terms of scale statistics.

CHAPTER 1: LITERATURE REVIEW

It is difficult to discuss a phenomenon that has no clear definition. In the field of psychology, this is sometimes the task we are faced with. Consider your emotions. How would you define an emotion? Would you describe it as a subjective feeling or a physiological change? Does it occur as a result of your cognitive appraisal of a situation or does it lead to changes in your behaviour? Research in psychology is concerned with this definition because the study of emotion is ubiquitous throughout its practice - clinical psychologists often help clients take control of their emotions; cognitive psychologists examine the connection between emotions, judgements, and decisions; and social psychologists are interested in the impact of emotion on human relationships (Shiota & Kalat, 2012). At present, researchers and theorists in the field are engaged in debate over 3 "big" theories of emotion. Three common aspects are shared among these theories in terms of their definitions of emotion. 1) Emotions are functional in an evolutionary sense. Meaning that emotions serve a purpose in our fitness and survival. For example, fear leading to physiological changes that prepare us for action (Clark, 2013). 2) Emotions are reactions to external stimuli and are distinct from internal drives (Razavi et al., 2016). 3) Emotions include all of the above-mentioned aspects, including cognition/appraisal, subjective feeling, physiological changes, and subsequent behaviour (Shiota & Kalat, 2012). Although the field is undecided upon a single definition of emotion, these commonalities are encouraging for future consensus. The 3 "big" theories of emotion include 1) The basic emotion model, 2) Dimensional models of emotion, and 3) Emotions as compounds of underlying process.

The basic emotion model suggests that there are distinct categories of emotions such as anger, fear, or happiness. These are suggested to be evolved responses to various types of

situations, having specific functions (Tracy & Randles, 2011). Every person should have the capacity to experience these basic emotions and they should produce consistent appraisals, feelings, physiological changes, and behaviours, regardless of human differences and individual interpretation of the eliciting events. The theory proposes that each aspect of emotion is equally important and will occur with consistency for a particular emotion (Shiota & Kalat, 2012). For example, anger will produce a challenge appraisal, feelings of power, physiological activation, and lead to violent behaviours.

Contrary to the basic emotion model, the dimensional model of emotion focusses on the subjective experiences associated with an emotion. Herein, the aspects of emotion do not predictably occur in equal amounts. Rather, one aspect of emotion is considered primary while the other aspects of emotion relate to it. For example, the challenge appraisal of an anger-like feeling might occur with feelings of helplessness rather than power. Although the physiological activation might be similar, there would likely not be violent behaviours associated with this feeling. One might label this emotion as "frustration." The dimensional models of emotion posit that the "basic emotions" are more like points on a spectrum of emotion. Here, considerations such as valence and intensity are considered (Lo, Hung, & Lin, 2016), while emotions are arranged along dimensions instead of in categories.

The final major theory of emotion suggests that emotions are compounds of underlying processes. The component process theory regards emotions as being subject to drastic change when a situation presents various components. These may include unexpectedness, control, displeasure, desire to change, and many more. Therefore, any emotion can be a compound of various appraisal components. These components are suggested to be situational in nature and

develop in sequence rather than simultaneously. Thus, according to the component process model, emotions are subject to change according to the situation (Shiota & Kalat, 2012).

Unlike emotions, moods are thought of as long-lasting, less intense, global phenomena, typically not related to an object (Spering, Wagner, & Funke, 2005). Restricting the discussion to the area of emotions, I will be iterating the importance of accurate, valid, and reliable measurement tools when assessing emotions. In the area of emotion and decision-making, Fredrickson (2006) discusses the broaden-and-build theory of positive emotion. Therein, negative emotions have high predictive specificity for ensuing actions. For example, fear often leads to the decision to fight, flee, or freeze in response to a situation, as is evolutionarily appropriate. By contrast, positive emotions have very little predictive specificity for immediate action. Consider the emotion of joy: It may lead to activation in play or a simple smile but has little immediate survival benefit in the face of threat. The broaden-and-build theory suggests that negative emotions often lead to quick and decisive action while positive emotions lead to the broadening of human thought and more deliberative decision making. In turn, this broadening of thought and choice considerations helps build personal resources that can be drawn upon when later faced with difficult decisions or threats to survival, hence the "build" component of the theory (Fredrickson, 2006). Compatible with this assumption, Isen, Rosenzweig, & Young (1991) discuss the findings that positive emotional experiences are related to increased dopamine levels. Dopamine is a neurotransmitter that has been found to activate brain regions associated with thoughtfulness and the ability to switch cognitive perspectives (Ashby, Isen, & Turken, 1999; Isen, 1991). Therefore, support for the broaden-and-build theory takes shape as positive emotions have been linked to increased elaboration and cognitive flexibility, particularly related

to open-mindedness, reduced defensiveness, and the ability to efficiently engage in the decision-making process (Ashby, Isen, & Turken, 1999; Isen, 1991).

Despite growing support for theories like that proposed by Fredrickson (2006), considerable research has deemed a valence approach to emotion and decision-making insufficient. Notably, Lerner & Keltner (2000) consider the potential for specific emotions of the same valence to produce very different responses. They suggest that research using the valence approach "sacrifices specificity for parsimony." For example, it is well researched that many emotions of the same valence will produce very different physiological responses and facial expressions. Why, then, should a judgement or choice be so universally directed by these same emotions?

This field is complicated because emotions are abstract concepts differentially defined by the individuals who experience them. Emotion as it relates to decision-making is just one example of the incoherence that permeates the study of emotion. The same is visible in areas of emotion and motivation. Emotions may lead to the mobilization of behaviours (e.g., Fear leading to the pursuit of security; Anger leading to destruction; Interest leading to exploration) (Cromwell, et al., 2020). Importantly, however, emotion also involves the experience of emotional feelings (affect). Affect helps to focus attention to personal needs and desires, thus motivating behaviours to satisfy them (Tomkins, 1984). Moreover, emotions play an integral role in social communication and the production of appropriate physiological responses such as facial expressions, gestures, posture, and more (Ekman, 2003). But how does one determine which emotions are motivating the ensuing actions? Cromwell, et al., (2020) annotated 3664 words that are commonly used to describe feelings of different valences and intensities, while prescribing to a dimensional model of emotion. Throughout their analyses of these emotion words, they found

that when an emotion and a motivation can be expressed using the same word, the value of an event or outcome is clearly communicated (Cromwell, et al., 2020). However, this type of top-down analysis of linguistics, motivation, and emotion does not tell us much about how to assess an individual's emotional experience from the bottom-up. Rather, due to the simple fact that nearly four thousand different emotions can be tied to motivation - and considering the possibility that endless combinations of emotional valence and intensity may exist - this feat appears to be more distant than previously thought.

The study of negative emotions has long been an area of considerable research with a focus on differentiation and the potential for discrete emotions to elicit a wide range of responses. There is, however, a de-emphasis on doing the same for positive emotions. Although, recent research, like that of Shiota, Campos, Oveis, Herenstein, & Simon-Thomas, & Keltner (2017), have proposed new frameworks for the examination of positive emotions. Just like the emotion of disgust if a response to pathogen threat and helps to keep us safe (Cumella, 2012), positive emotions are thought to have a common core in human evolution, but developed branches in response to fitness-relevant resources (Shiota et al., 2017). In this framework, positive emotions are crucial to social functioning and supporting the relationships that humans develop with one another. In other words, those things necessary for survival and reproductive success (Shiota et al., 2017). Whereas the Broaden and Build Theory suggests that positive emotions help to build and maintain internal resources that contribute to human success, Shiota et al. (2017), elaborate upon this notion by providing compelling evidence that our positive emotions are key in developing external resources specific to the survival of humans. To that end, the ways in which we communicate positive emotions to one another is crucial to our evolutionary success. Many studies have been conducted regarding the ways we communicate

emotions to other humans. For example, Ekman (1992) coined the Duchenne Smile which is visible in the expression of many positive emotions. Similarly, head movements and postural changes, touch, and tone of voice are all effective ways of communicating emotion. Interestingly, there are notable differences in each of these communicative methods depending on what emotion is experienced (Tracy & Robins, 2007). This holds true, not only for external expressions of emotion, but also for physiological changes that are adaptive in nature. For example, the closely related emotions of love and sexual desire are commonly expressed in research settings with very different physical tendencies. Love is often expressed by a self-hug, smiling, mutual gaze, open posture, forward leaning, and head tilting. Sexual desire, on the other hand is expressed by lip licking and tongue protrusions. Although love and sexual desire are closely related in a Social Functioning perspective for reproductive interests (Shiota et al., 2017), the emotions themselves, and perhaps the associated body movements, produced different physiological responses. Gonzaga, Turner, Keltner, Campos, & Altemus (2006) found that when women were asked to recount a close relationship experience and non-verbally display the emotion of love, there was a predicted release of oxytocin. Conversely, thinking about and displaying sexual desire is associated with the peripheral release of testosterone (Shiota, 2017).

At this level of research, it becomes intuitive that positive emotions can be differentiated. Humans are able to communicate love, gratitude, and happiness through touch alone with strong reliability (Hertenstein, Holmes, McCollough, & Keltner, 2009). The same is true for the communication of amusement, interest, enthusiasm, pleasure, awe, & triumph when using only tone of voice and the absence of any semantics (Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abrahamson, 2009). Here we can see that non-verbal expression of positive emotion is likely as diverse and sophisticated as that of negative emotions. Negative emotions have been known to

evoke autonomic nervous system (ANS) responses. For example, fear leading to shallow breathing, heart racing, and muscle tightening (Sinha & Parsons, 1996), as is appropriate for a fight or flight response. Recent research has examined ANS responses as they relate to positive emotions, finding that enthusiasm leads to cardiac, vascular, and electrodermal system activation (Maruskin, Thrash, & Elliot, 2012). Awe, on the other hand, is associated with a withdrawal of cardiac influence, consistent with a preparation for stillness, as well as getting "the chills," or piloerection.

Despite an abundance of disagreement in the field of emotion, researchers continue to attempt the measurement of emotions at an individual level because it offers invaluable information about how humans operate. We may be able to use this information to help improve communication between individuals and groups, build stronger relationships in psychological practice, and maximize teaching efficiency in the classroom, among other benefits. Although we have made incredible advances broadly in the realm of emotion, there appears to be various inconsistencies and shortcomings when it comes to specificity. Measurement of emotion is often achieved using self-report scales. As noted above, the vast lexicon that can be used to describe how individuals feel creates some inherent confusion about what we are actually measuring. This begs the question of how accurate we can be in this pursuit, and certainly how we can improve upon our strategy for obtaining such pertinent information. King, Haagsma, Delfabbro, Gradisar, & Griffiths (2013) conducted a study that iterated the importance of responsible measurement and reporting. How we are measuring it can tell us something about the aspects of a given construct we are focused on or prioritize. It is researchers' responsibility to self-regulate our definitions of constructs and the quality of measurement we have available to advance knowledge.

Although emotions may be differentially defined by the various camps who attempt to study them, one of the crucial elements in this type of research is consistency of measurement. We may not have a complete understanding of emotions, but we are able to interpret the aspects of emotion and find some common ground on which to build the foundation of emotion research. We can measure emotions using self-report data, physiological measurements, and behavioural observation (Shiota & Kalat, 2012). Interestingly, these types of measurement include the four aspects of emotion that each theoretical standpoint agrees upon: cognitive appraisal, subjective feeling, physiological changes, and ensuing behaviour (self-report); physiological changes (physiological measurement); and ensuing behaviours (behavioural observation). Because measurement of any phenomenon is crucial to its study, it is encouraging that there is agreement in the field at its most basic level.

Specifically, researchers are interested to see in which contexts a measure of emotion can demonstrate reliability and validity. The reliability of a measure speaks to the level of consistency and repeatability of scores. This is important because we want to know if the measures used are able to honestly measure something across individuals. The validity of a measure speaks to the level of accuracy a measure achieves. This is important because we must be cautious that a measure could reliably achieve some data, but it may be measuring something completely different than what it set out to.

As noted above, self-report measures access information from all four aspects of emotion. Moreover, it is the only measure that can access the important subjective feeling aspect. For this reason, the current study will be assessing the level of reliability and validity of various self-report measures of emotion. Self-report measures usually ask the individual to choose a number on a Likert scale (e.g., 0 representing no agreement with a statement or question and 7

representing total agreement) which suggests their level of emotional feeling at that time. Notably, there are some common pros and cons of self-report measures in any field of research. For one, an individual reporting may differ in their level of anxiety, for example, from one time to another. Although this may suggest a lack of reliability, it could also reflect that the person's anxiety is dynamic and may actually be different from one time to the next. Another potential drawback of self-report measures is that someone may assess their anger as a 5 in a particular situation while another person may assess that same situation as evoking an anger level of 2. Importantly, however, these individuals have likely developed distinct personal resources over the course of their lives which mediate their emotional responses to various situations. Despite these drawbacks, self-report measures generate the majority of data we see in emotion research today. That is why it is important to assess these measures and become responsible researchers by using the best tools available to us.

Self-report scales are widely used to measure emotions, yet there is little consensus regarding which emotions should be considered distinct in theories that adopt a discrete emotions perspective. Thus, the reality of distinct emotion research is sustained inconsistency and imprecision of measurement. There are likely two reasons why this is the case. First, theories that adopt a discrete emotion perspective have not developed a consistent taxonomy of discrete emotions (Weidman, Steckler, & Tracy, 2017). Therefore, various scales might be measuring the same construct while using different labels for those emotions. Secondly, the lack of a consistent taxonomy means that researchers in the field of discrete emotions cannot reliably develop scales to measure a specific set of emotions (Weidman, Steckler, & Tracy, 2017). Rather, they are developing scales to examine whichever emotions they deem pertinent to the application of their research. As such, the measurement of these emotions likely occurs via scales that have not

undergone a rigorous validation process. Ideally, we would see the development of a consistent

taxonomy for discrete emotions, meaning that each emotion identified would have consistent and

unique definitions across studies. Such a taxonomy would likely expedite the development of

reliable and valid scales for measuring each of these state emotions, in turn contributing to our

knowledge base in emotion research.

At present, discrete emotion research usually speaks about 4 to 10 distinct emotions

(Weidman, Steckler, & Tracy, 2017). Interestingly, these have been generated from research that

does not involve self-report. For example, non-verbal emotion expression research postulates

that there are 10 core emotions (anger, contempt, disgust, embarrassment, fear, happiness, pride,

sadness, and surprise) (Tracy & Robins, 2008). Other fields examine observable emotions from a

cross-species neuroscientific perspective and deem 7 emotions most evolutionarily important

(care, fear, lust, panic, play, rage, and seeking) (Panksepp, 2007). Meanwhile, neuroimaging

research has identified the 4 emotions they consider to be appropriately discrete (anger, fear,

happiness, sadness) (Damasio et al., 2000). Although each of these taxonomies is based in logic

and good science, the lack of cooperation and consistency inevitably prevents the science from

progressing. In addition to the various discrete emotions suggested by the aforementioned

programs of study, self-report methods have identified a large number of emotions that may not

be visibly important from an evolutionary sense or be associated with visible changes in cerebral

activity, but that do serve important social functions. These self-report-generated taxonomies

identified 25 lower-order clusters of discrete emotions that belong to the families of 6 higher

order clusters, including anger, fear, joy, love, sadness, and surprise. Examples of lower-order

discrete emotions include jealousy and pride. Therefore, based upon this taxonomy (6 higher-

order and 25 lower-order emotions) we might expect to see a range of 6 to 31 distinct emotions

cited in the literature of discrete emotion research. Interestingly, Weidman, Steckler, & Tracy (2017) found that recent research in the field of discrete emotions has used over 65 different emotion words in their scales. This mismatch between the taxonomy of discrete emotions and the empirical measures of the same certainly propagates measurement error and various inconsistencies that may be corrected by some cohesion.

Objective.

The objective of this thesis is to synthesize approaches to emotion measurement. A cohesive summary of the range of measures used in research focused on emotion can help us understand discrepancies in the field and inform how to further improve measurement of emotion. Noting relative strengths and weaknesses of the psychometric properties of emotion measurement will facilitate psychologists' and educators' critical appraisal of research on emotion as related to human motivation, well-being, mental illness, decision-making and learning. With increased efforts to engage in evidence-based practice, it is important to communicate findings with a critical lens so as to facilitate accurate knowledge utilization efforts. This study asks: (1) What are the psychometric properties of published emotion scales? (2) Which demographic populations are represented in these studies? (3) In what contexts are the scales validated and in what contexts are they used for research purposes? (4) Which emotions are included across scales and which are rarely to never represented?

CHAPTER 2: CURRENT STUDY

Methods.

The focus of this scoping review is on studies that somehow assess the validity or reliability of self-report survey scales of discrete emotions. These studies came from peer-

reviewed journals, websites reporting about and sharing scales, and mental measurements

yearbook. Given the complexity of language around emotions, this review was limited to only

English-language studies and scales. Inclusion criteria also included reports of primary data that

provided indication of validity or reliability of the measurement in the context of specified

demographic populations. Critical to demographic data are age, sex, and ethnicity. Any

additional demographic or contextual details were also synthesized as reported in the studies.

Given emotions are discrete and contextualized physiological reactions with associated personal

experience, vernacular and attributions around them, this review excluded scales that measure

general mood or affect over a broad range of time. Scales that have psychometric property data

protected for proprietary reasons were also excluded. Surveys that required payment to be used

were included only if we had access to primary reports of the demonstrated psychometric

properties of the scale.

The search criteria were established in consultation with a university librarian adept at

complex databases and grey literature. Databases include PsycInfo, Education Research

Complete, Academic Search Premier, and Google Scholar. Although, the latter two databases

generated no unique "hits" and were thus excluded from our final search strategy. The exact

search criteria that provided the most specificity are reported as follows:

PsycINFO Search

DE "Emotional States" OR DE "Expressed Emotion" OR DE "Negative Emotions" OR DE
"Positive Emotions"

AND

DE "Diary Measure" OR DE "Self-Report" OR DE "Data Collection" OR DE "Methodology"
OR DE "Likert Scales" OR DE "Retrospective Assessment" OR DE "Self-Evaluation"

NOT

DE "Psychoanalytic Personality Factors" OR DE "Emotional Instability" OR DE "Aromatherapy" OR DE "Mood Stabilizers" OR DE "Disruptive Mood Dysregulation Disorder" OR DE "Affective Disorders" OR DE "Euthymia" OR DE "Bipolar Disorder" OR DE "Seasonal Affective Disorder" OR DE "Premenstrual Dysphoric Disorder" OR DE "Mania" OR DE "Major Depression" OR DE "Hypothyroidism" OR DE "Epilepsy" OR DE "Endogenous Opiates" OR DE "Cyclothymic Disorder" OR DE "Carbamazepine" OR DE "Borderline Personality Disorder" OR DE "Affective Psychosis"

NOT

Disorder

AND

Validity OR Reliability

Limiters - Full Text; Scholarly (Peer Reviewed) Journals; Published Date: 20000101-20201231
Expanders - Apply related words; Apply equivalent subjects
Search modes - Boolean

Education Research Complete

DE "EMOTIONS" OR DE "Emotion States"
AND
Self-Report OR Scale OR questionnaire
NOT
DE "AFFECTIVE disorders" OR DE "CONTEXT effects (Psychology)" OR DE "COGNITIVE therapy"
NOT
Disorder OR diagnosis OR diagnostics OR alexithymia or alexithymic
AND
Validity OR Reliability
NOT
Trait OR mood OR regulation
NOT
nursing or nurses or nurse or nursing care or nursing practice
NOT
regulation OR self-control OR behavioral assessment
NOT
attitudes or wellbeing or well-being or well being or mental health
NOT
respiration OR heartbeat OR skin conductivity

Limiters - Full Text; Scholarly (Peer Reviewed) Journals; Published Date: 20000101-20201231
Expanders - Apply related words; Apply equivalent subjects

Search modes – Boolean

Once the search was complete, each result was screened according to the inclusion and exclusion criteria. Any articles from this search that referenced other studies assessing psychometric properties of the measure they used were also retrieved through the available library, or through a request for interlibrary loans and summarized as part of this scoping review.

Analysis.

Once screened for inclusion, each measure was critically appraised by the collection of existing evidence on it for key features of contextualized validity and reliability, including information regarding dimensionality, demographics, sample size, details regarding scale specificity, and theoretical basis. This information was reported in a table categorized by measure and reported by each relevant study for transparency (See Appendices A and B).

Results.

Following the search strategy, 59 articles from PsycINFO and Education Research Complete were screened for inclusion and exclusion criteria. Of these 59 articles, 17 met the criteria, leaving us with 12 scales that assess discrete emotions to be reviewed. Of these, 6 assessed dimensionality using a latent variable analysis to assess factor structure of subscales within the survey. Eight of the 12 surveys explicitly assessed validity, and all surveys had reliability estimates, primarily being Cronbach's alpha internal consistency estimates. Sample sizes of studies assessing the psychometric properties of surveys ranged from $N = 54$ to $N = 4250$.

Demographics.

The vast majority of studies collected data from students (university -8, secondary -2, middle -2, & elementary - 1) while others conducted their research with working populations over the age of 18. Many of the studies reported that their sample consisted primarily of women living in the US. Few studies focused on participants in the UK and other European countries. Of the studies in the US and UK, the sample was largely Caucasian.

Emotions measured.

Emotions measured in surveys ranged from high specific focus on just one emotion, to a range of up to 20 different emotion terms listed for participants to rank. All items to measure emotions were Likert scales, ranging from 5-to-7-point items. One survey (Lichtenfeld, Pekrun, Stupnisky, Reiss, & Murayama, 2012) was for children and used faces to help the children make the rankings that were coded on a Likert-scale. Many of the emotions being measured in the articles appraised deviated from the above-mentioned 31 emotions expected in discrete emotions research. The lack of a consistent taxonomy and matching scale development leaves something to be desired in terms of a consensus.

Subscales.

The number of items in the surveys ranged from 12 to 232. With measurement goals of parsimony and user-friendly surveys, 232 items might be overwhelming, but 12 might be too simplistic. Future studies are warranted regarding the assessment of parsimony and what is best for this type of self-report survey (e.g. optimal brevity while still being comprehensive and aligned to operational definitions of complex emotions).

Reliability and validity.

In terms of reliability, all internal consistency estimates reported were over 0.7, with most being above 0.85. Validity was assessed in a number of ways, including expert ratings of items

17

and assessment of the agreement between them, self-other reports of an individual's emotional tendencies to verify consistency, comparisons to pre-existing scales such as the PANAS, assessment of the relationship between a scale and an established index of the emotion in question, and various correlational data. In many cases, however, there was little-to-no discussion of scale validity. These results are indicative of the pervasive incoherence that epitomizes emotion research.

Theoretical approach.

Of the 17 studies assessed, only 9 made clear their theoretical approach. One study mentioned having a basic emotion model background while another discussed their subscription to the dimensional models of emotion. However, each of these conducted their study in terms of discrete emotion measurement. Interestingly, this exemplifies the disconnect between theory and empirical research yet again.

Discussion.

Though various scales for assessing state emotions are being used somewhat often in psychology and education research, it is clear that little has been done in the past twenty years to ensure measurement quality. Thus, inferences derived from these studies, whether it be in research involving test anxiety or response to disgust, must be received with skepticism. First, the demographics discussed above have implications for the generalizability of this research. As these studies have primarily focused on white university students, we see a sample that is not representative of the broader population. Leaving out minority groups contrives the research to only one small piece of the *human* emotional experience. When we conduct research using samples such as we see here, we not only lose credibility in the statistical sense, but we contribute to the ethical issue of systematically dehumanizing minority groups by leaving them

out of the conversation. Additionally, good research seeks to inform practice. If emotion research seeks to inform mental health practitioners, are we doing justice to those minority groups by basing our practice upon research that is applicable only to a specific ethnic and socioeconomic group?

Second, this research uses highly variable approaches to latent scale analysis with discussions of validity rarely going beyond assumptions that the scales developed a-priori are as sufficient. We see insufficient sample sizes and at no point did we see an attempt to replicate results on an independent sample. In terms of reliability, internal consistency was often reported, but there is rarely enough information generated to determine test-retest reliability or possible dual-coder reliability (e.g., Self-report versus peer report of your feelings; Research rating from facial expression). Moreover, Cronbach's alpha, as a statistical analysis, is inflated by a high number of survey items and large sample sizes. Thus, internal consistency is the low-hanging fruit of psychometrics and researchers in any discipline should be held to a higher standard of reporting.

Scale development should consider the number of items to include. In this review, we examined scales that had either far too many or far too few items (the latter being a more common trend). This is important because the emotions we are attempting to investigate and inherently nuanced constructs that involve a variety of complex appraisals. Take the emotion of joy, for example, which may involve feeling friendly toward other, physical animation, and feeling a sense of belonging (Weidman, Steckler, & Tracy, 2017). To capture the emotion of joy, there is a need to have an appropriately large number of items in a given scale to account for these various nuances. Scales that are too short leave a lot to be accounted for by single items. Because we are conducting research on individuals who are all unique, single items may be

19

subject to responses that vary based upon idiosyncrasies and not actually reflect the emotion itself. Importantly, scales that are too long also run the risk of losing parsimony and adding complexity that need not be, which makes replicability difficult and takes away from the goal of creating a consensus of emotions in the field's taxonomy.

Finally, we see a lack of transparency when it comes to the theoretical foundations of the research being conducted. Those who do state which frameworks their scales are based upon appear to have adopted an approach that is inconsistent with a discrete emotions model, yet attempt to measure state emotions, nonetheless. As discussed above, it is crucial to the development of scales that theoretical approaches develop consistent definitions of emotion and taxonomies of the like. Concerning the current review, there is a veritable smorgasbord of language and theory that has been thrown together. The research examined appears to bounce around from one theoretical framework to the next in an attempt to cover all of their bases. This occurrence is likely not the intention of individual researchers, but a product of the current state of emotion research.

Recommendations.

First and foremost, as discussed throughout this text, it is of utmost importance for the field of emotion research to arrive at a consensus taxonomy of discrete emotions in theory and to begin developing scales that match accordingly. Weidman, Steckler, and Tracy (2017) arrived at the number 31 when discussing 6 higher-order emotions and 25 supplemental, more nuanced emotions. Perhaps theorists of emotions should begin by examining these 31 emotions, assessing their uniqueness and evolutionary appropriateness. Arriving at a consistent taxonomy of discrete emotions is paramount to the development of scales that are meaningful. Otherwise, the word

discrete becomes a particularity of language rather than a true and meaningful description of the emotions being measured.

Important to the development of these scales is the transparency of statistics reported. Beyond reporting internal consistencies with Cronbach's alphas, researchers could make more progressive contributions by focusing on replicability and conducting trials on independent samples, searching to establish test-retest reliability as well as dual-coder information of the scales discussed. Significantly, we are seeing a lack of evidence throughout the research that supports the validity of these scales. With the initiative of creating consistent taxonomies and replicable scales, it will be possible to acquire more meaningful information about the scales' validity. Moreover, the number of items in a given scale should reflect the number and complexity of the emotions being measured. It is important to create scales that are neither insufficient, nor too cumbersome, to adequately capture the essence of what it is we are attempting to measure.

Conclusion.

What we know about any given phenomenon is linked to quality of measurement tools available to researchers. How can we know anything about the role of emotion on human performance, a sad child taking a diagnostic test with a school psychologist for example, if we do not measure sadness well in research? What we know about human emotions from an empirical sense is only as good as the tools we use to measure it. We need to identify strengths and room for further development. We need to be aware of which emotions are measured, and who these scales are (and thus are not) standardized for in order to make improvements in our measurement efforts moving forward.

With this scoping review of discrete emotion measures, the disconnect between theory and practice has become evident. While one remains ambiguous, the other cannot develop. At this point in time, discrete emotions have found their place floating between various theories without a foothold on common ground. Perhaps the most alarming discovery of this process was the lack of reporting throughout these studies. Many articles claiming that their scales are valid or reliable without providing evidence can only lead to the tendency of others to do the same. Here, I will revisit the importance of responsible research. Those measures that we use to draw conclusions about any phenomenon are critical to the truths that we are attempting to uncover. With inconsistent measures of emotion, we are inevitably creating culture of dishonest research. It is the responsibility of researchers in the field of emotions to take an interest in arriving at a consensus taxonomy and developing scales that may begin to tell us something meaningful.

References

Ashby, G. F., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106(3)*, 529-550.

Barabadi, E., & Khajavy, G. H. (2020). Perfectionism and foreign language achievement: The mediating role of emotions and achievement goals. Studies in Educational Evaluation, 65, 100874. https://doi.org/10.1016/j.stueduc.2020.100874

Barrett, L. F., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional Experience. Journal of Personality and Social Psychology, 87(5), 684–697. https://doi.org/10.1037/0022-3514.87.5.684

Bates, B. R., & Chadwick, A. E. (2015). Measuring State Disgust: Initial Evidence for the Reliability and Validity of an English-Language Version of the Ekel-State-Fragebogen. Communication Research Reports, 32(1), 73–82. https://doi.org/10.1080/08824096.2014.989978

Bongard, S., Martin, N. M., Seip, M., & al'Absi, M. (2011). Evaluation of a Domain-Specific Anger Expression Assessment Strategy. Journal of Personality Assessment, 93(1), 56–61. https://doi.org/10.1080/00223891.2010.513705

Clark, J. (2013). Integrating basic and higher-cognitive emotions within a common evolutionary framework: Lessons from the transformation of primate dominance into human pride. *Philosophical Psychology, 26(3),* 437-460.

Crawford, J., Cayley, C., Lovibond, P. F., Wilson, P. H., & Hartley, C. (2011). Percentile Norms and Accompanying Interval Estimates from an Australian General Adult Population Sample for Self-Report Mood Scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21,

STAI-X, STAI-Y, SRDS, and SRAS): Percentile norms for mood scales. *Australian*

Psychologist, 46(1), 3–14. https://doi.org/10.1111/j.1742-9544.2010.00003.x

Cromwell, H. C., Nobuhito, A., Barrett, K. C., Caldwell-Harris, C., Gendolla, G. H. C., Koncz,

R., & Sachdev, P. S. (2020). Mapping the interconnected neural systems underlying

motivation and emotion: A key step toward understanding the human affectome.

Neuroscience & Biobehavioural Reviews, 113, 204-226.

https://doi.org/10.1016/j.neubiorev.2020.02.032

Cumella, E. J. (2012). Fear factor unveiled: The biology and psychology of disgust.

PsycCRITIQUES, 57(24).

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., &

Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated

emotions. Nature Neuroscience, 3, 1049–1056.

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve*

Communication and Emotional Life. New York, NY: Times Books/Henry Holt and Co.

Fredrickson, B. (2006). *The broaden-and-build theory of positive emotions.* New York, NY:

Oxford University Press.

Fuller-Tyszkiewicz, M., Hartley-Clark, L., Cummins, R. A., Tomyn, A. J., Weinberg, M. K., &

Richardson, B. (2017). Using dynamic factor analysis to provide insights into data

reliability in experience sampling studies. *Psychological Assessment, 29*(9), 1120–1128.

https://doi.org/10.1037/pas0000411

Gonzaga, G. C., Turner, R. A., Keltner, D., Campos, B., & Altemus, M. (2006). Romantic love

and sexual desire in close relationships. *Emotion, 6, 163–179.*

Hertenstein, M. J., Holmes ,R., McCullough, M.,& Keltner, D.(2009). The communication of

emotion via touch. *Emotion, 9,* 566–573.

Isen, A. M., Rosenzweig, A. S., & Young, M. J. (1991). The influence of positive affect on
clinical problem solving. *Medical Decision Making, 11*(3), 221–227.
https://doi.org/10.1177/0272989X9101100313

Jovanović, V., Lazić, M., Gavrilov-Jerković, V., & Molenaar, D. (2019). The Scale of Positive
and Negative Experience (SPANE): Evaluation of Measurement Invariance and
Convergent and Discriminant Validity. *European Journal of Psychological Assessment*,
1–11. https://doi.org/10.1027/1015-5759/a000540

Kelley, T. L. (1927). *Interpretation of Educational Measurement*. Yonkers, NY: World Book.

King, D., Haagsma, M. C., Delfabbro, P. H., Gradisar, M., Griffiths, M. D. (2013). Toward a
consensus definition of pathological video-gaming: A systematic review of psychometric
assessment tools. *Clinical Psychology Review, 33*(3), 331 – 342. doi:
10.1016/j.cpr.2013.01.002

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific
influences on judgement and choice. Cognition & Emotion, 14(4), 473–493.
https://doi.org/10.1080/026999300402763

Lichtenfeld, S., Pekrun, R., Stupnisky, R. H., Reiss, K., & Murayama, K. (2012). Measuring
students' emotions in the early years: The Achievement Emotions Questionnaire-
Elementary School (AEQ-ES). Learning and Individual Differences, 22(2), 190–201.
https://doi.org/10.1016/j.lindif.2011.04.009

Lischetzke, T., Cuccodoro, G., Gauger, A., Todeschini, L., & Eid, M. (2005). Measuring
Affective Clarity Indirectly: Individual Differences in Response Latencies of State.
Emotion, 5(4), 431–445. https://doi.org/10.1037/1528-3542.5.4.431

Lo, L. Y., Hung, N. L., & Lin, M. (2016). Angry versus furious: A comparison between valence and arousal in dimensional models of emotion. The Journal of Psychology: Interdisciplinary and Applied, 150(8), 949-960.

Maruskin, L. A., Thrash, T. M., & Elliot, A. J. (2012). The chills as a psychological construct: Content universe, factor structure, affective composition, elicitors, trait antecedents, and consequences. Journal of Personality and Social Psychology, 103, 135–157.

Mason, L., Scrimin, S., Tornatora, M. C., & Zaccoletti, S. (2017). Emotional reactivity and comprehension of multiple online texts. Learning and Individual Differences, 58, 10–21. https://doi.org/10.1016/j.lindif.2017.07.002

McCarthy, C. J., Moller, N. P., & Fouladi, R. T. (2001). Continued Attachment to Parents: Its Relationship to Affect Regulation and Perceived Stress Among College Students. Measurement and Evaluation in Counseling and Development, 33(4), 198–213. https://doi.org/10.1080/07481756.2001.12069011

Okamoto, S. K. (2001). Instrument development: Practitioner fear of youthful clients: An instrument development and validation study. *Social Work Research*, *25*(3), 173–184. https://doi.org/10.1093/swr/25.3.173

Olderbak, S. & Wilhelm, O. (2017). Emotion Perception and Empathy: An Individual Differences Test of Relations. *Emotion, 17*, 1092-1106.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. Perspectives on Psychological Science, 2, 281–296.

Peixoto, F., Mata, L., Monteiro, V., Sanches, C., & Pekrun, R. (2015). The Achievement Emotions Questionnaire: Validation for Pre-Adolescent Students. *European Journal of*

Developmental Psychology, *12*(4), 472–481.

https://doi.org/10.1080/17405629.2015.1040757

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, *36*(1), 36-48.

Razavi, P., Zhang, J. W., Hekiert, D., Yoo, S. H., & Howell, R. T. (2016). Emotion, 16(8), 1097-1101.

Shiota, M. N., Campos, B., Oveis, C., Herenstein, M. J., Simon-Thomas, E., & Keltner, D. (2017). Beyond happiness: Building a science of discrete positive emotions. American Psychologist, 72(7), 617-643.

Shiota, M. N. & Kalat, J. W. (2012). *Emotion (second edition).* Linda Schreiber-Ganster.

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays*. Emotion, 9,* 838–846.

Sinha, R. & Parsons, O. A. (1996). Multivariate response patterning of fear and anger. *Cognition and Emotion, 10(2),* 173-198.

Spering, M. N., Wagener, D., & Funke, J. (2005). The role of emotions in complex problem-solving. Cognition and Emotion, 19(8), 1252–1261.

https://doi.org/10.1080/02699930500304886

Stewart, M. E., Watson, R., Clark, A., Ebmeier, K. P., & Deary, I. J. (2010). A hierarchy of happiness? Mokken scaling analysis of the Oxford Happiness Inventory. *Personality and Individual Differences*, *48*(7), 845–848. https://doi.org/10.1016/j.paid.2010.02.011

Thorndike, E. L. (1904). An introduction to the theory of mental and social measurements. New

      York, NY: Teacher's College, Columbia University. http://dx.doi.org/10.1037/13283-000

Tomkins, S. S. (1984). Affect Theory. *Approaches to Emotion, 11,* 163-195.

Tracy, J. L. & Randles, D. (2011). Four models of basic emotions: A review of Ekman and

      Cordero, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3(4*), 397-405.

Tracy, J. L., & Robins, R. W. (2007). The prototypical pride expression: Development of a

      nonverbal behavior coding system. Emotion, 7, 789– 801.

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-

      cultural recognition. Journal of Personality and Social Psychology, 94, 516–530.

Walden, T. A., Harris, V. S., & Catron, T. F. (2003). How I Feel: A Self-Report Measure of

      Emotional Arousal and Regulation for Children. *Psychological Assessment*, *15*(3), 399–

      412. https://doi.org/10.1037/1040-3590.15.3.399

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(2),* 267-295.

Survey Description

Table 1. Description of self-report discrete emotion scales and participants studied.

Source	Number of items	Demographics	Sample Size	Theoretical Grounding
Achievement Emotions Questionnaire				
Pekrun, Goetz, Frenzel, Barchfeld, & Perry (2011)	324 (24 subscales of 9 emotions in 3 contexts)	University students In Canada taking undergraduate psychology courses, mean age = 20.63 (SD = 3.48). Gender represented as binary with 60.15% female.	389	Control-Value Theory
Achievement Emotions Questionnaire for Pre-Adolescents				
Pekrun, Goetz, Titz, & Perry (2002)	8	University and School-Aged Students	452	Control-Value Theory
Barabadi, & Khajavy (2020)	24	Secondary students in Iran, taking English courses	2008	Control-Value Theory
Achievement Emotions Questionnaire: Elementary School				
Lichtenfield, Pekrun, Stupinsky, Reiss, & Murayama (2012)	28	Third grade students in USA Mean age = 8.69. Predominantly Caucasian, specific minority identities not specified	163	Control-Value Theory
Positive and Negative Affect Scale				
Mason, Scrimin, Tornotora, & Zaccoletti (2017)	30	Seventh grade students in Italy	104	Not reported
Barrett, Quigley, Bliss-Moreau, & Aronson (2004)		Undergraduate psychology majors in USA	54	Not reported
Crawford, & Henry (2004)		Non-clinical sample of adults broadly representative of the UK population	1003	Tripartite Theory

Source	Number of items	Demographics	Sample Size	Theoretical Grounding
Practitioner Fear of Youthful Clients				
Okamoto (2001)	42	Practitioners from 15 agencies with varied educational backgrounds and ethnicities in Hawaii	248	Not reported
German State Disgust Scale				
Bates, & Chadwick (2015)	12	Undergraduate students age 18 to 22, 74 of 78 participants were Caucasian	78	Not reported
Trait Affect Scale				
Fuller-Tyszkiewicz, Hartley-Clark, Cummins, Tomyn, Weinber, & Richardson (2017)	5	University students	127	Multiple Discrepancies Theory
Scale of Positive and Negative Experience (SPANE)				
Jovanovic, Lazic, Gavrilov-Jerkovic, & Molenaar (2017)	12	Adults from Serbia, mean age = 28.56, 61.95% female	4250	Not reported
Oxford Happiness Inventory				
Stewart, Watson, Clark, Ebmeier, & Deary (2010)	12	University students in Edinburgh, 17 years and older	1024	Not Reported
Multidimensional Mood Questionnaire				
Todeschini, & Eid (2005)		University students in Germany, mean age = 22.3 years, 86% women	85	Dimensional Model
State-Trait Anger Expression Inventory				
Bongard, Martin, Seip, & al'Absi (2011)	24	Adults (18 years and older) from Germany who had an intimate partner and was employed for at least 6 months	86	Not reported

Source	Number of items	Demographics	Sample Size	Theoretical Grounding
How I Feel				
Walden, Harris, & Carton (2003)	30	Seven to thirteen-year-old students in USA from schools in low SES neighbourhoods	628	Basic Emotions Model
Perceived Stress Scale				
McCarthy, Moller, Fouladi (2001)	14	University students, mean age = 21.6 years	200	Attachment Theory

****Deomgraphics reports who participants were of studies assessing the psychometric properties of the scales. This includes age, sex, ethnicity, and any other demographics reported.

Dimensionality, Validity, and Reliability

Table 2. Psychometric properties of self-report discrete emotion scales.

Source	Emotions Assessed	Dimensionality	Validity	Reliability
Achievement Emotions Questionnaire				
Pekrun, Goetz, Frenzel, Barchfeld, & Perry (2011)	Enjoyment, hope, pride, shame, relief, anger, anxiety, hopelessness, boredom	Multiple CFAs to test the structure for each of 9 emotions with CFI = 0.88 to 0.99; GFI = 0.83 top 0.99; and RMSEA = 0.049 to 0.15	Construct validity assessing the emotions relation to theorized antecedents and consequences. Convergent validity, correlating to anxiety scales.	Internal Consistency estimate Cronbach's alpha for subscales ranged from = 0.75 (moderate) to 0.93 (strong)
Achievement Emotions Questionnaire for Pre-Adolescents				
Pekrun, Goetz, Titz, & Perry (2002)	Enjoyment, pride, hope, relief, anger, anxiety, shame, hopelessness, boredom	CFA moderate fit (r = 0.58 for university sample and r = 0.64 for school-aged sample)	Conclusions drawn about validity, but no statistics	Internal Consistency estimate Cronbach's alpha = 0.86 and 0.87
Barabadi, & Khajavy (2020)	Boredom, hopelessness, anger, anxiety, enjoyment, pride	CFA very strong fit to data (factor loadings from 0.57 to 0.96, CFI = 0.966, TLI = 0.961, RMSEA = 0.038, SRMR = 0.039)	Not reported	Not reported

Source	Emotions Assessed	Dimensionality	Validity	Reliability
Achievement Emotions Questionnaire: Elementary School				
Lichtenfield, Pekrun, Stupinsky, Reiss, & Murayama (2012)	Enjoyment, anxiety, boredom	CFA-Hierarchical model good fit (SRMR = 0.62, CFI = 0.91, TLI = 0.90, RMSEA = 0.084)	Elementary aged students demonstrated ability to distinguish between different emotions	Internal Consistency estimate Cronbach's alpha for subscales ranged from = 0.75 (moderate) to 0.95 (strong)
Positive and Negative Affect Scale				
Mason, Scrimin, Tornotora & Zaccoletti (2017)	Positive affect (including happy, cheerful, enthusiastic, peppy, satisfied, calm, quiet, still), Negative affect (including sad, fear, sluggish, disappointed, nervous, afraid). 20 items	Not reported	Demonstrated significant difference between positive and negative affect	Internal consistency estimate Cronbach's alpha, positive affect = 0.88; negative affect = 0.92
Barrett, Quigley, Bliss-Moreau, & Aronson (2004)		Two dimensional MDS solution from INDSCAL procedure, $R^2 = 0.75$)	Discriminant validity demonstrated interoceptive sensitivity predicted emotional experience	Results replicated in two studies reported in this paper Congruence coefficients were strong (0.92 and 0.94).
Crawford, & Henry (2004)		Correlated two-dimensional model (positive and negative affect), RCFI = 0.94, RMSEA = 0.058, SRMR = 0.052	Demonstrated significant linear relation with depression and anxiety scales	Internal consistency estimate Cronbach's alpha, positive affect = 0.85; negative affect = 0.89

Source	Emotions Assessed	Dimensionality	Validity	Reliability
Practitioner Fear of Youthful Clients				
Okamoto (2001)	Fear	EFA with a four-factor solution accounting for 41 % of variance.	Predicted relationships between subscales were confirmed.	Internal Consistency estimate Cronbach's alpha for subscales ranged from = 0.73 to 0.87 (moderate)
German State Disgust Scale				
Bates, & Chadwick (2015)	Disgust	CFA demonstrated adequate fit, CFI = 0.94, RMSEA = 0.10, SRMR = 0.06	Strong positive correlation with another disgust index, r = 0.75, p < 0.05	Cronbach's alpha for subscales ranged from = 0.91 to 0.95 (strong)
Trait Affect Scale				
Fuller-Tyszkiewicz, Hartley-Clark, Cummins, Tomyn, Weinber, & Richardson (2017)	Happy, content, excited, satisfied, discontent	CFA demonstrated weak fit, CFI = 0.90, NFI = 0.88, RMSEA = 0.09, SMC = 0.53	Not reported	Cronbach's alpha = 0.88
Scale of Positive and Negative Experience (SPANE)				
Jovanovic, Lazic, Gavrilov-Jerkovic, & Molenaar (2017)	Satisfaction with life, depression, meaning in life, search for meaning	CFA Demonstrated 2-factor model, CFI = 0.95	Strong positive correlations with PANAS, Life Satisfaction and Depression	Internal consistency estimate Cronbach's alpha, positive affect = 0.90; negative affect = 0.84
Oxford Happiness Inventory				
Stewart, Watson, Clark, Ebmeier, & Deary (2010)	Happiness	Not Reported	Not tested, though authors suggest to measure this scale against other measures of happiness for validation	Hierarchical scale H > 0.3, 0.12. Good internal consistency and high over time

Source	Emotions Assessed	Dimensionality	Validity	Reliability
Multidimensional Mood Questionnaire				
Todeschini, & Eid (2005)	Happy, uncomfortable, contended, tense, relaxed, uneasy, rested, sleepy, energetic	Not reported	Not reported	Internal consistency estimate Cronbach's alpha ranged from 0.90 to 0.94
State-Trait Anger Expression Inventory				
Bongard, Martin, Seip, & al'Absi (2011)	Anger	Not reported	Compared self-reports to reports of work colleagues or intimate partners.	Internal consistency estimate Cronbach's alpha ranged from 0.74 to 0.91
How I Feel				
Walden, Harris, & Carton (2003)	Happy, Excited, Sad, Scared, Mad	CFA with three independent samples supported a three-factor model of a) frequency and intensity of positive emotion, b) frequency and intensity of negative emotion, and c) emotion-control	Ten experts rated the scale with 96.4% agreement with all the items.	Internal consistency estimate Cronbach's alpha ranged from 0.86 to 0.90 Moderately stable scores for 120 participants over 2 years
Perceived Stress Scale				
McCarthy, Moller, & Fouladi (2001)	Stress	Not reported	Correlated with social anxiety, $r = 0.52$ and physiological symptom, $r = 0.76$	Internal consistency estimate Cronbach's alpha = 0.87

*Dimensionality is noted with goodness of fit indexes when a factor analysis is reported.

**Validity is specified in each cell based on what is reported in primary studies (Convergent, Predictive, Construct, Expert, Community-voice)

***Reliability includes internal consistency estimates, inter-rater or intra-rater reliability, and test-retest reliability when they are reported in the literature.